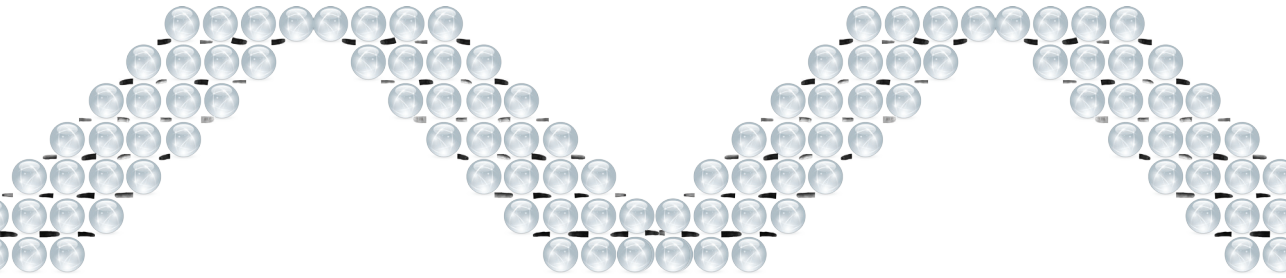
A phosphoproteomics view at human pluripotent stem cells



Adja Zoumaro-Djayoon

Colophon

The research in this thesis was performed in the Biomolecular Mass Spectrometry and Proteomics group, Utrecht University, the Netherlands.

Financial support from the Jurriaanse stichting for printing this thesis is kindly acknowledged.

Cover photo: Design by Adja Zoumaro-Djayoon

Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress

ISBN: 978-90-8891-608-3

Dedicated to my Parents Lantame and Ikpindi

A mes parents Lantame et Ikpindi

A phosphoproteomics view at human pluripotent stem cells

Inzicht in humane pluripotente stamcellen m.b.v. phosphoproteomics

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdediging op woensdag 15 mei 2013 des middags te 12.45 uur

door

Adja Damba Zoumaro-Djayoon

geboren op 12 juni 1982 te Sokodé, Togo

Promotor: Prof. dr. A.J.R. Heck

Co-promotor: Dr. J. Muñoz

Table of contents

Chapter I

General introduction

1. **Human pluripotent stem cells**

Stem cells are unspecialized cells that possess two unique properties which distinguish them from other cells. They have the potential to self-replicate through cell division and to differentiate into other cell types. Several types of stem cells exist which are classified according to their differentiation potential. The fertilized egg (zygote) is a totipotent stem cell as it has the ability to differentiate in all cells of an adult organism. In embryogenesis the zygote gives rise to the extra-embryonic tissues (e.g. placenta) and to pluripotent stem cells (inner cell mass). The pluripotent stem cells can differentiate into the three embryonic germ layers, mesoderm, endoderm and ectoderm. The aforementioned cells can differentiate into multipotent or unipotent stem cells which are cells that can further differentiate respectively into diverse types of somatic cells or one specific cell lineage as illustrated in Figure 1.

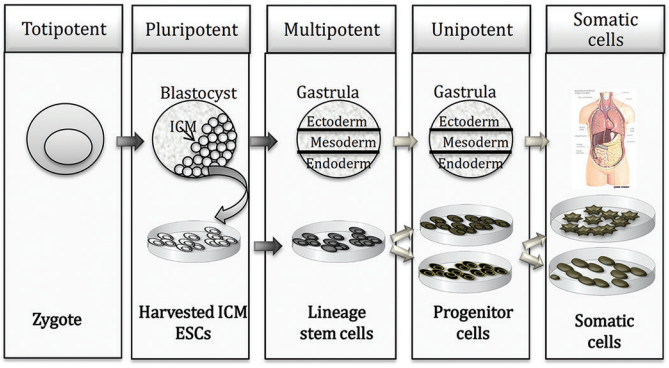


Figure 1: *Derivation of embryonic stem cells from the inner cell mass (ICM) of a blastocyst (embryo). The potency of each cell type is indicated and the arrow point to the differentiation direction of stem cell type. The zygote generates a blastocyst containing an inner cell mass (ICM). The blastocyst then differentiates into the gastrula containing the three germ layers; ectoderm, mesoderm and endoderm. The multipotent stem cells in these cell layers can further differentiate into somatic cells. The ICM can also be isolated and cultured in vitro to generate embryonic stem cells (ESCs) that are able to differentiate into progenitor cells and eventually into somatic cells.*

In 1981, pluripotent embryonic stem cells (ESCs) were extracted from the mouse blastocyst (*1*) to be used as an *in vitro* model for early mammalian development. The ability of mouse ESCs (mESCs) to develop into an embryo and to proliferate in an undifferentiated state, has led to the derivation of

hESCs from 3-5 days-old embryos (*2*). To this end, the inner cell mass (ICM) of these embryos (blastocysts) are isolated and plated *in vitro* as embryonic stem cells (Figure 1). The self-renewal and pluripotency characteristics of hESCs enable their unlimited expansion and allow the generation of multiple cell types from a single cell source and as such to meet the potential demand for tissue replacement. hESCs can also be used as *in vitro* models for certain human diseases and these models can be used, in turn, to develop drugs against the disease. However, the use of human embryos in the derivation of these cells implies an important ethical debate. Moreover, there are other issues concerning the effectiveness of these cell therapies such as teratoma formation by residual undifferentiated cells and immune rejection of foreign cells (*3*). Incomplete differentiation or differentiation to an undesired cell type may lead to a pathophysiologic state or non-functional tissue construct.

In 2006, Takahashi and Yamanaka (*4*) showed that mouse fibroblasts can be reprogrammed to an embryonic-like state by introducing defined factors, namely *Oct4, Sox2, Klf-4*, and *c-Myc (4)*, for which Prof. Yamanaka received the Nobel Prize in 2012. These cells, termed induced pluripotent stem cells (iPSCs), exhibit a similar developmental potential as their ESC counterparts and offer the advantage that embryonic material is no longer required. In addition, iPSCs allow the generation of patient-specific cells for autologous transplantation, thereby preventing immune rejection. In addition, they can be used to model human diseases and as a tool for the development of patient and disease-specific drugs (Figure 2). Later, other reports described the successful generation of hiPSCs from human somatic cells (*5, 6*) using a slightly different set of human genes (i.e. *OCT4, SOX2, NANOG, LIN28*) or by employing other types of reprogramming methods different than the initially used retroviral expression systems such as drug-inducible systems (*7*), virus-free transposon-mediated system (*8*), and recombinant proteins (*9*). MicroRNAs (miRNAs) can also be used to generate iPSCs potentially with a higher efficiency (*10*). Up to now multiple somatic cell types have been used to generate iPSCs such as fibroblasts (*5, 11, 12*), blood cells (*13*) and neural progenitors (*14*).

Since the derivation of the first hiPS cell lines, a fundamental question that has arisen but remains unanswered is whether or not hiPSCs can replace hESCs. Many research groups have attempted to address the question of the similarity of hiPSCs and hESCs at the molecular level (*5, 15*). Guenther et

al. (*16*) used microarray-based gene expression profiling and genome-wide maps of histones H3K4me3 and H3K27me3 to compare the transcriptional states of ESCs and iPSCs. This analysis revealed very few variations in the chromatin structure between reprogrammed cells and ESCs. However, these differences were not consistent across different ES and iPS cell lines and therefore could not serve to distinguish hiPSCs from hESCs (*16*). Regarding the gene expression in hiPSCs compared to hESCs, another study by Ohi et al. showed that low-passage hiPSCs seems to possess a transcriptional hallmark from their donor somatic cells that cannot be erased during the reprogramming process (*17*). Likewise, Chin et al. reported the presence of a transcriptional memory in hiPSCs by using genome-wide expression approach which revealed a recurrent gene expression signature in low-passage hiPSCs, regardless of their parental origin or the reprogramming method (*18*). However, in this study late passage hiPSCs showed a more similar gene expression profile to hESCs (*18*). Since DNA methylation plays an important role in the regulation of gene expression in ESCs (*19*), DNA methylation comparisons between hiPSCs and hESCs (*17, 20-22*) have been also carried out. In the study of Ohi et al. incomplete DNA methylation was found at the promoter regions of somatic genes in hiPSCs (*17*). Another study re-

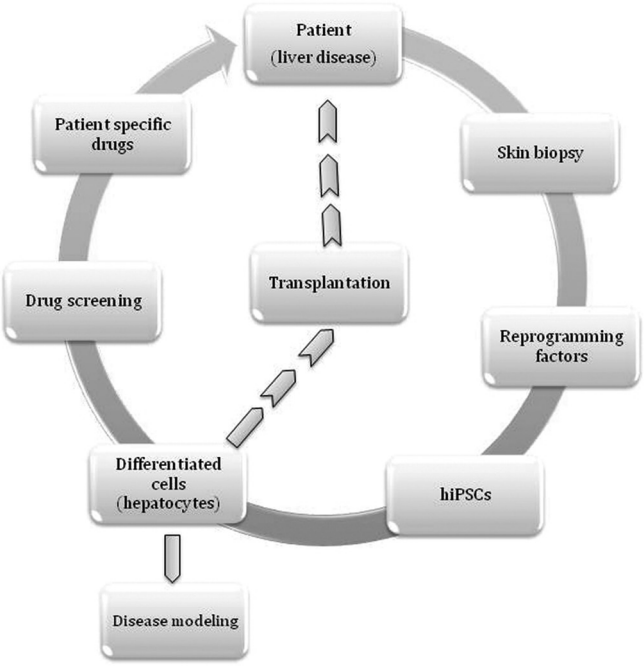


Figure 2: *The potential use of hiPSCs. Patient specific derived iPSCs can be used to generate model systems for diseased cell types. The iPS and differentiated cells can also be used for drug screening or for transplantation with regenerative purposes.*

ported that differentially methylated regions in hiPSCs were enriched in tissue-specific genes (*21*) and also a methylation signature was observed in iPSCs that is consistent with their tissue of origin (*20*). It was further suggested that the unique gene expression profile found in iPSCs can, to some extent, be explained by differences in the DNA methylation patterns between iPSCs and ESCs (*20, 21*). However, some of the differences observed between hESCs and hiPSCs might be explain by lab-specific culture conditions (*23*). Hence, published microarray data sets from different research groups were re-analyzed revealing lab-specific signatures (*23*). The similarity of hiPSCs and hESCs was further addressed by miRNA profiles. Due to the importance of miRNAs in the maintenance of pluripotency in ESCs (*24*), miRNA profiles were compared between hESCs and hiPSCs and revealed a high degree of similarity between the two. Though, a signature in the expression of a number of miRNAs was found in hiPSCs (*25*). More recently, comparative studies of hiPSCs and hESCs at the proteome and phosphoproteome level (*26, 27*) have shown that both cell lines have practically identical phosphorylation and proteome levels (*26, 27*). Despite all the above efforts, the question of how similar hiPSCs are to hESCs remains a subject of active debate.

2. **Molecular mechanisms controlling the pluripotent state**

To avoid maladaptive responses, hESCs have developed a complex transcriptional circuitry that triggers responses to several types of signals only in an appropriate biological environment. Micro-environmental factors such as growth factors and cytokines are used explicitly to control stem cell differentiation and self-renewal. However, adhesive cues and stress can modify the response of stem cells to these signals (*28*). ESCs depend on defined sets of growth factors and are regulated through multiple molecular mechanisms of cell-matrix interaction, transcriptional regulation, chromatin-modifying enzymes, regulatory RNA molecules and specific signal transduction pathways such as FGF2, actvin/nodal, and Wnt (*29*).

a. Transcriptional regulation

Several transcription factors known to play essential roles in the maintenance of hESC pluripotency also have essential roles in early development (*30*). OCT4, SOX2 and NANOG are the central transcriptional regulators

that specify ES cell identity (*30-34*). Studies have shown that OCT4, SOX2 and NANOG co-localize in the ESC chromatin together with additional transcription factors and miRNA encoding genes (*35-38*). SOX2 is known to heterodimerize with OCT4 and they together mediate the transcription activity of several genes specific for ESCs. NANOG is thought to stabilize the pluripotent state of ESCs but it appears not to be essential for the pluripotency (*39*). Together, OCT4, SOX2 and NANOG interact to regulate common targets which include both genes that promote self-renewal and differentiation (*30, 40*). Through auto-regulatory and feed-forward loops, they control their own expression and regulate that of several hundred of genes to maintain the pluripotent transcriptional network (*30*). They activate genes that promote pluripotency and repress genes that are necessary for lineage commitment and differentiation (*30*).

Epigenetics has been associated with chromatin states and the transcriptional status of genes. In general, histone acetylation and methylation of histone H3K4 correlates with active transcription (*41, 42*) whereas methylation of H3K9 and H3K27 correlates with gene repression (*41*). Trimethylation of H3K4 (H3K4me3) is found on the majority of genes in hESCs (*43*), in contrast to H3K27 trimethylation (H3K27me3) (*44*). A few promoters in ESCs carry both modifications, termed bivalent chromatin mark (*45, 46*). This chromatin mark is suggested to poise key developmental genes for lineage-specific activation or repression (*46*). To promote activation and repression of genes, OCT4, SOX2 and NANOG recruit multiple chromatin regulatory factors or complexes (*42*). Polycomb repressive complexes (PRC), for instance, bind many H3K27me3 domains at the promoter regions of target genes in ESCs, to thereby repress their transcription (*47*).

b. Growth factors and signaling pathways promoting self-renewal and pluripotency

Several growth factors have been involved in the control of the pluripotent state. Leukaemia inhibitory factor (LIF) and bone morphogenetic protein (BMP) are known to maintain mESC pluripotency. In mESCs, LIF mediated JAK-STAT signaling promotes pluripotency in the presence of serum or BMP (*48*) in addition to the PI3K/AKT pathway (*49*). However, LIF or BMP activation of JAK-STAT signaling in hESCs does not maintain them pluripotent in long term expansion, instead, BMP signaling induces the differen-

tiation of hESCs (*50, 51*). The fact that hESCs and mESCs require different growth factors may reside in the different states of pluripotency of both cell lines since hESCs appear to correspond to mouse derived epiblast stem cells (*52*). Another key growth factor is the basic fibroblast growth factor (bFGF or FGF-2) which has been widely used to promote hESC self-renewal for long-term culture (*51, 53*). Exogenous FGF-2 is capable of maintaining hESC self-renewal in the absence of feeder cells (*51, 54, 55*). Through receptor tyrosine kinases (RTKs), the FGF-2 signal triggers the activation of several intracellular signaling cascades including PI3K/AKT, Ras/ERK, and PLC-γ (*56-58*). Other factors have also been involved in self-renewal and differentiation of hESCs and include transforming growth factor-β (TGF-β) (*59*), bone morphogenetic protein (BMP) (*50*), platelet derived growth factor (PDGF) (*60*) and insulin growth factor (IGF) (*61*) most of which act through RTKs (*62*).
RTKs induce the activation of a variety of signaling proteins controlling numerous fundamental cellular processes including cell cycle, cell migration, cell metabolism and survival, as well as cell proliferation and differentiation (*62, 63*). The binding of specific ligands to these transmembrane receptors results in dimerization causing autophosphorylation of tyrosines located in the cytoplasmic domains of cells (*62*). In hESCs, the RTKs transduce extrinsic signals to their downstream targets including signaling proteins belonging to the FGF-2¸ IGF and PDGF pathways for the maintenance of their undifferentiated state (*60, 61*). Likewise, TGF-β superfamily members have been implicated in the maintenance of hESC pluripotency and differentiation (*64*) and signals through SMADs via two main branches. The TGF-β1, activin and nodal branch, signals via the activation through phosphorylation, of both TGF-β receptors serine/threonine kinases ALK4, 5 and 7 (*65*) subsequent activation of SMAD2/3 is associated with SMAD4 (*65*) which ultimately maintains pluripotency. On the other hand, the BMP branch is involved in differentiation of hESCs and signals by the activation of receptors ALK1, 2, 3, and 6. This triggers SMAD1/5, inducing their association with the common SMAD4 (*66*). Upon association, the SMAD complexes translocate to the nucleus to regulate gene expression. In addition, WNT signaling proteins are believed to be implicated in controlling hESC pluripotency and differentiation (*67*). The canonical WNT signaling involves the binding of WNT ligands to the receptor called Frizzled, leading to the activation of dishevelled family proteins, which, in turn, destabilize glycogen synthase kinase 3β (GSK-3β) and the axin/adenomatous polyposis coli complex.

This process blocks the degradation of β-catenin, causing its accumulation in the cytoplasm and, subsequently, the translocation of β-catenin into the nucleus for the activation of target genes. Despite many studies performed on hESCs, more investigation is necessary to fully understand the molecular mechanisms controlling the differentiation and self-renewal of hESCs and hiPSCs. Especially important is the link between transcription regulation and growth factor signaling, in order to facilitate the manipulation of these cells and produce an unlimited source of cells for both basic research and clinical applications.

3. **Proteomics, and its application into stem cell research**

Traditionally, the approaches used to investigate molecular mechanisms of complex biological systems are predominantly based on gene expression profiling such as microarray analysis, and low-throughput protein analyses such as western blots and immunofluorescence microscopy. However, determination of mRNA levels are insufficient to predict protein levels since gene expression regulation in cells does not directly correlate with protein levels (*68*). Differences in the gene expression levels and protein levels are associated to post-translational mechanisms controlling protein translational rates, the half-lives of proteins or mRNAs. Therefore, investigation of cells and organisms at the protein level is important to gain more insight into biological processes. Though, global protein analysis is challenging because of the complexity and dynamic range of the proteome (defined as the entire set of proteins present in a given time under defined conditions in a cell, tissue or organism). Cellular processes including protein-protein interactions, protein trafficking and localization further contribute to the complexity of the proteome. Further, the total number of proteins is estimated to be much larger than the genes that encode them, due to both splice variants and post translational modifications (PTMs) (*69, 70*). The use of mass spectrometry (MS)-based techniques has facilitated the identification and quantification of proteins and PTMs (*71-73*). Regarding the stem cell field, several groups have already applied MS-based proteomic approaches to characterize the embryonic stem cell proteome (*74, 75*) and to follow the regulatory networks of self-renewal (*61*) and differentiation (*76-79*).

The identification of proteins by intact mass is often compromised by the presence of PTMs; therefore, the analysis of intact protein mixtures by mass

spectrometry becomes complex. A modification on a protein changes the protein mass and therefore the measured mass of that protein will not match with the mass predicted from the protein sequence (*80*) resulting in incorrect identification of proteins in a database search. For proteomic analysis by mass spectrometry (MS), where protein identification plays a central role, a bottom-up approach is usually preferred. The bottom-up approach allows for sequence based identification of proteins in simple and complex biological samples. For this, cells are typically lysed to extract the proteins. Proteins are then unfolded during a denaturation process to ease their enzymatic digestion by preferably a site-specific protease. MS analysis of the resulting peptide mixture is subsequently performed. An advantage of peptide analysis by MS is that peptides fragment more efficiently than intact proteins, producing simple fragmentation spectra to interpret. Peptide identification is subsequently performed by matching the experimental fragmentation spectra against the theoretical spectra of proteins included in a certain type of database. In this section of the introduction, different techniques in the workflow for proteomic analysis (Figure 3) will be briefly described.

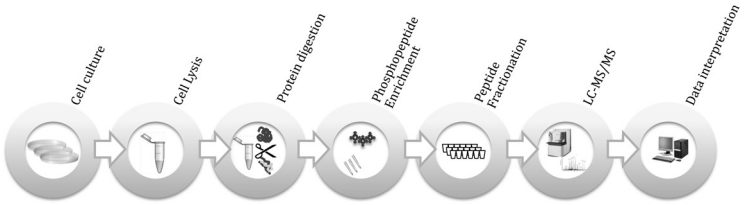


Figure 3: *A general proteomics workflow including an optional phosphopeptide enrichment step.*

a. Proteolytic digestion

Several proteolytic enzymes are available for protein digestion. The large majority of proteomic analyses employ trypsin (*81*) since the cleavage characteristics of trypsin are more suitable for the subsequent collision-induced dissociation (CID) fragmentation (*82*). CID is by far the most commonly used fragmentation technique for peptide identification. The ideal peptide for analysis by CID should have an amino acid length of 7-35 , a low charge state (z) and a high mass-to-charge ratio (*m*/*z*) (*83*). Trypsin cleaves to the C-terminal of lysine and arginine residues. The resulting tryptic peptides contain a primary amino group on the N-terminus and a basic lysine or argi-

nine on the C-terminus. Therefore, the peptides becomes multiply charged (doubly charged in the majority of cases) in an acidic buffer as well as in gas phase in the mass spectrometer. Other proteases used in proteomics such as Lys-C, Lys-N, Glu-C and Asp-N can be utilized, and are beneficial especially when electron transfer dissociation (ETD) is chosen as fragmentation method (*84-88*).

b. Peptide separation by liquid chromatography

In proteomic analyses, peptide separation is essential to reduce sample complexity. A wide range of separation techniques exist which exploit the physicochemical properties of peptides such as hydrophobicity (reverse phase chromatography), hydrophilicity (hydrophilic interaction liquid chromatography) and charge state (ion exchange chromatography).
Reversed phase (non-polar stationary phase and polar liquid mobile phase) is one of the most commonly used stationary phases in High Performance Liquid Chromatography (HPLC), to separate peptides prior to mass spectrometry analysis. The advantage of reversed phase LC is that the mobile phase is composed of acidified water and organic solvents such as acetonitrile that is compatible with electrospray ionization. When HPLC system is coupled with electrospray ionization and connected to the mass spectrometer, the peptides eluting from the LC column will then be directly infused into the mass spectrometer. The introduction of nanoelectrospray allows the formation of smaller droplet sizes permitting a higher amount of ionized peptides to enter the mass spectrometer. During nanoelectrospray, a low flow rate (nanoliter range) through a narrow spray orifice is preferred (*89*). This allows a higher efficiency in the generation of ions and an overall higher amount of ions to be transferred into the mass spectrometer. The use of nanoelectrospray resulted in the miniaturization of LC columns. Typically, fused silica capillaries of 50-100 μm of internal diameter are used. The capillaries are filled with the stationary phase consisting of silica particles of a diameter size between 1-5μm with C-18 alkyl groups attached to the surface.
For peptide separation, the peptides are in our laboratory typically loaded on a trap column o pre-column to concentrate the peptides and eliminate other molecules present in the sample. Subsequently the peptides are eluted from the trap column onto a longer (analytical) column for separation by the increase of organic solvent. The peptides then elute from the analytical column through the spray emitter directly into the mass spectrometer.

c. Ionization techniques

Ionization of peptides is required for their analysis by mass spectrometry to transform them into gas-phase ions. The two types of ionization methods that are predominantly used for biomolecules are matrix-assisted laser desorption/ionization (MALDI) (*90*) and electrospray ionization (ESI) (*91*). MALDI and ESI are termed soft ionization techniques since they can produce gas phase ions keeping the analyte intact and possibly even maintain non-covalent interactions during the ionization process at atmospheric pressure (*92*). Soft ionization techniques generally form charged peptides ions by the addition of a proton $[M+H]^+$ or by deprotonation $[M-H]^-$which are stable and generate, upon MS fragmentation, spectra that are simpler to interpret (*93*).

i. MALDI

In MALDI, peptides are first dissolved in an organic matrix solution. The mixture is then spotted onto a metal target plate and left to dry so that the peptides co-crystallize with the matrix. For the analysis of the peptides in the mass spectrometer, a pulsed laser is used to excite the crystalized sample on the plate. Subsequently, the desorption of charged ions into the gas phase occurs.

One important advantage of MALDI over ESI is that it can be used for samples in buffers containing salts and detergents. However, MALDI is mostly used for less complex peptide mixtures when separation of the peptide mixtures is unnecessary whereas ESI performs better on more complex samples, thereby making ESI, the ionization technique of choice for a complex cell lysate digest. For that reason ESI has been the predominant ionization method used for the work described in this thesis.

ii. ESI

During ESI ionization, which takes place at atmospheric pressure, the dissolved peptide mixtures flow through a thin needle at the end of the analytical chromatographic column and form small droplets by the application

of a voltage of 1-3kV to the spray capillary. The voltage provides an electric field between the capillary and the counter-electrode, separated by 0.3-2cm. This electric field is required for the electrostatic dispersion of the liquid effluent containing the peptides that are eluting from the column. In positive mode, positive ions drift towards the meniscus of the liquid and form a cone, known as the Taylor cone and the negative ions drift away. The droplets dispersed by Coulomb forces into aerosols that move toward the counter electrode. The solvent in these fine droplets evaporates, the droplets decrease in size with the charge density increasing.
The exact mechanism of how the gas phase ions are formed is not fully understood, however, two mechanisms have been proposed. The ion evaporation model (IEM) assumes that the gas phase ions are formed by exiting the larger droplets when the radius of droplets decreases below a certain size (<10 nm) due to droplet fission and solvent evaporation (*94*). The charge residue model (CRM) proposes that gas phase ions are formed due to multiple droplet fissions and solvent evaporation, until no further solvent evaporation occurs, with retention of some of the charges in the droplet (*95*).

d. Mass analyzers

The produced gas phase ions are transferred into the mass spectrometer. Once inside, they are guided by electric fields into the high vacuum of the mass spectrometer for the determination of their *m*/*z* ratio. The separation of ions according to their *m*/*z* can be achieved by different types of analyzers such as quadrupoles, time-of-flight (TOF), Fourier transformer ion cyclotron resonance (FTICR), ion traps, Orbitraps and, in addition, hybrid mass analyzers (*93*). Peptide analysis by mass spectrometry is performed in all instruments in two steps: MS mode and MS/MS mode, in which the peptides are fragmented. Both types of information are used to deduce the amino acid sequence of the peptides. The different mass analyzers used in proteomics are briefly described below (from (*93*)).

i. Quadrupole and ion trap mass analyzers

The quadrupole mass analyzer consists of four metal rods placed in parallel. A combination of direct current (DC) and radio frequency (RF) voltages is applied to adjacent rods of the quadrupole assembly whereas opposite rods

are electrically connected. The instrument allow only ions of a certain *m*/*z* to stabilize and pass through toward the detectors and ions with a different *m*/*z* which are not stable at the applied voltages to collide with the rods and become neutralized or expelled out of the quadrupole. By changing the voltages, the mass analyzer can be used as a mass filter and to trap ions (*96*). As mass analyzer, it can filter specific *m*/*z* range by ramping the voltages to produce MS spectra. The resolution of such a mass analyzer is relatively low (< 10000 (FWHM)).

As part of a more complex MS instrument, quadrupoles are used to transport/guide ions of a certain *m*/*z* range through the different components of the instrument or to trap ions (i.e. quadrupole ion traps). As an ion trap, the instrument first allows the passage of only preselected masses which can be subsequently trapped. For fragmentation, ions corresponding to a selected mass are retained in the trap while ions of different masses are rejected. The ions are trapped by the use of electric fields and subsequently transferred to a detector in order of increasing *m*/*z* ratio to record a spectrum. By the use of an inert gas, typically helium, in the ion trap at a high pressure, the ions are focused towards the center of the trap increasing the resolution, the sensitivity and the detection limit. Different types of quadrupole ion traps exist. A 3-D quadrupole ion trap has a central ring electrode and two adjacent end-cap electrodes with hyperbolic surfaces (*97*). In contrast, a linear quandrupole ion trap consists essentially of four hyperbolic rods around a central axis (*98*). The linear trap has higher injection efficiencies and higher ion storage capacities resulting in increased sensitivity an improved precision in mass assignment.

ii. Time of-flight (TOF) mass analyzers

The pulsed nature of MALDI produces ions in packets and makes it suitable for TOF mass analyzers. During TOF mass analysis, ions are pulsed into the flight tube in short and well-defined packets at an exact starting point and follow a trajectory through the mass analyzer toward the detector where a signal is produced at an end time point. Since all ions acquire the same energy in the source, light ions travel faster toward the detector than heavy ions making the mass of the ions proportional to the recorded flight time. The design of the instrument often includes an ion reflector. The ion reflector can be utilized to compensate for the initial spread of the kinetic energy of ions

with the same *m*/z. In the reflector mode, increasing voltage is applied to reflectron lenses that "reflect" the ions toward a second reflectron detector resulting in higher resolution (20,000 FWHM)(*99*).

iii. Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry

FT-ICR (*93*) is a high resolution instrument (> 750 000 FWHM). The ICR cell consists of plates arranged as a capped cylinder surrounding by a strong magnetic field (Penning trap). Ions generated in the source are directed into the ICR cell which is made of two trapping plates, two excitation plates and two detector plates. The ions are first trapped in the cell by the electric fields applied to the trapping electrodes and the magnetic field wherein the ICR cell is contained.. The ions in the ICR cell start to move in a cyclotron motion with orbital frequencies which are inversely proportional to their *m*/*z*. As an RF is applied to the excitation plates, the ions with frequencies in resonance with the applied RF are accelerated. These ions subsequently increase their orbit radius bringing them closer to the detector plates, which record the image current. The complex image is deconvoluted to the frequency domain by Fourier transformation (FT) from which the *m*/*z* can be derived.

iv. Orbitrap mass analyzer

In 2005, Alexander Makarov developed a new type of mass analyzer with a trapping mode based on the Kingdon trap. In an Orbitrap, (*100*) ions enter the trap in line with an inner axial electrode. The voltage of the axial electrode is increased as the ion enters, causing it to spiral. As the spiral moves toward the center of the axis the voltage increase stops and the spiral becomes a ring of ions. These ions cycle at different frequencies and the frequencies are converted to ion masses by Fourier transformation of the ion current. An advantage of the Orbitrap over FT-ICR is its ability to trap more ions

v. Hybrid mass analyzers

Many mass spectrometers are built by hyphenation of multiple mass analyzers to combine the strengths of the analyzers such as a triple quadrupole

(QqQ), quadrupole-TOF (Q-TOF) or the LTQ-ion trap-Orbitrap (Figure 4). The triple quadrupole is a linear series of three quadrupoles (Q1, q2, Q3) in which the first and third quadrupoles (Q1 and Q3) are typically used to filter specific *m*/*z* ions and q2 is used as a collisional cell. In MS mode, generated ions are transmitted to Q1 which select the ions within a mass range that pass through q2 and are trapped in Q3. In q2, the ion kinetic energy is set as such to minimize fragmentation of ions passing through. The ions are ejected from Q3 according to their *m*/*z* ratio to produce a mass spectrum (*101*). In MS/MS mode, Q1 is used for the selection and isolation of the precursor ion and the selected ion is fragmented in q2 followed by the trapping and mass analysis of the ions in Q3 (fragmentation in space). A Q-TOF mass spectrometer (*93*) consist of a triple quadrupole mass analyzer in which the Q3 is replaced by a TOF mass analyzer. In MS mode, the two quadrupoles are used as ion guides (using RF only) and the TOF is the only mass analyzer. In MS/MS mode, Q1 is used to transmit and select the precursor ions. The selected ions are then accelerated before they enter q2 for fragmentation. The fragment ions are subsequently analyzed in the TOF.

An LTQ-Orbitrap (*102, 103*), consists of a linear quadrupole ion trap (LTQ) and an Orbitrap, both of which can be used as mass analyzers. This allows the simultaneous analysis of ions in both parts of the instrument. While a MS spectrum is acquired in the Orbitrap with a high resolution and high mass accuracy, the fragmentation of selected ions occurs with a faster scan speed in the LTQ. In the MS mode, ions are pre-scanned in the LTQ to determine the optimum amount of ions to trap (automatic gain control). The ions are transferred via an octapole ion guide into a quadrupole trap named the C-trap. The ions trapped in the C-trap move toward the centre of the C-trap by ramping the voltages on the electrodes. Subsequently, the ions are accelerated to a higher kinetic energy and injected into the Orbitrap. The C-trap is located to the exit of the Orbitrap to allow a fast injection of the ions into the Orbitrap. On their way from the C-trap, ions pass through steps of differential pumping until they reach the high vacuum compartment of the Orbitrap. The resolution of the instrument is 60 000 at *m*/*z* 400 when acquiring the spectrum in 1 second, and increases linearly with the detection time to a maximum of 100 000. The mass accuracy of the instrument is 5 ppm using external calibration and 2 ppm with internal calibration. In the C-trap, the "lock mass" which can be a background ion (from the ambient or solvent), can be employed for real time corrections of mass shifts, improving

the mass accuracy. Another version of an LTQ-Orbitrap (LTQ-Orbitrap velos) which is illustrated in Figure 4 has an additional collision cell for high energy C-trap dissociation (HCD) and a dual linear ion trap. The dual ion trap enables a more efficient trapping of ions and a faster scanning and detection (*104*) by using the first linear ion trap for trapping and fragmentation of ions with the second ion trap scanning the fragments to generate the MS/MS spectra. The LTQ-Orbitrap is the mass spectrometer used for the mass spectrometry analysis in this thesis.

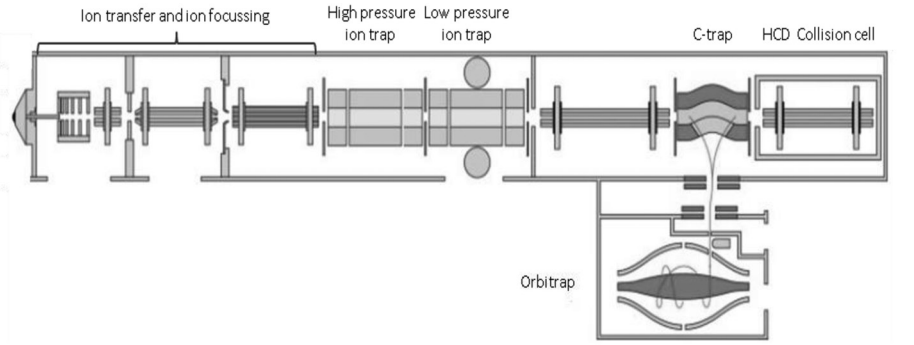


Figure 4: *Schematic representation of the LTQ Orbitrap Velos mass spectrometer adapted from Olsen et al (104). In this instrument, the quadrupoles and multipoles guide ions to the linear ion traps in which the high pressure ion trap is used to trap and fragment the selected precursor ions and the fragments are subsequently scan for detection in the low pression ion trap. The C-trap is used to trap and inject ions into the Orbitrap for high resolution MS with a multipole collision cell next to the C-trap that is used for HCD fragmentation.*

e. Fragmentation techniques

Principally three types of fragmentation techniques are used for peptide fragmentation in the mass spectrometer: collision induced dissociation (CID) (*92*), electron transfer dissociation (ETD) (*105*) and high energy C-trap dissociation (HCD) (*106*). Prior to fragmentation, a precursor ion is first selected, then isolated and subsequently fragmented to its constituting product ions. Different types of ions specific to a fragmentation method are obtained. These ions are annotated following a nomenclature described by Roepstorff and Fohlman (*107*) and further modified by Biemann *et al.* (*108*) (Figure 5). Cleavage of the amide back bone leads to two types of fragmentations depending on the localization of the charges. If the charge is retained on the N-terminus of the ion, a, b, and c-type ions are generated. C-terminal fragments termed x, y, z ions are generated when the charge is located at the

C-terminus. Thus, fragmentation of multiply charged ions generates a mixture of complementary ion pairs. CID involves collisions of excited peptide ions with neutral atoms or molecules thereby inducing the cleavage of peptide amide bonds (CO-NH) generating predominately b and y ions (Figure 5). CID is usually performed in linear quadrupole ion traps using nitrogen, helium or argon as collisional gas. With CID, the energy transferred to an ion is limited. For large ions, the energy is distributed on a greater number of amide bonds which leads to a slow reaction rate of the fragmentation and thus the degree of fragmentation is limited. On the other hand, HCD collision occurs with a higher energy level, resulting in extensive fragmentation. In Orbitrap mass spectrometers, HCD can be performed in a collision cell close to the C-trap and the ions can be transmitted to the Orbitrap through the C-trap for detection, resulting in high resolution and high mass accuracy of the fragment ions. In the third fragmentation type, ETD fragmentation occurs in an ETD cell (usually a linear quadrupole ion trap) by the transfer of electrons from singly charged aromatic anions to multiply protonated peptide ions causing the fragmentation of the peptide backbone to generate c and z ions by cleavage of the NH-CH amide peptide bond (Figure 5). Similar types of fragment ions are generated by the use of electron capture dissociation (ECD) in a FT-ICR mass spectrometer. During CID, cleavage of the weakest bonds occurs which is the peptide amide bond that has the lowest energy barrier to fragment, however PTMs that are often present on peptides/proteins have lower energy barrier than amide bonds and are therefore often cleaved during CID fragmentation. On the other hand, ECD and ETD induce fragmentation along the peptide backbone, dependent on charge of the ion thereby preserving labile modifications on the peptides which is an advantage for the identification of PTMs on peptides.

CID is most frequently used as fragmentation method in proteomic analysis. However, alternative fragmentation methods such as HCD and ETD are known to improve the identification of long and highly charged peptides, including peptides with high number of basic residues and peptides containing PTMs (*82, 109-111*). Moreover, the choice of the fragmentation type often depends on the protease used for enzymatic digestion. Trypsin generates peptides with sizes suitable for CID and HCD fragmentation. However, Lys-C or Lys-N generates larger peptides suitable for ETD fragmentation (*87, 88*). Further, it has been shown that alternating fragmentation of the precursor ion can sustainably increase the identification of the matching

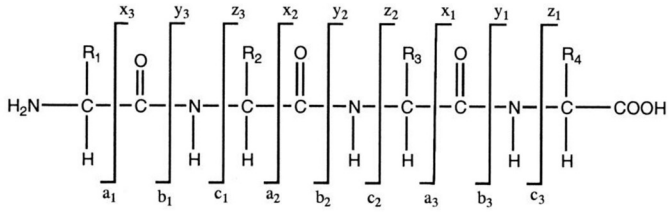


Figure 5: *Fragmentation nomenclature of peptide ions as proposed by Roepstorff and Fohlman (107). Three types of fragmentation can be observed in MS/MS spectra: N-terminal ion fragments are named with a,b,c and the C-terminal fragments are denoted x,y,z, depending on the cleavage site.*

peptide (*82*, *112*).

f. Detectors

Several types of detectors exist and the choice depends on the design and application of the instrument. The most widely used ion detector in mass spectrometry is the electron multiplier. In this detector, ions from the mass analyzer are accelerated to a high velocity in order to enhance the detection efficiency. In principal, the ions strike the detector surface which releases electrons from the surface. The electrons in turn strike the surface emitting more electrons. That procedure is repeated so that the final stream of electrons strikes the cathode (*93*).

g. Database search and data processing

Various search engines are available for protein identification from an MS/MS spectrum which all use different scoring algorithms such as Mascot (*113*) and Sequest (*114*). These software packages compare the fragmentation spectra acquired by mass spectrometry against the theoretical fragmentation, generated *in silico*, from peptides in the database. Other experimental parameters such as the type of protease, and the mass tolerance are necessary for an efficient search. Furthermore, peptide modifications (e.g. chemical modifications, PTMs) can also be included. A score value is assigned to each peptide, which is a measure of how well the experimental spectrum is matched to the theoretical spectrum. Thus, the peptide score is dependent on the quality of the fragmentation spectrum. A low quality fragmentation spectrum with few fragmentation ions could provide false positive identifi-

cations. A way to estimate false positives is to perform the database search using a reversed or decoy database containing reversed or random (fake) sequences and forward (real) peptide sequences to calculate a false discovery rate (FDR) (*115*, *116*). With this method, estimation is made on the relative amount of false positives in the whole identification dataset. For further validation of peptide identifications from database searches, tools such as Percolator are used. The Percolator algorithm (*117*) uses a number of parameters such as mass deviation, ion intensity, retention time, number of residues and charge. These parameters are related to the quality of the match of the target and decoy peptide spectral match (PSM). Subsequently, the set of target and decoy PSMs are distinguished by the most relevant parameter and filtered to a fixed FDR. This training is then applied to all target and decoy PSMs in the data (*117*).

h. Quantification approaches

Mass spectrometry is not inherently quantitative, mainly due to the different physicochemical properties that proteolytic peptides possess such as charge, size and hydrophobicity that influence their ionization and their detection by mass spectrometry. Many different approaches have been described to quantify peptides and therefore proteins in biological systems (e.g. cells, tissues, organisms). Depending on the information desired, two types of quantitative MS strategies can be applied. Relative quantitation compares the amount of proteins between samples reporting the quantitative ratio or relative changes. Absolute quantitation provides information about the absolute amount or concentration of a protein in a sample.

i. Stable isotope labeling

Stable isotope labeling methods can be applied to obtain relative quantitative information on the abundance of proteins or the level of site-specific PTMs between samples. With this method, samples can be simultaneously analyzed minimizing variations arising from sample preparations and LC-MS analyses. Labeling methods can be applied to samples originating from cells, tissues or whole organisms. Isotopes ^{2}H, ^{13}C, ^{15}N and ^{18}O are commonly used for labeling and compared with their naturally, more abundant light forms ^{1}H, ^{12}C, ^{14}N, and ^{16}O. Stable isotope labeling approaches are based on

the fact that a peptide labeled with heavy stable isotopes differs in mass from the unlabeled or light labeled peptides while having the same chemical properties during LC/MS analysis. However, a shift in a LC retention time can be observed with deuterated peptides which typically elute earlier than the non-labeled counterparts (if reversed phase LC is used). The relative abundance is determined by integrating the extracted ion chromatograms of the labeled and unlabeled ions at the MS level. For quantitation at the MS/MS level, the relative abundance of peptides is determined from the intensities of the so-called reporter ions. The labels can be introduced at various steps in the workflow (Figure 1).

1. *Chemical labeling*

During chemical labeling, labels are added at the peptide level by a chemical reaction. There are mainly two groups of chemical labeling approaches which are: labeling by isotopic tags and labeling by isobaric tags. Labeling by isotopic tags (e.g. dimethyl, $^{18}O/^{16}O$, ICAT, ICPL) is used for quantification at the MS level where mass differences are observed between the labels. Chemical labeling with isobaric tags are used for quantification at the MS/MS level (TMT (*118*), iTRAQ (*119*)).

Isobaric tags are composed of mass reporters, linkers and reactive groups. The mass reporters have ^{15}N and ^{13}C substitutions that give them variable masses. The differences in mass of the reporters are normalized by linkers which therefore vary in mass. The labeling is achieved by the reaction of peptides primary amines or cysteine residues cross-linked with the reactive group. The reporter dissociates from the peptide upon fragmentation and the relative intensities of the reporter ions are used for peptide quantification.

In this thesis, dimethyl labeling (isotopic tag) has been primary used as a way to perform relative quantification between different samples. With dimethyl labeling, the primary amines on the N-terminus of peptides and the ε-amino group of lysine residues are labeled with dimethyl groups using isotopomers of formaldehyde and cyanoborohydride allowing for double or triple labeling (*120*). The resulting mass difference between labels is 4 Da. ^{18}O labeling can be achieved by performing the digestion of one of the samples in heavy water ($H_2{}^{18}O$) (*121*). During the digestion, labeling occurs by the replacement of ^{16}O with ^{18}O atoms at the C-terminus of the peptides. In

the isotope-coded affinity tag (ICAT) labeling (*122*) method, cysteine residues react with an iodacetamide moiety that is attached to an isotopically coded linker with a biotin molecule. Only cysteine containing peptides will be labeled and therefore quantified. A similar labeling method is the isotope-coded protein labeling (ICPL) that is used for the labeling of N-termini of proteins and lysine residues instead of cysteine residues thereby increasing the number of labeled peptides in the sample (*123*).

2. *Metabolic labeling*

In metabolic labeling, stable isotopes are introduced using the protein synthesis machinery of cells. A popular example of such method is the stable isotope labeling by amino acids in cell culture (SILAC) (*124*). In SILAC, cells are grown in media containing either the naturally existing essential amino acid or the isotopically (heavy) labeled form of an amino acid. After several passages, the heavy amino acids are incorporated in the newly synthesized proteins. Preferentially, essential amino acids such as methionine or lysine are used for this purpose to ensure that the medium is the only source of that particular amino acid. Yet, non-essential amino acids are also used such as arginine. A disadvantage of using heavy arginine labeling is that a conversion of arginine to proline is often observed (*124*). This method cannot be applied for the efficient labeling of autotrophic cells that are able to synthesize all amino acids by themselves using inorganic compounds. In those cases, ^{15}N labeling represents an excellent alternative, especially for complex multicellular model organisms such as *Caenorhabditis elegans* and *Drososphila melanogaster* (*125*). Using this method, one set of cells is grown using normal medium containing the natural abundance of the isotopes of nitrogen (^{14}N at 99.6% and ^{15}N at 0.4%) and the second set of cells is grown in a medium enriched in ^{15}N. For labeling of multicellular organisms, these organisms need to be grown with ^{15}N labeled nutrient. Depending on the organisms, one or more generations are required for complete labeling. However, the uncertainty of complete labeling remains, since different tissues in the organism have different protein turnover rates and therefore, need to be determined.

ii. Label free

In label free quantification, relative protein abundance is determined based

on the corresponding peptides intensities or spectral counting (*126*, *127*). Based on mass accuracy and retention time, peptides are quantified by their intensities across individual LC/MS runs of samples to be compared. With the spectral counting method, the number of spectra matched to peptides of a particular protein is used as a substitute to measure protein abundance. The main disadvantages of label free quantification are the low accuracy due to variability introduced during sample preparation, and variation in retention time across LC-MS runs for MS1 intensity-based quantitation.

i. Enrichment methods for phosphorylated peptides

PTMs of proteins in eukaryote organisms determine their tertiary and quaternary structures and play an important role in the modulation of protein function. The combination of separation and fractionation techniques, as well as enrichment methods, with mass spectrometry assists in the characterization of peptides with PTMs. Phosphorylation is the most abundant and ubiquitous PTM. Reversible protein phosphorylation is an important mechanism involved in many cellular processes such as protein regulation and signal transduction. Protein phosphorylation mostly occur on serine, threonine and, to a lesser extent, on tyrosine residues. The process is catalyzed by kinases whereas dephosphorylation is carried out by phosphatases. Around 500 genes code for protein kinases and 150 phosphatase genes exist in the human genome (*128*). Of these, two thirds are serine/threonine kinases and less than 40 are serine/threonine phosphatases (*129*). Importantly, reversible protein phosphorylation has attracted attention due to its involvement in many diseases such as cancer, diabetes and Alzheimer diseases. Mass spectrometry has been used to identify and quantify phosphorylated peptides in a large scale fashion. However, the analysis of phosphorylated peptides poses some challenges since phosphorylation is a labile PTM. During CID fragmentation, phosphorylated residues can lose their phosphate group in the form of H_3PO_4 (from serine and threonine) and HPO_3 (from tyrosine) (*130*). In addition, tyrosine phosphorylated peptides generate an immonium ion with an m/z ratio of 216.043 during CID fragmentation (*131*). Furthermore, due to their low stoichiometry, the mass spectrometric detection of phosphopeptides presents additional difficulties. Therefore, pre-enrichment of these low abundant peptides is necessary. The enrichment methods used nowadays are based on immunoaffinity purification

and electrostatic properties of phosphate groups.

i. Metal-ion based affinity chromatography

Metal-based affinity chromatography exploits the strong affinity of the negatively charged phosphate groups to positively charged metal ions. One of the most commonly used method is immobilized metal affinity chromatography (IMAC) (*132*) which is composed of an immobilized chelating group, for instance, nitrilotriacetic acid (NTA), bound to a metal group as stationary phase. The metal ions bind to phosphate groups of peptides, thereby separating the phosphorylated peptides from the non-phosphorylated species. Ga^{3+} (*133*), Fe^{3+} (*134*), and Zr^{3+} (*135*) have been successfully utilized for the enrichment of phosphopeptides. Another popular method is metal oxide affinity chromatography (MOAC) that exploits the phosphate and metal oxide interaction. Titanium dioxide (TiO_2) microspheres (*136, 137*) are used for the enrichment of phosphopeptides, which is proven to have higher selectivity for phosphopeptides than IMAC (*138*). However, the selectivity of these methods is often not optimal, since non-phosphorylated peptides containing a high number of acidic residues (aspartic acid and glutamic acid) are usually co-enriched. To reduce the non-specific binding of these peptides, organic acids such as dihydrobenzoic acid (DHB) are used to decrease the non-specific binding by competing with the non-phosphorylated peptides (*139*). A third approach used for the enrichment of phosphorylated peptides is Ti^{4+}-IMAC which is more efficient in phosphopeptide enrichment than the above described methods (*140*). Ti^{4+}-IMAC consists of a immobilized Ti^{4+} on a phosphate polymer (*140*). This method allows for the enrichment of phosphopeptides with high number of basic residues (*140*). In the Ti^{4+}-IMAC enrichment method, the use of higher concentration of TFA in the reaction solution reduces the interaction between basic residues and phosphate group within peptides. This is because TFA is known to interact with basic residues, thereby allowing the pool of positive phosphate group to bind to the Ti^{4+} (*141*).

ii. Immuno-affinity purification of Tyrosine phosphorylated peptides

The methods described above are mostly suitable for the detection of ser-

ine and threonine phosphorylated peptides by MS analysis while tyrosine phosphorylated peptides are barely detected using these methods. The low sub-stochiometry of tyrosine phosphorylated proteins *in vivo* is not an indication of a less important function since abnormalities in tyrosine phosphorylation are directly linked to numerous cancers and other diseases (*142*). An effective approach to specifically enrich tyrosine phosphorylated peptides is the use of immuno-affinity chromatography. Here, specific antibodies generated against phosphorylated tyrosine are used for enrichment. The antibodies are usually cross-linked to a non-metal resin (agarose or sepharose). Whole cell digest dissolved in a buffer at physiological pH is applied to the antibody column for the binding of phosphotyrosine peptides. The release of the enriched peptides occurs by disrupting the antigen-antibody interaction with a suitable elution buffer. The detection of tyrosine phosphorylated peptides by mass spectrometry significantly increases using immuno-affinity purification in comparison with other phosphopeptide enrichment methods (*143*).

iii. Targeting phosphopeptides with various separation methods

Cation and anion exchange chromatography are based on the separation of peptides by their charge state in solution. In cation exchange chromatography, e.g. strong cation exchange (SCX) the stationary phase is negative and therefore interacts with ions of positive charge and the opposite principle is applied in anion exchange chromatography, e.g. strong anion exchange (SAX) and weak anion exchange (WAX). The bound peptides are eluted by increasing the ionic strength with salt containing buffers to displace the peptides from the stationary phase. In proteomics, SCX is often used as the first dimensional pre-fractionation method prior to the reversed phase separation in a second dimension and MS analysis. The use of SCX as a phosphopeptide enrichment method is based on the fact that, at pH 2.7, most tryptic peptides (without missed-cleavages) carry a positive net charge state of +2 in solution. On the other hand, most phosphorylated peptides carry a positive net charge state of one (without missed-cleavages), since the phosphate group maintains a negative charge at that pH. Therefore, low pH SCX separation of tryptic peptides can be used to separate phosphorylated peptides from the bulk of non-phosphorylated peptides (*144*).
Similarly, hydrophilic interaction liquid chromatography (HILIC), based on

a hydrophilic stationary phase and an organic mobile phase, can be used for first dimension of separation in proteomics. The separation power of HILIC is shown to be higher than that of SCX where usually the peptides of the same charge states elute together in one cluster (*145*). With HILIC separation, peptides elute more evenly during the chromatography and thereby reduce the sample complexity in a more efficient manner than the SCX. Therefore, HILIC can also be used as a fractionation method for samples enriched for phosphorylated peptides. Using HILIC, the separation of the phosphopeptides will be equally spread over multiple fractions allowing the detection of higher number of phosphorylated peptides than in the case of the SCX separation.

4. **Scope and outline of the thesis**

The focus of this thesis is on extending our knowledge on stem cell biology by using advanced phosphoproteomics technologies. Reversible protein phosphorylation has been one of the most investigated post-translational modifications in the decade, made possible by advances in enrichment technologies and MS-based proteomics. PTMs are generally involved in the regulation of many biological processes, including processes that regulate pluripotency and self-renewal of stem cells. In recent years, several enrichment methods have been introduced that helped to improve the detection and quantitation of phosphorylated peptides by MS enabling the study of phosphorylation dynamics in cells and tissues. The work described in this thesis is focused on the application of a selection of such proteomic strategies, for the investigation of phosphorylation events in pluripotent stem cells. A global phosphoprotemic analysis was performed to investigate the role of the ligand FGF-2 on hESCs focusing on its contribution in the maintenance of hESCs pluripotency (chapter 2). A number of phosphorylated proteins were found to be differentially phosphorylated upon FGF-2 stimulation. The result of this study revealed that there is a direct connection between FGF-2 signaling and the transcriptional regulatory circuitry of hESCs suggesting that FGF-2 may directly regulate the OCT4/SOX2/NANOG trimeric complex.

Although FGF-2 is a ligand for receptor tyrosine kinases, tyrosine phosphorylated peptides belonging to proteins which are known to play a crucial role in FGF-2 signaling were barely detected in the study reported in

Chapter 2, since the enrichment methods used, principally enrich for all serine, threonine, and tyrosine phosphorylated peptides. As the latter are less abundant, they are easily missed in these global analyses. In chapter 3, we therefore explored and set-up a protocol for a more targeted strategy by the use of specific antibodies raised against phosphotyrosine peptides to enrich through immunoprecipitation (IP) phosphotyrosine peptides prior to their detection by MS. In chapter 4, this method is applied in combination with quantitative mass spectrometry-based proteomics in a study where tyrosine phosphorylation levels of hESCs were compared to hiPSCs. A previous study from our laboratory revealed that globally the proteomes of hiPSCs and hESCs are nearly indistinguishable. However, speculations have been raised that differences between these cells may be apparent at the protein PTM level. Employing the antibody based tyrosine phosphorylated peptide enrichment; we detected several tyrosine phosphorylated proteins with differential levels of phosphorylation in two hiPS cell lines when compared to their hESC counterparts. Proteins we identified as having a different phosphorylation level in hiPSCs, included SRC family kinases, which play a role in the maintenance of hESCs pluripotency. Interestingly, our antibody based phosphotyrosine peptide dataset showed a very small overlap (<3%) when compared to a recent global phosphoproteomic comparison of hiPSCs and hESCs (*27*), indicating the complementarity of both approaches. In chapter 5, we explored the complementarity of these two methods further and made a comparison between tyrosine phosphorylated peptides enriched and identified using two different enrichment strategies, namely a) Ti^{4+}IMAC enrichment followed by fractionation using HILIC and b) antibody based enrichment. The result of this analysis suggested that antibodies used for phosphotyrosine peptide enrichment have a bias toward hydrophobic peptides and singly phosphorylated peptides. On the other hand, Ti^{4+}IMAC mainly enriched for phosphotyrosine peptides containing a high number of acidic residues and multiply phosphorylated peptides. Complementary methods might thus be necessary to further increase the detection of tyrosine phosphorylated peptides by mass spectrometry.
I conclude the thesis with a future outlook discussing the need for advancing further, enabling tools for quantitative phosphoproteomic analyses, and how they could be employed to further investigate stem cell biology.

5. References

[1] G. R. Martin, Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells, Proc Natl Acad Sci U S A 78 (1981) 7634-7638.

[2] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, J. M. Jones, Embryonic stem cell lines derived from human blastocysts, Science 282 (1998) 1145-1147.

[3] R. Passier, L. W. van Laake, C. L. Mummery, Stem-cell-based therapy and lessons from the heart, Nature 453 (2008) 322-329.

[4] K. Takahashi, S. Yamanaka, Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors, Cell 126 (2006) 663-676.

[5] J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, Slukvin, II, J. A. Thomson, Induced pluripotent stem cell lines derived from human somatic cells, Science 318 (2007) 1917-1920.

[6] J. Yu, K. Hu, K. Smuga-Otto, S. Tian, R. Stewart, Slukvin, II, J. A. Thomson, Human induced pluripotent stem cells free of vector and transgene sequences, Science 324 (2009) 797-801.

[7] D. Hockemeyer, F. Soldner, E. G. Cook, Q. Gao, M. Mitalipova, R. Jaenisch, A drug-inducible system for direct reprogramming of human somatic cells to pluripotency, Cell Stem Cell 3 (2008) 346-353.

[8] K. Woltjen, I. P. Michael, P. Mohseni, R. Desai, M. Mileikovsky, R. Hamalainen, R. Cowling, W. Wang, P. Liu, M. Gertsenstein, K. Kaji, H. K. Sung, A. Nagy, piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells, Nature 458 (2009) 766-770.

[9] D. Kim, C. H. Kim, J. I. Moon, Y. G. Chung, M. Y. Chang, B. S. Han, S. Ko, E. Yang, K. Y. Cha, R. Lanza, K. S. Kim, Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins, Cell Stem Cell 4 (2009) 472-476.

[10] N. Miyoshi, H. Ishii, H. Nagano, N. Haraguchi, D. L. Dewi, Y. Kano, S. Nishikawa, M. Tanemura, K. Mimori, F. Tanaka, T. Saito, J. Nishimura, I. Takemasa, T. Mizushima, M. Ikeda, H. Yamamoto, M. Sekimoto, Y. Doki, M. Mori, Reprogramming of mouse and human cells to pluripotency using mature microRNAs, Cell Stem Cell 8 (2011) 633-638.

[11] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, Induction of pluripotent stem cells from adult human fibroblasts by defined factors, Cell 131 (2007) 861-872.

[12] I. H. Park, R. Zhao, J. A. West, A. Yabuuchi, H. Huo, T. A. Ince, P. H. Lerou, M. W. Lensch, G. Q. Daley, Reprogramming of human somatic cells to pluripotency with defined factors, Nature 451 (2008) 141-146.

[13] Y. H. Loh, S. Agarwal, I. H. Park, A. Urbach, H. Huo, G. C. Heffner, K. Kim, J. D. Miller, K. Ng, G. Q. Daley, Generation of induced pluripotent stem cells from human blood, Blood 113 (2009) 5476-5479.

[14] S. Eminli, J. Utikal, K. Arnold, R. Jaenisch, K. Hochedlinger, Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression, Stem Cells 26 (2008) 2467-2474.

[15] K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, Cell 126 (2006) 663-676.

[16] M. G. Guenther, G. M. Frampton, F. Soldner, D. Hockemeyer, M. Mitalipova, R. Jaenisch, R. A. Young, Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells, Cell Stem Cell 7 (2010) 249-257.

[17] Y. Ohi, H. Qin, C. Hong, L. Blouin, J. M. Polo, T. Guo, Z. Qi, S. L. Downey, P. D. Manos, D. J. Rossi, J. Yu, M. Hebrok, K. Hochedlinger, J. F. Costello, J. S. Song, M. Ramalho-Santos, Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells, Nat Cell Biol 13 (2011) 541-549.

[18] M. H. Chin, M. J. Mason, W. Xie, S. Volinia, M. Singer, C. Peterson, G. Ambartsumyan, O. Aimiuwu, L. Richter, J. Zhang, I. Khvorostov, V. Ott, M. Grunstein, N. Lavon, N. Benvenisty, C. M. Croce, A. T. Clark, T. Baxter, A. D. Pyle, M. A. Teitell, M. Pelegrini, K. Plath, W. E. Lowry, Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures, Cell Stem Cell 5 (2009) 111-123.

[19] G. Altun, J. F. Loring, L. C. Laurent, DNA methylation in embryonic stem cells, J Cell Biochem 109 (2010) 1-6.

[20] K. Kim, A. Doi, B. Wen, K. Ng, R. Zhao, P. Cahan, J. Kim, M. J. Aryee, H. Ji, L. I. Ehrlich, A. Yabuuchi, A. Takeuchi, K. C. Cunniff, H. Hong-

guang, S. McKinney-Freeman, O. Naveiras, T. J. Yoon, R. A. Irizarry, N. Jung, J. Seita, J. Hanna, P. Murakami, R. Jaenisch, R. Weissleder, S. H. Orkin, I. L. Weissman, A. P. Feinberg, G. Q. Daley, Epigenetic memory in induced pluripotent stem cells, Nature 467 (2010) 285-290.

[21] A. Doi, I. H. Park, B. Wen, P. Murakami, M. J. Aryee, R. Irizarry, B. Herb, C. Ladd-Acosta, J. Rho, S. Loewer, J. Miller, T. Schlaeger, G. Q. Daley, A. P. Feinberg, Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts, Nat Genet 41 (2009) 1350-1353.

[22] K. Kim, R. Zhao, A. Doi, K. Ng, J. Unternaehrer, P. Cahan, H. Huo, Y. H. Loh, M. J. Aryee, M. W. Lensch, H. Li, J. J. Collins, A. P. Feinberg, G. Q. Daley, Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells, Nature biotechnology 29 (2011) 1117-1119.

[23] A. M. Newman, J. B. Cooper, Lab-specific gene expression signatures in pluripotent stem cells, Cell Stem Cell 7 (2010) 258-262.

[24] Y. H. Loh, L. Yang, J. C. Yang, H. Li, J. J. Collins, G. Q. Daley, Genomic approaches to deconstruct pluripotency, Annu Rev Genomics Hum Genet 12 (2011) 165-185.

[25] K. D. Wilson, S. Venkatasubrahmanyam, F. Jia, N. Sun, A. J. Butte, J. C. Wu, MicroRNA profiling of human-induced pluripotent stem cells, Stem cells and development 18 (2009) 749-758.

[26] J. Munoz, T. Y. Low, Y. J. Kok, A. Chin, C. K. Frese, V. Ding, A. Choo, A. J. Heck, The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells, Mol Syst Biol 7 (2011) 550.

[27] D. H. Phanstiel, J. Brumbaugh, C. D. Wenger, S. Tian, M. D. Probasco, D. J. Bailey, D. L. Swaney, M. A. Tervo, J. M. Bolin, V. Ruotti, R. Stewart, J. A. Thomson, J. J. Coon, Proteomic and phosphoproteomic comparison of human ES and iPS cells, Nature methods 8 (2011) 821-827.

[28] D. M. Cohen, C. S. Chen, in StemBook. (Cambridge (MA), 2008).

[29] R. Jaenisch, R. Young, Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming, Cell 132 (2008) 567-582.

[30] L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner,

D. K. Gifford, D. A. Melton, R. Jaenisch, R. A. Young, Core transcriptional regulatory circuitry in human embryonic stem cells, Cell 122 (2005) 947-956.

[31] A. A. Avilion, S. K. Nicolis, L. H. Pevny, L. Perez, N. Vivian, R. Lovell-Badge, Multipotent cell lineages in early mouse development depend on SOX2 function, Genes Dev 17 (2003) 126-140.

[32] I. Chambers, D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, A. Smith, Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells, Cell 113 (2003) 643-655.

[33] A. H. Hart, L. Hartley, M. Ibrahim, L. Robb, Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human, Developmental Dynamics 230 (2004) 187-198.

[34] J. Nichols, B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Schöler, A. Smith, Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4, Cell 95 (1998) 379-391.

[35] I. Chambers, S. R. Tomlinson, The transcriptional foundation of pluripotency, Development 136 (2009) 2311-2322.

[36] X. Chen, H. Xu, P. Yuan, F. Fang, M. Huss, V. B. Vega, E. Wong, Y. L. Orlov, W. Zhang, J. Jiang, Y.-H. Loh, H. C. Yeo, Z. X. Yeo, V. Narang, K. R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W.-K. Sung, N. D. Clarke, C.-L. Wei, H.-H. Ng, Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells, Cell 133 (2008) 1106-1117.

[37] M. F. Cole, S. E. Johnstone, J. J. Newman, M. H. Kagey, R. A. Young, Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells, Genes Dev 22 (2008) 746-755.

[38] J. Kim, J. Chu, X. Shen, J. Wang, S. H. Orkin, An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells, Cell 132 (2008) 1049-1061.

[39] I. Chambers, J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, A. Smith, Nanog safeguards pluripotency and mediates germline development, Nature 450 (2007) 1230-1234.

[40] Y.-H. Loh, Q. Wu, J.-L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K.-Y. Wong, K. W. Sung, C. W.

H. Lee, X.-D. Zhao, K.-P. Chiu, L. Lipovich, V. A. Kuznetsov, P. Robson, L. W. Stanton, C.-L. Wei, Y. Ruan, B. Lim, H.-H. Ng, The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells, Nat Genet 38 (2006) 431-440.

[41] B. Li, M. Carey, J. L. Workman, The Role of Chromatin during Transcription, Cell 128 (2007) 707-719.

[42] Stuart H. Orkin, K. Hochedlinger, Chromatin Connections to Pluripotency and Cellular Reprogramming, Cell 145 (2011) 835-850.

[43] M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, R. A. Young, A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells, Cell 130 (2007) 77-88.

[44] B. L. Sha K, The chromatin signature of pluripotent cells, In: StemBook [Internet]. (2009 May 31).

[45] V. Azuara, P. Perry, S. Sauer, M. Spivakov, H. F. Jorgensen, R. M. John, M. Gouti, M. Casanova, G. Warnes, M. Merkenschlager, A. G. Fisher, Chromatin signatures of pluripotent cell lines, Nat Cell Biol 8 (2006) 532-538.

[46] B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, E. S. Lander, A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells, Cell 125 (2006) 315-326.

[47] T. I. Lee, R. G. Jenner, L. A. Boyer, M. G. Guenther, S. S. Levine, R. M. Kumar, B. Chevalier, S. E. Johnstone, M. F. Cole, K.-i. Isono, H. Koseki, T. Fuchikami, K. Abe, H. L. Murray, J. P. Zucker, B. Yuan, G. W. Bell, E. Herbolsheimer, N. M. Hannett, K. Sun, D. T. Odom, A. P. Otte, T. L. Volkert, D. P. Bartel, D. A. Melton, D. K. Gifford, R. Jaenisch, R. A. Young, Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells, Cell 125 (2006) 301-313.

[48] Q.-L. Ying, J. Nichols, I. Chambers, A. Smith, BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3, Cell 115 (2003) 281-292.

[49] M. P. Storm, H. K. Bone, C. G. Beck, P.-Y. Bourillot, V. Schreiber, T. Damiano, A. Nelson, P. Savatier, M. J. Welham, Regulation of Nanog Expression by Phosphoinositide 3-Kinase-dependent Signaling in Murine Embryonic Stem Cells, Journal of Biological Chemistry 282

(2007) 6265-6273.
[50] R. H. Xu, X. Chen, D. S. Li, R. Li, G. C. Addicks, C. Glennon, T. P. Zwaka, J. A. Thomson, BMP4 initiates human embryonic stem cell differentiation to trophoblast, Nat Biotechnol 20 (2002) 1261-1264.
[51] R. H. Xu, R. M. Peck, D. S. Li, X. Feng, T. Ludwig, J. A. Thomson, Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells, Nat Methods 2 (2005) 185-190.
[52] K. Greenow, A. R. Clarke, Controlling the Stem Cell Compartment and Regeneration In Vivo: The Role of Pluripotency Pathways, Physiological Reviews 92 (2012) 75-99.
[53] S. Avery, K. Inniss, H. Moore, The regulation of self-renewal in human embryonic stem cells, Stem Cells Dev 15 (2006) 729-740.
[54] V. M. Ding, L. Ling, S. Natarajan, M. G. Yap, S. M. Cool, A. B. Choo, FGF-2 modulates Wnt signaling in undifferentiated hESC and iPS cells through activated PI3-K/GSK3beta signaling, J Cell Physiol 225 417-428.
[55] M. E. Levenstein, T. E. Ludwig, R. H. Xu, R. A. Llanas, K. VanDenHeuvel-Kramer, D. Manning, J. A. Thomson, Basic fibroblast growth factor support of human embryonic stem cell self-renewal, Stem Cells 24 (2006) 568-574.
[56] P. Dvorak, A. Hampl, Basic fibroblast growth factor and its receptors in human embryonic stem cells, Folia Histochem Cytobiol 43 (2005) 203-208.
[57] V. P. Eswarakumar, I. Lax, J. Schlessinger, Cellular signaling by fibroblast growth factor receptors, Cytokine Growth Factor Rev 16 (2005) 139-149.
[58] D. M. Ornitz, N. Itoh, Fibroblast growth factors, Genome Biol 2 (2001) REVIEWS3005.
[59] L. Vallier, M. Alexander, R. A. Pedersen, Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells, J Cell Sci 118 (2005) 4495-4509.
[60] A. Pebay, R. C. Wong, S. M. Pitson, E. J. Wolvetang, G. S. Peh, A. Filipczyk, K. L. Koh, I. Tellis, L. T. Nguyen, M. F. Pera, Essential roles of sphingosine-1-phosphate and platelet-derived growth factor in the maintenance of human embryonic stem cells, Stem Cells 23 (2005) 1541-1548.
[61] S. C. Bendall, M. H. Stewart, P. Menendez, D. George, K. Vijayaraga-

van, T. Werbowetski-Ogilvie, V. Ramos-Mejia, A. Rouleau, J. Yang, M. Bosse, G. Lajoie, M. Bhatia, IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro, Nature 448 (2007) 1015-1021.

[62] J. Schlessinger, Cell signaling by receptor tyrosine kinases, Cell 103 (2000) 211-225.

[63] S. R. Hubbard, J. H. Till, Protein tyrosine kinase structure and function, Annu Rev Biochem 69 (2000) 373-398.

[64] D. James, A. J. Levine, D. Besser, A. Hemmati-Brivanlou, TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells, Development 132 (2005) 1273-1282.

[65] Y. Shi, J. Massague, Mechanisms of TGF-beta signaling from cell membrane to the nucleus, Cell 113 (2003) 685-700.

[66] B. J. Conley, S. Ellis, L. Gulluyan, R. Mollard, BMPs regulate differentiation of a putative visceral endoderm layer within human embryonic stem-cell-derived embryoid bodies, Biochem Cell Biol 85 (2007) 121-132.

[67] V. M. Ding, L. Ling, S. Natarajan, M. G. Yap, S. M. Cool, A. B. Choo, FGF-2 modulates Wnt signaling in undifferentiated hESC and iPS cells through activated PI3-K/GSK3beta signaling, Journal of cellular physiology 225 (2010) 417-428.

[68] S. P. Gygi, Y. Rochon, B. R. Franza, R. Aebersold, Correlation between Protein and mRNA Abundance in Yeast, Mol Cell Biol 19 (1999) 1720-1730.

[69] M. R. Wilkins, E. Gasteiger, A. A. Gooley, B. R. Herbert, M. P. Molloy, P.-A. Binz, K. Ou, J.-C. Sanchez, A. Bairoch, K. L. Williams, D. F. Hochstrasser, High-throughput mass spectrometric discovery of protein post-translational modifications, J Mol Biol 289 (1999) 645-657.

[70] K. Chandramouli, P.-Y. Qian, Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity, Human Genomics and Proteomics 1 (2009).

[71] G. A. Valaskovic, N. L. Kelleher, D. P. Little, D. J. Aaserud, F. W. McLafferty, Attomole-Sensitivity Electrospray Source for Large-Molecule Mass Spectrometry, Anal Chem 67 (1995) 3802-3805.

[72] M. Mann, O. N. Jensen, Proteomic analysis of post-translational

modifications, Nature biotechnology 21 (2003) 255-261.

[73] B. D. Halligan, V. Ruotti, W. Jin, S. Laffoon, S. N. Twigger, E. A. Dratz, ProMoST (Protein Modification Screening Tool): a web-based tool for mapping protein modifications on two-dimensional gels, Nucleic Acids Research 32 (2004) W638-644.

[74] R. Lu, F. Markowetz, R. D. Unwin, J. T. Leek, E. M. Airoldi, B. D. MacArthur, A. Lachmann, R. Rozov, A. Ma/'ayan, L. A. Boyer, O. G. Troyanskaya, A. D. Whetton, I. R. Lemischka, Systems-level dynamic analyses of fate change in murine embryonic stem cells, Nature 462 (2009) 358-362.

[75] D. L. C. van den Berg, T. Snoek, N. P. Mullin, A. Yates, K. Bezstarosti, J. Demmers, I. Chambers, R. A. Poot, An Oct4-Centered Protein Interaction Network in Embryonic Stem Cells, Cell Stem Cell 6 (2010) 369-381.

[76] L. M. Brill, W. Xiong, K. B. Lee, S. B. Ficarro, A. Crain, Y. Xu, A. Terskikh, E. Y. Snyder, S. Ding, Phosphoproteomic analysis of human embryonic stem cells, Cell Stem Cell 5 (2009) 204-213.

[77] X. Guo, W. Ying, J. Wan, Z. Hu, X. Qian, H. Zhang, F. He, Proteomic characterization of early-stage differentiation of mouse embryonic stem cells into neural cells induced by all-trans retinoic acid in vitro, Electrophoresis 22 (2001) 3067-3075.

[78] K. Nagano, M. Taoka, Y. Yamauchi, C. Itagaki, T. Shinkawa, K. Nunomura, N. Okamura, N. Takahashi, T. Izumi, T. Isobe, Large-scale identification of proteins expressed in mouse embryonic stem cells, Proteomics 5 (2005) 1346-1361.

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

[80] S.-W. Lee, S. J. Berger, S. Martinović, L. Paša-Tolić, G. A. Anderson, Y. Shen, R. Zhao, R. D. Smith, Direct mass spectrometric analysis of intact proteins of the yeast large ribosomal subunit using capillary LC/FTICR, Proceedings of the National Academy of Sciences 99 (2002) 5942-5947.

[81] J. V. Olsen, S.-E. Ong, M. Mann, Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues, Molecular & Cellular Proteomics 3 (2004) 608-614.

[82] A. Guthals, N. Bandeira, Peptide identification by tandem mass spectrometry with alternate fragmentation modes, Molecular & Cellular Proteomics (2012).

[83] J. J. Thelen, J. A. Miernyk, The proteomic future: where mass spectrometry should be taking us, Biochemical Journal 444 (2012) 169-181.

[84] W. Ni, S. Dai, B. L. Karger, Z. S. Zhou, Analysis of isoaspartic Acid by selective proteolysis with Asp-N and electron transfer dissociation mass spectrometry, Anal Chem 82 (2010) 7485-7491.

[85] H. Molina, D. M. Horn, N. Tang, S. Mathivanan, A. Pandey, Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry, Proc Natl Acad Sci U S A 104 (2007) 2199-2204.

[86] N. Taouatas, M. M. Drugan, A. J. Heck, S. Mohammed, Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase, Nature methods 5 (2008) 405-407.

[87] N. Taouatas, A. J. Heck, S. Mohammed, Evaluation of metalloendopeptidase Lys-N protease performance under different sample handling conditions, J Proteome Res 9 (2010) 4282-4288.

[88] D. L. Swaney, C. D. Wenger, J. J. Coon, Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics, J Proteome Res 9 (2010) 1323-1329.

[89] M. Wilm, M. Mann, Analytical properties of the nanoelectrospray ion source, Anal Chem 68 (1996) 1-8.

[90] M. Karas, D. Bachmann, U. Bahr, F. Hillenkamp, Matrix-assisted ultraviolet laser desorption of non-volatile compounds, International Journal of Mass Spectrometry and Ion Processes 78 (1987) 53-68.

[91] J. Fenn, M. Mann, C. Meng, S. Wong, C. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules, Science 246 (1989) 64-71.

[92] X. Han, A. Aslanian, J. R. Yates Iii, Mass spectrometry for proteomics, Current Opinion in Chemical Biology 12 (2008) 483-490.

[93] E. d. Hoffmann, Mass spectrometry : principles and applications. V. Stroobant, Ed., (J. Wiley, Chichester, West Sussex, England ;, ed. 3rd ed., 2007).

[94] J. V. Iribarne, B. A. Thomson, On the evaporation of small ions from charged droplets, The Journal of Chemical Physics 64 (1976) 2287-

2294.

[95] P. Kebarle, U. H. Verkerk, Electrospray: From ions in solution to ions in the gas phase, what we know now, Mass Spectrometry Reviews 28 (2009) 898-917.

[96] R. E. March, An Introduction to Quadrupole Ion Trap Mass Spectrometry, Journal of Mass Spectrometry 32 (1997) 351-369.

[97] G. C. Stafford Jr, P. E. Kelley, J. E. P. Syka, W. E. Reynolds, J. F. J. Todd, Recent improvements in and analytical applications of advanced ion trap technology, International Journal of Mass Spectrometry and Ion Processes 60 (1984) 85-98.

[98] D. J. Douglas, A. J. Frank, D. Mao, Linear ion traps in mass spectrometry, Mass Spectrometry Reviews 24 (2005) 1-29.

[99] R. Aebersold, M. Mann, Mass spectrometry-based proteomics, Nature 422 (2003) 198-207.

[100] A. Makarov, Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis, Anal Chem 72 (2000) 1156-1162.

[101] G. Hopfgartner, E. Varesio, V. Tschappat, C. Grivet, E. Bourgogne, L. A. Leuthold, Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules, J Mass Spectrom 39 (2004) 845-855.

[102] A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat, S. Horning, Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer, Anal Chem 78 (2006) 2113-2120.

[103] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, R. Graham Cooks, The Orbitrap: a new mass spectrometer, J Mass Spectrom 40 (2005) 430-443.

[104] J. V. Olsen, J. C. Schwartz, J. Griep-Raming, M. L. Nielsen, E. Damoc, E. Denisov, O. Lange, P. Remes, D. Taylor, M. Splendore, E. R. Wouters, M. Senko, A. Makarov, M. Mann, S. Horning, A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed, Molecular & Cellular Proteomics 8 (2009) 2759-2769.

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

[106] J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, M. Mann,

Higher-energy C-trap dissociation for peptide modification analysis, Nat Meth 4 (2007) 709-712.

[107] P. Roepstorff, J. Fohlman, Proposal for a common nomenclature for sequence ions in mass spectra of peptides, Biomed Mass Spectrom 11 (1984) 601.

[108] K. Biemann, Appendix 5. Nomenclature for peptide fragment ions (positive ions), Methods Enzymol 193 (1990) 886-887.

[109] R. J. Chalkley, A. Thalhammer, R. Schoepfer, A. L. Burlingame, Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides, Proc Natl Acad Sci U S A 106 (2009) 8894-8899.

[110] J. J. Coon, B. Ueberheide, J. E. Syka, D. D. Dryhurst, J. Ausio, J. Shabanowitz, D. F. Hunt, Protein identification using sequential ion/ion reactions and tandem mass spectrometry, Proc Natl Acad Sci U S A 102 (2005) 9463-9468.

[111] N. L. Young, P. A. Dimaggio, B. A. Garcia, The significance, development and progress of high-throughput combinatorial histone code analysis, Cell Mol Life Sci 67 (2010) 3983-4000.

[112] C. K. Frese, A. F. Altelaar, M. L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A. J. Heck, S. Mohammed, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos, J Proteome Res 10 (2011) 2377-2388.

[113] D. N. Perkins, D. J. Pappin, D. M. Creasy, J. S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis 20 (1999) 3551-3567.

[114] J. K. Eng, A. L. McCormack, J. R. Yates Iii, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, Journal of the American Society for Mass Spectrometry 5 (1994) 976-989.

[115] L. Käll, J. D. Storey, M. J. MacCoss, W. S. Noble, Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases, J Proteome Res 7 (2007) 29-34.

[116] A. Keller, A. I. Nesvizhskii, E. Kolker, R. Aebersold, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search, Anal Chem 74 (2002) 5383-5392.

[117] M. Brosch, L. Yu, T. Hubbard, J. Choudhary, Accurate and Sensi-

tive Peptide Identification with Mascot Percolator, J Proteome Res 8 (2009) 3176-3181.

[118] A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, C. Hamon, Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS, Anal Chem 75 (2003) 1895-1904.

[119] P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D. J. Pappin, Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents, Molecular & Cellular Proteomics 3 (2004) 1154-1169.

[120] P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A. J. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics, Nat Protoc 4 (2009) 484-494.

[121] I. I. Stewart, T. Thomson, D. Figeys, 18O Labeling: a tool for proteomics, Rapid Communications in Mass Spectrometry 15 (2001) 2456-2465.

[122] A. Schmidt, J. Kellermann, F. Lottspeich, A novel strategy for quantitative proteomics using isotope-coded protein labels, Proteomics 5 (2005) 4-15.

[123] J. Kellermann, ICPL--isotope-coded protein label, Methods Mol Biol 424 (2008) 113-123.

[124] S.-E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, M. Mann, Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics, Molecular & Cellular Proteomics 1 (2002) 376-386.

[125] J. Krijgsveld, R. F. Ketting, T. Mahmoudi, J. Johansen, M. Artal-Sanz, C. P. Verrijzer, R. H. A. Plasterk, A. J. R. Heck, Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics, Nat Biotech 21 (2003) 927-931.

[126] H. Liu, R. G. Sadygov, J. R. Yates, 3rd, A model for random sampling and estimation of relative protein abundance in shotgun proteomics, Anal Chem 76 (2004) 4193-4201.

[127] J. C. Silva, R. Denny, C. A. Dorschel, M. Gorenstein, I. J. Kass, G. Z. Li, T. McKenna, M. J. Nold, K. Richardson, P. Young, S. Geromanos, Quantitative proteomic analysis by accurate mass retention

time pairs, Anal Chem 77 (2005) 2187-2200.

[128] P. T. W. Cohen, Protein phosphatase 1 – targeted in many directions, Journal of Cell Science 115 (2002) 241-256.

[129] D. Barford, Molecular mechanisms of theprotein serine/threonine phosphatases, Trends in Biochemical Sciences 21 (1996) 407-412.

[130] E. Salih, Phosphoproteomics by mass spectrometry and classical protein chemistry approaches, Mass Spectrometry Reviews 24 (2005) 828-846.

[131] H. Steen, M. Fernandez, S. Ghaffari, A. Pandey, M. Mann, Phosphotyrosine Mapping in Bcr/Abl Oncoprotein Using Phosphotyrosine-specific Immonium Ion Scanning, Molecular & Cellular Proteomics 2 (2003) 138-145.

[132] L. Andersson, J. Porath, Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography, Anal Biochem 154 (1986) 250-254.

[133] M. C. Posewitz, P. Tempst, Immobilized gallium(III) affinity chromatography of phosphopeptides, Anal Chem 71 (1999) 2883-2892.

[134] D. C. Neville, C. R. Rozanas, E. M. Price, D. B. Gruis, A. S. Verkman, R. R. Townsend, Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry, Protein Sci 6 (1997) 2436-2445.

[135] S. Feng, M. Ye, H. Zhou, X. Jiang, H. Zou, B. Gong, Immobilized zirconium ion affinity chromatography for specific enrichment of phosphopeptides in phosphoproteome analysis, Molecular & cellular proteomics : MCP 6 (2007) 1656-1665.

[136] Y.-Q. Yu, J. Fournier, M. Gilar, J. C. Gebler, Phosphopeptide enrichment using microscale titanium dioxide solid phase extraction, Journal of Separation Science 32 (2009) 1189-1199.

[137] M. W. Pinkse, S. Mohammed, J. W. Gouw, B. van Breukelen, H. R. Vos, A. J. Heck, Highly robust, automated, and sensitive online TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster, J Proteome Res 7 (2008) 687-697.

[138] M. W. H. Pinkse, P. M. Uitto, M. J. Hilhorst, B. Ooms, A. J. R. Heck, Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns, Anal Chem 76 (2004) 3935-3943.

[139] M. R. Larsen, T. E. Thingholm, O. N. Jensen, P. Roepstorff, T. J. D. Jørgensen, Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns, Molecular & Cellular Proteomics 4 (2005) 873-886.

[140] H. Zhou, M. Ye, J. Dong, G. Han, X. Jiang, R. Wu, H. Zou, Specific Phosphopeptide Enrichment with Immobilized Titanium Ion Affinity Chromatography Adsorbent for Phosphoproteome Analysis, J Proteome Res 7 (2008) 3957-3967.

[141] H. Zhou, T. Y. Low, M. L. Hennrich, H. van der Toorn, T. Schwend, H. Zou, S. Mohammed, A. J. R. Heck, Enhancing the Identification of Phosphopeptides from Putative Basophilic Kinase Substrates Using Ti (IV) Based IMAC Enrichment, Molecular & Cellular Proteomics 10 (2011).

[142] T. P. Conrads, T. D. Veenstra, An enriched look at tyrosine phosphorylation, Nat Biotech 23 (2005) 36-37.

[143] T. Hunter, 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 (1998) 583-605.

[144] S. A. Beausoleil, M. Jedrychowski, D. Schwartz, J. E. Elias, J. Villen, J. Li, M. A. Cohn, L. C. Cantley, S. P. Gygi, Large-scale characterization of HeLa cell nuclear phosphoproteins, Proc Natl Acad Sci U S A 101 (2004) 12130-12135.

[145] P. J. Boersema, S. Mohammed, A. J. Heck, Hydrophilic interaction liquid chromatography (HILIC) in proteomics, Anal Bioanal Chem 391 (2008) 151-159.

Chapter II

Investigating the role of FGF-2 in stem cell maintenance by global phosphoproteomics profiling

Adja D. Zoumaro-Djayoon[1], Vanessa Ding[2], Leong-Yan Foong[2], Andre Choo[2], Albert J. R. Heck[1], Javier Muñoz[1]

1-Biomolecular Mass Spectrometry and Proteomics Group. University of Utrecht, Padualaan 8, 3584 CH, Utrecht (The Netherlands) and Netherlands Proteomics Center.

2-Stem Cells Group. Bioprocessing Technology Institute, 20 Biopolis Way. Singapore.

Based on Proteomics. 2011 Oct; 11(20):3962-3971

Abstract

Human embryonic stem cells (hESCs) are of immense interest for regenerative medicine as a source of tissue replacement. Expansion of hESCs as a pluripotent population requires a balance between survival, proliferation and self-renewal signals. One of the key growth factors that maintains this balance is fibroblast growth factor-2 (FGF-2). However, the underlying molecular mechanisms are poorly understood. We recently profiled specifically tyrosine phosphorylation events that occur during FGF-2 stimulation of hESCs (*1*). Here, we complement this phosphoproteome profiling by analyzing temporal dynamics of mostly serine and threonine protein phosphorylation events. Our multi-dimensional strategy combines strong cation exchange (SCX) chromatography to reduce the sample complexity followed by titanium dioxide (TiO2) off-line for the enrichment of phosphopeptides, and dimethylation based stable isotope labeling for quantification. This approach allowed us to identify and quantify 3,261 unique proteins from which 1,064 proteins were found to be phosphorylated in one or more residues (representing 1,653 unique phosphopeptides). Approximately 40% of the proteins (553 unique phosphopeptides) showed differential phosphorylation upon FGF-2 treatment. Amongst those are several members of the canonical pathways involved in pluripotency and self-renewal (e.g. Wnt, PI3K/AKT), hESC-associated proteins such as SOX2, RIF1, SALL4, DPPA4, DNMT3B and 53 proteins that are target genes of the pluripotency transcription factors SOX2, OCT4 and NANOG. These findings complement existing pluripotency analyses and provide new insights into how FGF-2 assists in maintaining the undifferentiated state of hESCs.

1. **Introduction**

Human embryonic stem cells (hESCs) exhibit two exceptional properties, the ability to self-renew over long periods of time *in vitro* and the capability of differentiating into almost all the cell types of the human body. Pluripotency and self-renewal are regulated by a combination of extrinsic and intrinsic factors (*2*). Intrinsic factors regulating pluripotency in hESCs include a core of three transcription factors: OCT4, SOX2 and NANOG (*3-6*). OCT4, SOX2 and NANOG bind to a large number of genes in order to promote transcription (e.g. genes responsible for the maintenance of pluripotency) or repression (e.g. genes involved in the regulation of differentiation) (*7, 8*). Extrinsic factors that promote self-renewal and differentiation of hESCs involve several growth factors such as FGF-2 (*9*), transforming growth factor-β (TGF-β) (*10*), bone morphogenetic protein (BMP) (*11*), platelet derived growth factor (PDGF) (*12*) and insulin growth factor (IGF) (*13*), most of which act through receptor tyrosine kinases (RTKs) (*14*). These RTKs activate intracellular pathways via recruitment and activation of a variety of signaling proteins controlling most fundamental cellular processes including cell cycle progression, cell migration, cell metabolism and survival, as well as cell proliferation and differentiation (*14, 15*).

Amongst the growth factors, FGF-2 is known to be crucial for the *in vitro* expansion of hESCs in feeder-free culture systems, by maintaining hESCs in an undifferentiated state in the absence of serum and feeder cells (*16-18*). FGF-2 is also supplemented to support the growth of hESC using defined media such as StemPro and mTeSR (*19*). In addition, the prolonged withdrawal of FGF-2 in hESC results in loss of pluripotent marker expression (*18, 20*). *In-vivo*, FGF-2 is expressed in various cell types from early embryos to adult cell types and play a role in cell proliferation and migration (*21*) Although, it is clear that FGF-2 can support growth of undifferentiated hESCs, little is known about the underlying molecular mechanisms. To explore the phosphorylation networks triggered by exogenous FGF-2 in hESCs, we previously performed a targeted quantitative analysis of tyrosine phosphorylation events using immuno-affinity purification and mass spectrometry (MS)-based proteomics (*1*). This study showed that FGF-2 treatment resulted in the concomitant activation of the four different FGFRs (FGFR 1, 2, 3 and 4). We observed increased tyrosine phosphorylation of other RTKs such as ERBB2, ERBB3, EGFR, IGF1R, Ephrin receptors and VEGFR. Fur-

thermore, tyrosine phosphorylation was increased on downstream proteins such as PI3K, MAPK and several Src family members, suggesting that these pathways are activated upon exogenous FGF-2 stimulation.
However, tyrosine phosphorylation is less frequent in cells as the vast majority of phosphorylation affects serine and threonine residues (*22, 23*). Therefore, to complement our understanding about the role of exogenous FGF-2 in hESCs, we quantitatively analyzed temporal dynamics of protein phosphorylation during FGF-2 stimulation of hESCs. A global MS-based proteomics approach was used involving SCX and TiO_2 for phosphopeptide enrichment and dimethyl stable isotope labeling for quantification. A total of 3,261 proteins were identified, from which 1,064 were found to be phosphorylated. Interestingly, 363 proteins showed differential phosphorylation profiles upon FGF-2 stimulation in at least one of the four time points studied (1, 5, 15 and 60 min). We discuss these differential changes in Ser and Thr phosphorylation in relationship to our previous data on the targeted Tyr phosphorylation, focusing on proteins and pathways known to be involved in stem cell maintenance.

2. **Materials and methods**

Cell culture, lysis and in-solution digestion

Human ESC line, HES-3 (46,XX) was obtained from ES Cell International (ESI, Singapore). Briefly, the cells were cultured on Matrigel-coated (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) tissue culture dishes and supplemented with conditioned medium from immortalised mouse embryonic fibroblast (Δ-MEF) (*24*). For routine culture, the medium was supplemented with 10 ng/mL of FGF-2 (Invitrogen, Carlsbad, CA, USA), and the medium was changed daily. The cultures were passaged weekly following by enzymatic treatment as previously described (*18*). For FGF-2 starved cultures, cells were maintained in the absence of FGF-2 for 5-7 population doublings (PD, 1 PD = ~1 day). Cells were then stimulated with 10 ng/mL of FGF-2 at the indicated time points.
Cells were lysed on ice in 7 M urea, 2 M thiourea, 40 mM Tris, 50 μg/mL DNase, 50 μg/mL RNase, 1 mM sodium orthovanadate and 1x *PhosSTOP* (Roche Diagnostics, Switzerland) in the presence of protease inhibitors. Protein concentration was determined using Bradford Assay. Total protein lysate of 1.5 mg per time point were reduced with dithiothreitol (DTT) at a fi-

nal concentration of 10 mM at 56°C and subsequently alkylated with 55 mM iodoacetamide. Lysates were diluted 6-fold with 100 mM ammonium bicarbonate and digested overnight with trypsin.

Dimethyl labeling

Tryptic peptides were desalted using a Sep-Pak C18 column (Waters, USA, Massachusetts). The eluted peptides were lyophilized, and re-suspended in 100 μL of 100 mM triethylammonium bicarbonate. Stable isotope dimethyl labeling was performed as previously described (*25, 26*). Labeled samples were mixed in 1:1:1 ratio based on total peptide amount, determined by analyzing an aliquot of the labeled samples on a regular LC-MS/MS run and comparing overall peptide signal intensities.

Strong cation exchange

Prior to the mass spectrometric analysis, samples were fractionated using strong cation exchange (SCX) chromatography as previously described (*27*). SCX was performed using an Agilent 1100 HPLC system (Agilent Technologies, Waldbronn, Germany) with two C_{18} Opti-Lynx (Optimized Technologies, Oregon OR, USA) trapping cartridges and a polysulfoethyl A SCX column (PolyLC, Columbia, MD; 200 mm x 2.1 mm inner diameter, 5 μm, 200-Å). The labeled peptides were dissolved in 10% formic acid (FA) (Sigma-Aldrich), and in total, 1.5 mg of protein were loaded onto the trap columns at 100 μL/min and subsequently eluted onto the SCX column with 80% acetonitrile (ACN) (Biosolve, The Netherlands) and 0.05% FA. SCX buffer A was made of 5 mM KH_2PO_4 (Merck, Germany), 30% ACN and 0.05% FA, pH 2.7; SCX buffer B consisted of 350 mM KCl (Merck, Germany), 5 mM KH_2PO_4, 30% ACN and 0.05% FA, pH 2.7. The gradient was performed as follows: 0% B for 10 min, 0–85% B in 35 min, 85–100% B in 6 min and 100% B for 4 min. A total of 45 fractions were collected and dried in a vacuum centrifuge.

Phosphopeptide enrichment by offline TiO_2

Phosphopeptides were further enriched using home-made GELoader tips

(Eppendorf, Hamburg, Germany) packed with TiO_2 beads (5μm, INERTSIL) (*28*). Peptides were loaded in 10% FA and subsequently washed with either 20 μL of 80% ACN, 0.1% TFA (Fluka, Sigma-Aldrich) or with 20 μL of 80% ACN, 0.1% TFA and 7.7M 2,5-dihydroxybenzoic acid (Sigma-Aldrich) for the SCX fractions abundant in non-phosphorylated peptides. Phosphopeptides were then eluted twice with 20 μL of 1.25% ammonia solution (Merck, Germany), pH 10.5 and 3 μL of 100% FA was finally added to acidify the samples.

On-line nanoflow LC-MS/MS

Nanoflow LC-MS/MS was carried out by coupling an Agilent 1100 HPLC system (Agilent Technologies, Waldbronn, Germany) to an LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany). SCX fractions were dried, reconstituted in 10% FA and delivered to a trap column (Aqua™ C_{18}, 5 μm (Phenomenex, Torrance, CA); 20 mm x 100 μm inner diameter, packed in house) at 5 μL/min in 100% solvent A (0.1 M acetic acid in water). Next, peptides were eluted from the trap column onto an analytical column (ReproSil-Pur C_{18}-AQ, 3 μm (Dr. Maisch GmbH, Ammerbuch, Germany); 40 cm x 50-μm inner diameter, packed in house) at 100 nL/min in a 90 min or 3 h gradient from 0 to 40% solvent B (0.1 M acetic acid in 8:2 (v/v) acetonitrile/water). The eluent was sprayed via distal coated emitter tips butt-connected to the analytical column. The mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS. Full-scan MS spectra (from *m/z* 300 to 1,500) were acquired in the Orbitrap with a resolution of 60,000 at *m/z* 400 after accumulation to a target value of 500,000 in the linear ion trap. The three most intense ions at a threshold > 5,000 were selected for collision-induced fragmentation in the linear ion trap at a normalized collision energy of 35% after accumulation to a target value of 10,000.

Data analysis

Peak lists were created from RAW data files using Bioworks 3.3.1 (Thermo Fisher Scientific, San Jose, CA). Peptide identification was performed using Mascot search engine (version 2.2.2, Matrix Science, London, UK) against a concatenated forward-decoy SwissProt human database (version 56.2, 40,656

sequences) (*29*). Carbamidomethylation of cysteines was used as fixed modification whereas oxidation of methionine and phosphorylation of serine, threonine and tyrosine were set as variable modifications. A dimethyl-based quantitation method was used in Mascot, whereby a given peptide is set to carry a single type of modification ("light", "intermediate" and "heavy" labels) thus preventing hits containing inconsistent labelled peptides (e.g. "Heavy" and "Light" labels in the N-termini and K of a peptide). The database searches were initially performed using a peptide tolerance of 50 ppm and the MS/MS tolerance was set to 0.6 Da. Mascot .dat files were filtered by mass accuracy and Mascot Ion Scores using an in-house .dat-file filter (*30*) resulting in a global FDR of <1%. Quantitative analysis of peptide triplets and PTM scoring to assign the most probable phosphorylation site within the peptide sequence were performed using an in-house dimethyl-adapted version of MSQuant (*31*). Briefly, peptide ratios were determined using the extracted ion chromatograms (XICs) of the monoisotopic peaks of the "light", "intermediate" and "heavy" labeled peptides. To correct for errors during sample mixing, phosphopeptides ratios (log_2 scale) were normalized by substracting the median (log_2 scale) of all the non phosphopeptides ratios. Protein classification was performed using the PANTHER classification system (www.pantherdb.org). The human SwissProt database was used as a reference list to identify under and over-represented protein classes and the Bonferroni correction for multiple testing was applied. The representations of signaling pathways were created using Ingenuity Pathway Analysis tool (IPA; Ingenuity Systems, Mountain View, CA; www.ingenuity.com). The MS data associated with this manuscript can be downloaded from ProteomeCommons.org under the following Tranche hash: zyqKqgezqbO56Nd2yQyDiEeFYxFAksCzxpNYYFE7qy6eWOTrfoHQTq6XTXf7BgS0bYkN+McE1zeYt36TeYKM70jWScMAAAAAAAABr2Q==

3. **Results**

(Phospho)proteome analysis of FGF-2 stimulated hESCs

FGF-2 facilitates the long term culture of hESCs in feeder free culture conditions (*32*). To further understand the role of FGF-2 on hESCs, we extended our previous phosphotyrosine analysis and performed an unbiased phosphoproteomic analysis under identical biological conditions. HESCs deprived of FGF-2 for 5 days showed no expression changes of pluripotency

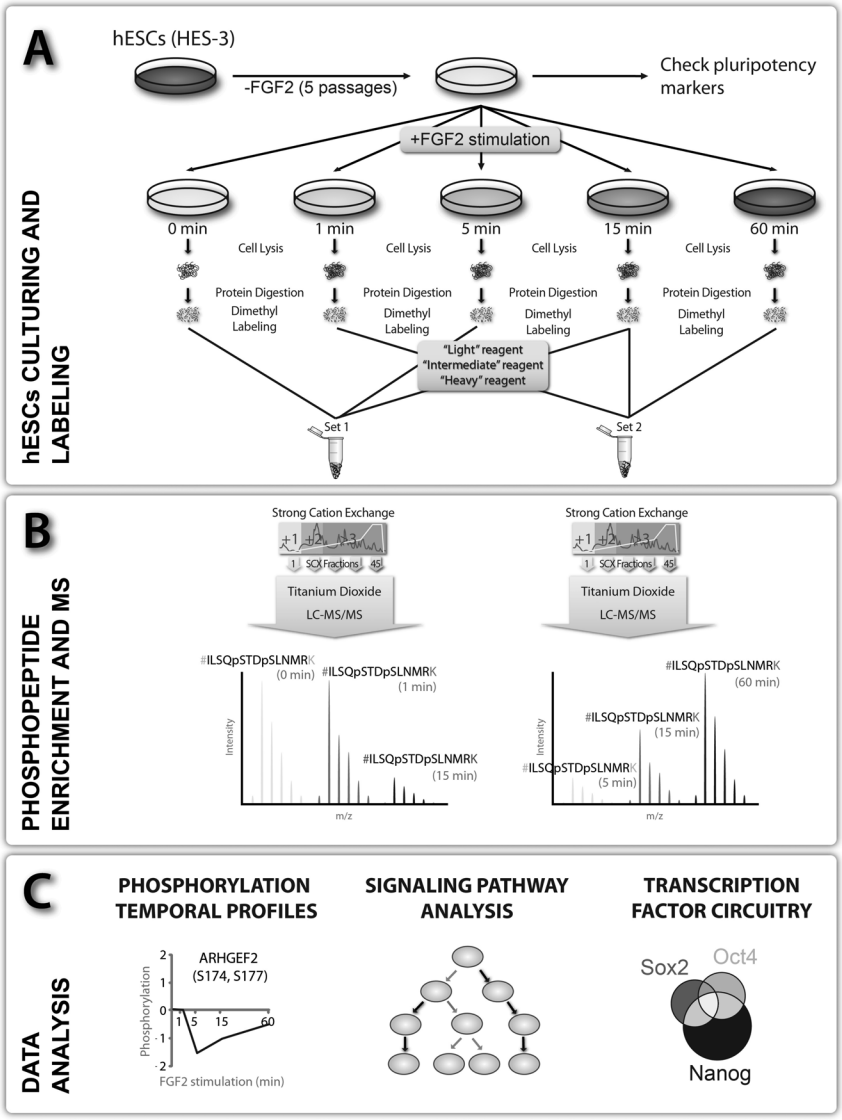


Figure 1: *Overview of the experimental design. A) FGF-2 stimulated hESCs at different time points were harvested and lysed where after the extracted proteins were digested. The resulting peptides were stable isotope labeled using triplex dimethylation (26) and mixed as depicted to create two sets of samples. B) SCX and TiO_2 were used to enrich for phosphopeptides prior to LC-MS/MS. C) The two sets combined provided dynamic profiles of differential phosphorylation during the time course of FGF-2 stimulation.*

markers OCT4, Podocalyxin, Tra1-60 and SSEA3 (*33*) compared to cells routinely cultured with FGF-2 (Figure S1). Additionally, when FGF-2 starved hESCs were stimulated with FGF-2, the cells were able to maintain their undifferentiated state for more than 10 passages (data not shown). These results suggest that FGF-2 activated signaling in the starved hESCs is still functional and maintains hESCs in the undifferentiated state. Subsequently, the FGF 2 starved hESCs were stimulated with 10 ng/mL FGF-2 for 0, 1, 5, 15 and 60 min (Figure 1A) and 1.5 mg of protein for each time point were digested using trypsin. Resulting peptides were labeled using stable isotope dimethyl labeling and combined in two sets as depicted in Figure 1A. Prior to LC-MS/MS, phosphopeptides were enriched using a combination of SCX and TiO_2 chromatography (Figure 1B).

When combined, the LC-MS/MS runs allowed us to identify 3,261 unique proteins (FDR < 1%) (Table S1). Interestingly, a total of 1,064 proteins were found to be phosphorylated at one or more residues from which 810 proteins (representing 1,653 unique phosphopeptides) could be quantified (Table S2). Non-phosphorylated peptides were additionally quantified and negligible variations in protein abundance were found (data not shown). We next carried out an analysis of variability of the quantified phosphopeptides. To this end, reproducibility of the ratios between technical and biological replicates was assessed (Supplementary Table S2) by calculating the relative standard deviations (RSD) for intra-experiment (technical error) and inter-experiment (biological variation) measurements (FigureS2). Around 90% of the quantified proteins and phosphopeptides had relative standard deviations below 30% (Figure S2). Therefore, an arbitrary ratio of ± 0.5 (log_2 scale) was used as a cut-off. On this basis, we determined 363 phosphoproteins (553 phophopeptides) to be differentially phosphorylated during FGF-2 stimulation (Table S2).

Recently, several groups have reported the phosphoproteome of different hESC lines; therefore, we compared our dataset to these studies (*34, 35*). 427 phosphorylated proteins were identified from our data that had not been reported previously in hESCs (Figure S3A). As expected, we found 1,036 phosphorylated proteins that were not identified in our phosphotyrosine analysis (*1*) (Figure S3B). These findings highlight that our analysis is complementary to the phosphotyrosine dataset and certainly adds novel knowledge to the existing hESCs phosphoproteome datasets making this study valuable to the field. To gain insights into the protein content of HES-

3 hESCs, we classified the identified proteins by molecular function (Figure S4). Our proteome falls into multiple functional classes including numerous proteins known to be expressed at low levels in mammalian cells such as kinases (122) and transcription factors (204). We subsequently analyzed our dataset (using all the identified proteins, 3,261) for over and under-represented categories. Nucleic acid binding proteins were found highly over-represented ($p < 0.05$, Bonferroni adjusted for multiple testing) (Figure S4). This finding is in agreement with our previous study in differentiating hESCs (*34*) and might reflect that chromatin remodeling and transcription processes are highly active in hESCs. On the other hand, receptors were found to be under-represented (Figure S4D). These plasma membrane proteins are either present at a low-copy number and/or difficult to detect and more targeted membrane protein enrichment protocols are required (*36*).

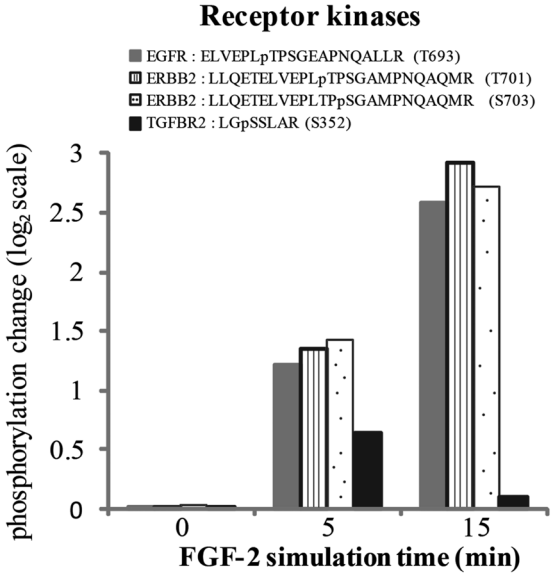


Figure 2. *Several receptor kinases display a change in phosphorylation upon FGF-2 stimulation. Two phosphorylation sites of ERBB2 (T701 and S703) showed an increase in phosphorylation after 5 and 15 min of FGF-2 addition. Another member of the EGFR family, EGFR1 (T693), displayed a similar phosphorylation pattern. TGFβRII (S332) on the other hand, exhibited a transient increase in phosphorylation. Additional information including the relative errors of peptide ratios is shown in Table S1 and Table S2.*

Simultaneous activation of multiple signaling pathways upon FGF-2 stimulation

Many growth factors, including FGF-2, bind to RTKs resulting in a cascade

of signaling events throughout the cell. Our previous study showed a rapid increase in tyrosine phosphorylation of all four FGFRs and ERBB2 after FGF-2 stimulation (*1*). In the present work, two other RTKs from the EGF pathway, ERBB2 (T701, S703) and EGFR1 (T693) showed a significant increase in phosphorylation (Figure 2). Similarly, TGF-βRII (S352) followed a transient activation upon FGF-2 stimulation (Figure 2).

Moreover, downstream signaling proteins involved in PI3K, MAPK, Wnt, and actin/cytoskeletal canonical pathways were also found to be differentially phosphorylated. PIK3C2A (S338), AKT1S1 (T246), TBC1D4 (S341; S318) and p53 (S313) were decreased in phosphorylation on indicated residues (Figure 3, Table S2). Additionally, MDM2 (S166) and CHK1 (S88) were increased in phosphorylation on the belonging phosphosites (Figure 3). Phosphorylation of MDM2 at S166 inhibits p53-mediated cell death by targeting p53 for proteosomal degradation (*37*) (Figure 3). The regulation in phosphorylation of numerous AKT1 substrates from our data suggested that PKB/AKT was activated during FGF-2 exposure. This is in agreement with our previous observations where FGF-2 stimulated hESCs activated both ERK/MAPK and PI3K/AKT pathways (*1*).

Additionally, changes in phosphorylation levels of GSK3A, GSK3B, CTNNB1 and CSNK1A1 (Figure S5), belonging to the Wnt signaling pathway, were observed. GS3KA and GSK3B phosphorylation at sites Y279 and Y216, respectively, which are located near their activation sites (*38*) were found to be increased after 15 min of FGF-2 stimulation for GSK3B, and 60 min for GSK3A (Figure 3). GSK3A displayed, in addition, an increase in phosphorylation on T347 (Table S2). CTNNB1 was found to be increased in phosphorylation on T551 within 5 min of FGF-2 treatment, however, a decrease in phosphorylation to basal levels was observed after 15 min (Figure S5). Moreover, the priming kinase that phosphorylates β-catenin, CK1 (CSNK1A1) was decreased in phosphorylation at T321 at 15 min (Figure S5). These results suggest activation of PI3K/GSK3 pathway as a consequence of FGF-2 treatment.

A large number of members of the actin/cytoskeletal pathway were also found to be differentially phosphorylated upon FGF-2 stimulation. In addition to their well-known structural function, cytoskeletal proteins play important roles in cellular processes including signal transduction (*39*). Phosphorylation of several proteins belonging to the focal adhesion assembly was found to be influenced by FGF-2 treatment. Structural components of focal

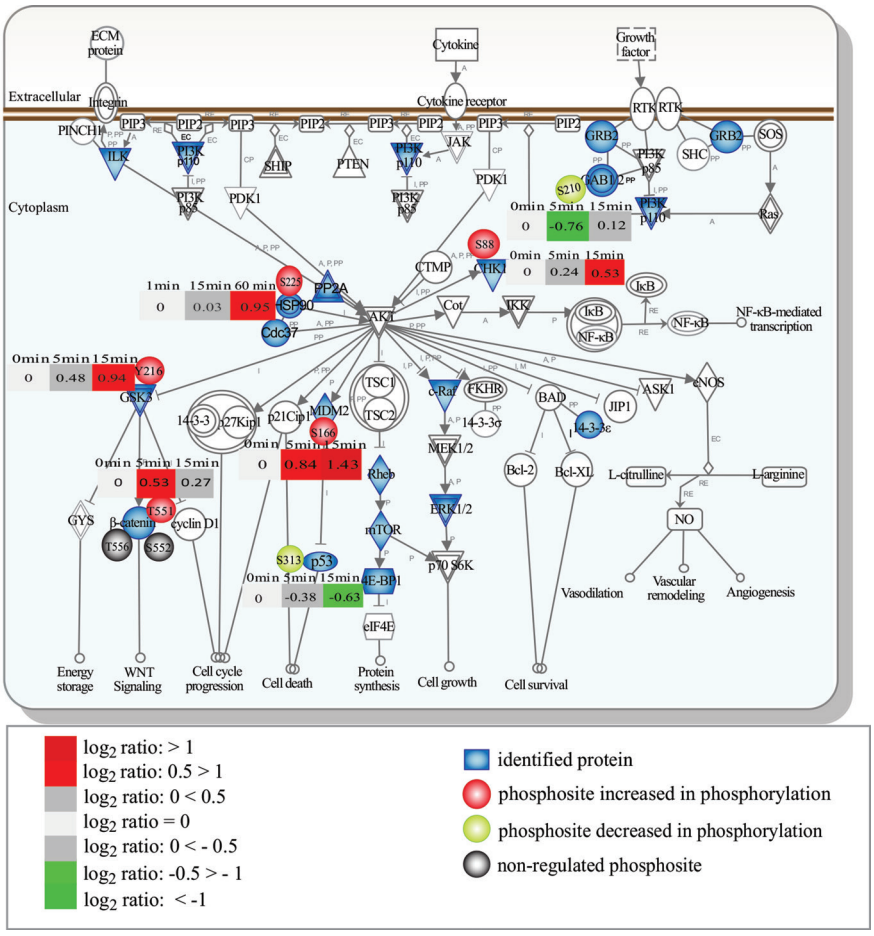


Figure 3: *Observed differential phosphorylation in the PI3K/AKT pathway. In blue are annotated the phosphoproteins that were identified in this study. Changes in phosphorylation are displayed as well using the log_2 ratios, which are depicted for each measured time point. Additional information including the relative errors of peptide ratios is displayed in Table S1 and Table S2. Increased phosphorylation is annotated in red, whereas decreased phosphorylation is displayed in green. Phosphosites that are not changing are shown in grey.*

adhesions such as vinculin (VCL), zyxin (ZYX) and filamin A/C (FLNA/C) were decreased in phosphorylation, whereas filamin B and binding partners of vinculin, Src kinase paxillin and SORBS3 were increased (Figure 4A). Scr substrates TJP2, PKP3 and PKP4 found in tight junctions were also changing in phosphorylation upon FGF-2 stimulation (Figure 4B). Proteins involved in actin/cytoskeletal remodeling undergoing a change in phosphorylation in our data set include: ABI2, VIM, STMN1, AFAP1 and ABLIM1 (Figure 4C).

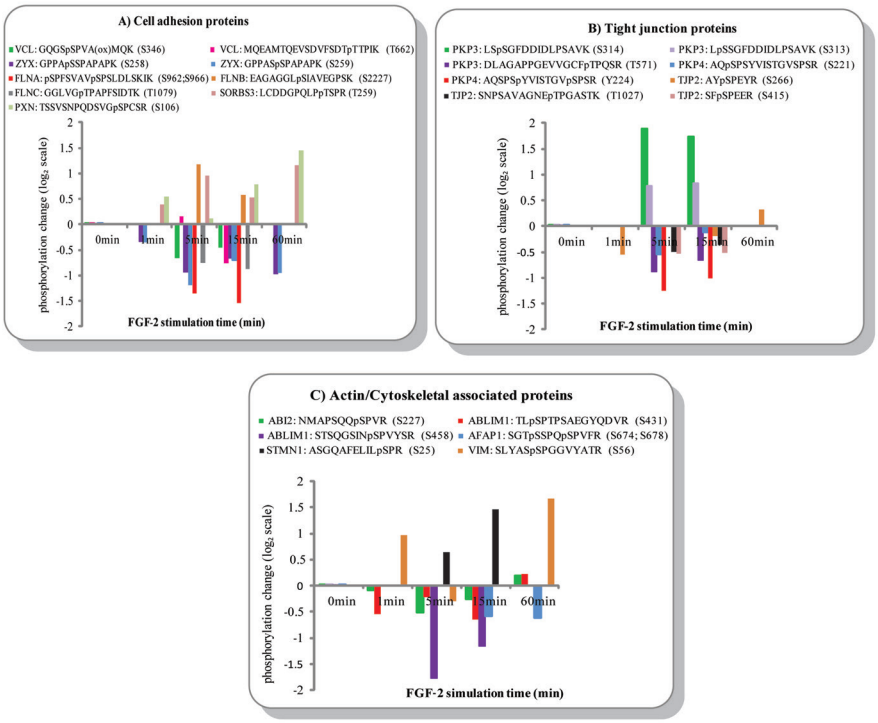


Figure 4: *Observed differential phosphorylation in cytoskeletal related proteins. Phosphorylation changes are displayed as log_2 ratios for a) focal adhesion proteins b) proteins located at the tight junction and c) proteins associated with actin/cytosketetal. Most of these proteins displayed a decrease in phosphorylation during FGF-2 stimulation. The full list of phosphoproteins including peptide ratios and RSDs can be found in Table S1 and S2.*

Phosphorylation of the STMN1 protein is known to cause instability resulting in the disassembly of microtubules (*40*). Throughout the regulated cytoskeleton interactors, several were shown in our previous study (*1*) including Src substrates to display an increased level of tyrosine phosphorylation indicating that FGF-2 treatment not only leads to the earlier mentioned signaling pathway activation, but might also be involved in the reorganization of the cytoskeleton and cell adhesion.

<u>Several pluripotency transcriptional regulators become differentially phosphorylated during early FGF-2 stimulation</u>

Nucleic acid binding proteins and transcription factors represent by far one of the largest classes of regulated phosphoproteins in our dataset (Figure

S4). Amongst them, two of the three most important pluripotency transcription factors, OCT4 and SOX2 (Table S1) were identified. These genes are known to be expressed at low copy number indicating the reasonably good sensitivity of our analysis. Most importantly, two phosphopeptides for SOX2 were found to be differentially phosphorylated by FGF-2. We detected a doubly phosphorylated peptide (S249, S251) which was increased in phosphorylation and a singly phosphorylated peptide (S249) which displayed a decrease in phosphorylation (Table S2). NANOG is a homeobox-containing transcription factor regulating cell fate in pluripotent hESCs. NANOG is regulated by binding to SOX2 and OCT4, and together the trimeric complex form the core transcriptional regulatory circuitry of hESCs that inhibit genes related to differentiation and express genes important for pluripotency (*7*). To find out if FGF-2 affected any downstream targets of these transcription regulators, we mapped our regulated phosphoproteins to chromatin immunoprecipitation data sets published by Boyer *et al.* Of the 363 regulated phoshoproteins, approximately 16% (57 proteins) were target genes of at least one of these transcription factors (Figure 5). Importantly, 10 phosphoproteins are target genes common to the three pluripotency associated proteins, including transcriptional regulators and stem cell associated proteins (e.g. DPPA4, SOX2, STAT3, RIF1), other DNA binding proteins (e.g. IRX2, SMARCAD1, TOP2A) and cytoskeletal remodeling and cell-cell signaling proteins (e.g. DPYSL2, DPYSL3, GJA1). STAT3 increased in phosphorylation on S727. Phosphorylation of this residue is required for its transcriptional activity (*41*). SALL4, another transcription factor involved in self-renewal (*42*) and pluripotency (*43*) showed a 2-fold increase in phosphorylation on S800. An additional transcription regulator associated with pluripotency and self-renewal, DPPA4 (*44*) was found to decrease in phosphorylation on S7 and increase in phosphorylation on T215 (Table S2). Moreover, DNMT3B, highly expressed in hESCs and essential for *de novo* methylation (*45*), was increased in phosphorylation on S137 and S387. Phosphorylation of the transcription factor ATF2 on T71 was increased within 15 min of stimulation (Table S2). This residue is phosphorylated through the MAPK/ERK pathway induced by EGF or insulin (*46*). ATF7, a homologue of ATF2, was increased at 15 and 60 min of FGF-2 stimulation at activation sites T51 and T53 (Table S2) which are required for its transcriptional activation processes (*47*). Although the function of most of these proteins is not fully established in stem cells, the observed changes in phosphorylation of 57 proteins con-

trolled by OCT4, SOX2 and NANOG emphasize the importance of FGF-2 signaling in undifferentiated hESCs.

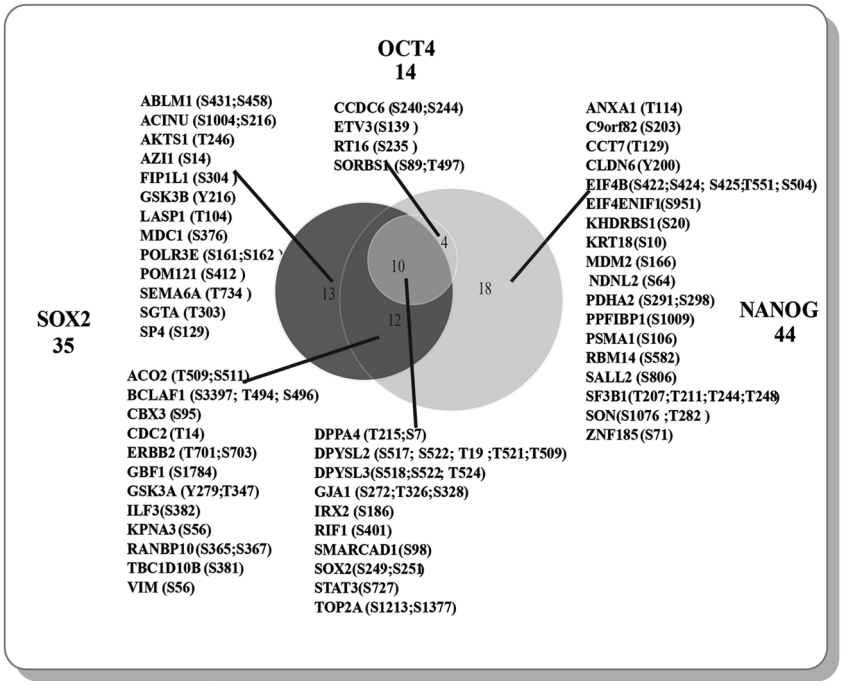


Figure 5: *Downstream targets of the core transcriptional regulatory circuitry regulated by phosphorylation. Venn diagram representing the overlap of target genes of SOX2, OCT4 and NANOG differentially phosphorylated upon FGF-2 exposure. Regulated phosphosites are displayed for every section of the diagram.*

4. **Discussion**

The analysis of molecular changes associated with exogenous FGF-2 supplementation in hESCs is essential although it is still unclear how FGF-2 assists in maintaining the undifferentiated state (*16, 17*). FGF-2 executes its biological function by first binding to the FGFRs (FGFR1-4), and subsequently activating the downstream signaling pathways. Since FGF-2 signaling initiates *via* RTKs, we initially performed a targeted tyrosine phosphorylation analysis of hESCs followed by FGF-2 stimulation (*1*), observing activation of multiple RTKs as well as downstream targets of PI3K, MAPK, and Src family members. We have now complemented this initial study by performing

a global unbiased phosphorylation analysis of the same hESCs stimulated by FGF-2 and observed more than 300 regulated phosphoproteins. Complementing our previous analysis, we show that FGF-2 may result in the activation of several other downstream signaling pathways including Wnt and actin/cytoskeletal pathways, which could regulate numerous transcription factors including several pluripotency-associated proteins such as SOX2, DNMT3B, RIF1, DPPA4, SALL4 and STAT3.

Our results showed that the stimulation of FGF-2 caused a significant increase in serine and threonine phosphorylation of two EGFRs: EGFR1 and ERBB2. This is consistent with our previous observations using phospho-RTK arrays whereby the addition of FGF-2 not only resulted in an increase in the phosphorylation level of the FGFRs but also EGFR family receptors. Heterodimerization of ERBB2 and ERBB3 has been proposed to be important in pluripotent hESCs (*48*). TGFBR2, a serine/threonine kinase receptor was also increased in phosphorylation. When activated by TGF-β superfamily members, TGFBR2 phosphorylates TGFBR1 (*49*), which transduces the signal by consecutive activation of SMAD proteins. This results in the complex assembly of transcription factors in the nucleus which includes SMAD4, regulating in this way specific gene expression (*50*). The activation of the TGF-β receptors by TGF-β/Activin A/Nodal promotes self-renewal in undifferentiated hESCs (*20*). Nevertheless, it is possible that the activation of FGFR family may co-operate synergistically and possibly transactivate with the other RTKs to maintain undifferentiated hESC.

Both the MAPK/ERK and PI3K/AKT pathways were recently shown to be activated by exogenous FGF-2 in hESCs (*18*). Moreover, activation of these pathways have been suggested to be involved in the maintenance of undifferentiated hESC (*18, 51, 52*). In this study, we identified MEKK1, MEK2, MEK3, ERK1 and ERK2 in our datasets and found ERK3 (MAPK6) and MEKKK4 (MAP4K4), an activator of the JUNK pathway (*53*), to be decreased in phosphorylation (Table S2). In our quest to profile the phosphorylation pattern of FGF-2 stimulated hESCs, we also observed numerous substrates of AKT differentially phosphorlyated upon FGF-2 stimulation. These substrates are key components of the Wnt signaling pathway. For instance, GSK3A and GSK3B were increased in phosphorylation on their activation sites (*38*). The regulation of Wnt signaling in hESCs could potentially promote either self-renewal or differentiation depending on the presence or absence of pluripotency supporting factors such as FGF-2 or conditioned medium (*54*).

CK1 and GSK3 phosphorylate β-catenin, activating its rapid degradation by the proteasome (*38*, *55*, *56*). In the presence of Wnt, GSK3 is inhibited causing the stabilization of β-catenin and inducing its accumulation and translocation into the nucleus where it forms a complex with TCF and provokes transcription of the Wnt target genes. Previously, we observed that the inhibition of PI3K/AKT pathway resulted in the loss of pluripotency marker expression (*18*). In addition, stimulation of undifferentiated hESC with FGF-2 resulted in the accumulation of β-catenin in the nucleus. Importantly, a three-fold increase in TCF/LEF activity was also observed after 4 h of FGF-2 stimulation (*18*). Taking these results together, we propose that FGF-2 can modulate the biphasic Wnt signaling pathway in order to maintain hESCs at its undifferentiated state.

Apart from the Wnt, PI3K/AKT and MAPK/ERK pathways, we postulate that the exposure of FGF-2 on hESCs concomitantly results in cytoskeleton rearrangement since a large number of cytoskeletal proteins and interactors were regulated in phosphorylation. Previously, we had also observed a large number of Src kinase substrates that had shown a distinct increase in tyrosine phosphorylation, and demonstrated the involvement of Src family kinases in the maintenance of pluripotency. Results from the current dataset corroborates that FGF-2 stimulation could be involved in regulation of cytoskeletal/actin dependent processes including modulation of cell-adhesion, cell-cell interaction, and formation of tight junction. Therefore, both datasets imply that FGF-2 stimulation regulates these processes for the maintenance of hESC pluripotency.

The importance of FGF-2 in the maintenance of pluripotent hESC has been widely reported. However, none of these reports had made any direct link between FGF-2 mediated signaling and the regulation of hESC pluripotency. In this study, we found that SOX2 is regulated through phosphorylation (S249, S251) upon FGF-2 stimulation. Interestingly, we previously showed that these phosphosites enhance the sumoylation of the protein (*34*) which is known to inhibit its DNA binding properties (*57*) suggesting a potential molecular mechanism for SOX2 activation/deactivation. To determine the effect of FGF-2 on the downstream targets of the core pluripotency circuitry, we mapped the regulated phosphoproteins to ChIP data for OCT4, SOX2 and NANOG (*7*). We found that 16% of the 363 regulated phosphoproteins were targets of at least one of these transcriptional factors. This result establish-

es a connection between FGF-2 signaling and the transcriptional regulatory circuitry of hESCs and provides evidence that FGF-2 may directly regulate the OCT4/SOX2/NANOG trimeric complex to maintain the undifferentiated hESC phenotype. Our observation is in agreement with previous results from our group where the pluripotency circuitry was also found to be regulated by phosphorylation upon BMP-4 induced differentiation of hESCs (*34*) and highlights the link between external growth factors-mediated signaling and the transcriptional circuitry involved in the control of pluripotency and self-renewal.

5. **Conclusion**

Multiple signaling pathways have been suggested to be required for the maintenance of undifferentiated hESCs (*48*), however the underlying molecular mechanisms induced by these pathways remain poorly understood. The results of our study emphasize that FGF-2 plays a key role in signaling pathway controlling the maintenance of undifferentiated hESCs through the phosphorylation of: (i) FGFRs and other RTKs which in-turn activate proteins belonging to different downstream signaling pathways, such as ERK, Wnt and PI3K/AKT, (ii) cytoskeleton associated proteins, and (iii) pluripotency-associated proteinsand transcription regulators. Therefore, it is our hope that our data will serve as a resource for future investigations into stem cell maintenance.

6. **Acknowledgements**

We thank Dr. Shabaz Mohammed for mass spectrometry support; Dr. Andreas Helbig and Marco Hennrich for SCX; Henk van den Toorn for in-house software development and others members from Heck Lab for fruitful discussions. This work was supported by the Netherlands Proteomics Centre, which is part of the Netherlands Genomics Initiative and by the Agency for Science Technology and Research (A*STAR), Singapore.

Supplementary data is available at **http://tinyurl.com/azdchapter2**

7. References

[1] V. M. Y. Ding, P. J. Boersema, L. Y. Foong, C. Preisinger, G. Koh, S. Natarajan, D.-Y. Lee, J. Boekhorst, B. Snel, S. Lemeer, A. J. R. Heck, A. Choo, Tyrosine Phosphorylation Profiling in FGF-2 Stimulated Human Embryonic Stem Cells, PLoS ONE **6** (2011) e17538.

[2] H. Zhang, Z. Z. Wang, Mechanisms that mediate stem cell self-renewal and differentiation, J Cell Biochem **103** (2008) 709-718.

[3] I. Chambers, D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, A. Smith, Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells, Cell **113** (2003) 643-655.

[4] M. Boiani, H. R. Scholer, Regulatory networks in embryo-derived pluripotent stem cells, Nat Rev Mol Cell Biol **6** (2005) 872-884.

[5] J. Nichols, B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Scholer, A. Smith, Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4, Cell **95** (1998) 379-391.

[6] A. A. Avilion, S. K. Nicolis, L. H. Pevny, L. Perez, N. Vivian, R. Lovell-Badge, Multipotent cell lineages in early mouse development depend on SOX2 function, Genes Dev **17** (2003) 126-140.

[7] L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, R. A. Young, Core transcriptional regulatory circuitry in human embryonic stem cells, Cell **122** (2005) 947-956.

[8] N. Ivanova, R. Dobrin, R. Lu, I. Kotenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, I. R. Lemischka, Dissecting self-renewal in stem cells with RNA interference, Nature **442** (2006) 533-538.

[9] S. Avery, K. Inniss, H. Moore, The regulation of self-renewal in human embryonic stem cells, Stem Cells Dev **15** (2006) 729-740.

[10] L. Vallier, M. Alexander, R. A. Pedersen, Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells, J Cell Sci **118** (2005) 4495-4509.

[11] R. H. Xu, X. Chen, D. S. Li, R. Li, G. C. Addicks, C. Glennon, T. P. Zwaka, J. A. Thomson, BMP4 initiates human embryonic stem cell differentiation to trophoblast, Nat Biotechnol **20** (2002) 1261-1264.

[12] A. Pebay, R. C. Wong, S. M. Pitson, E. J. Wolvetang, G. S. Peh, A. Filipczyk, K. L. Koh, I. Tellis, L. T. Nguyen, M. F. Pera, Essential roles of sphin-

gosine-1-phosphate and platelet-derived growth factor in the maintenance of human embryonic stem cells, Stem Cells **23** (2005) 1541-1548.
[13] S. C. Bendall, M. H. Stewart, P. Menendez, D. George, K. Vijayaragavan, T. Werbowetski-Ogilvie, V. Ramos-Mejia, A. Rouleau, J. Yang, M. Bosse, G. Lajoie, M. Bhatia, IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro, Nature **448** (2007) 1015-1021.
[14] J. Schlessinger, Cell signaling by receptor tyrosine kinases, Cell **103** (2000) 211-225.
[15] S. R. Hubbard, J. H. Till, Protein tyrosine kinase structure and function, Annu Rev Biochem **69** (2000) 373-398.
[16] M. E. Levenstein, T. E. Ludwig, R. H. Xu, R. A. Llanas, K. VanDenHeuvel-Kramer, D. Manning, J. A. Thomson, Basic fibroblast growth factor support of human embryonic stem cell self-renewal, Stem Cells **24** (2006) 568-574.
[17] R. H. Xu, R. M. Peck, D. S. Li, X. Feng, T. Ludwig, J. A. Thomson, Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells, Nat Methods **2** (2005) 185-190.
[18] V. M. Ding, L. Ling, S. Natarajan, M. G. Yap, S. M. Cool, A. B. Choo, FGF-2 modulates Wnt signaling in undifferentiated hESC and iPS cells through activated PI3-K/GSK3beta signaling, J Cell Physiol **225** 417-428.
[19] A. C. Chin, J. Padmanabhan, S. K. Oh, A. B. Choo, Defined and serum-free media support undifferentiated human embryonic stem cell growth, Stem Cells Dev **19** 753-761.
[20] B. Greber, H. Lehrach, J. Adjaye, Fibroblast growth factor 2 modulates transforming growth factor beta signaling in mouse embryonic fibroblasts and human ESCs (hESCs) to support hESC self-renewal, Stem Cells **25** (2007) 455-464.
[21] R. T. Bottcher, C. Niehrs, Fibroblast growth factor signaling during early vertebrate development, Endocr Rev **26** (2005) 63-77.
[22] M. Mann, S. E. Ong, M. Gronborg, H. Steen, O. N. Jensen, A. Pandey, Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome, Trends Biotechnol **20** (2002) 261-268.
[23] S. Lemeer, A. J. Heck, The phosphoproteomics data explosion, Curr Opin Chem Biol **13** (2009) 414-420.
[24] A. Choo, J. Padmanabhan, A. Chin, W. J. Fong, S. K. Oh, Immortalized feeders for the scale-up of human embryonic stem cells in feeder and fe-

eder-free conditions, J Biotechnol **122** (2006) 130-141.
[25] P. J. Boersema, L. Y. Foong, V. M. Ding, S. Lemeer, B. van Breukelen, R. Philp, J. Boekhorst, B. Snel, J. den Hertog, A. B. Choo, A. J. Heck, In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling, Mol Cell Proteomics **9** 84-99.
[26] P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A. J. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics, Nat Protoc **4** (2009) 484-494.
[27] S. Gauci, A. O. Helbig, M. Slijper, J. Krijgsveld, A. J. Heck, S. Mohammed, Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach, Anal Chem **81** (2009) 4493-4501.
[28] M. W. Pinkse, S. Mohammed, J. W. Gouw, B. van Breukelen, H. R. Vos, A. J. Heck, Highly robust, automated, and sensitive online TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster, J Proteome Res **7** (2008) 687-697.
[29] J. E. Elias, S. P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, Nat Methods **4** (2007) 207-214.
[30] H. W. van den Toorn, J. Munoz, S. Mohammed, R. Raijmakers, A. J. Heck, B. van Breukelen, RockerBox: analysis and filtering of massive proteomics search results, J Proteome Res.
[31] P. Mortensen, J. W. Gouw, J. V. Olsen, S. E. Ong, K. T. Rigbolt, J. Bunkenborg, J. Cox, L. J. Foster, A. J. Heck, B. Blagoev, J. S. Andersen, M. Mann, MSQuant, an open source platform for mass spectrometry-based quantitative proteomics, J Proteome Res **9** 393-403.
[32] M. Amit, C. Shariki, V. Margulets, J. Itskovitz-Eldor, Feeder layer- and serum-free culture of human embryonic stem cells, Biol Reprod **70** (2004) 837-845.
[33] H. L. Tan, W. J. Fong, E. H. Lee, M. Yap, A. Choo, mAb 84, a cytotoxic antibody that kills undifferentiated human embryonic stem cells via oncosis, Stem Cells **27** (2009) 1792-1801.
[34] D. Van Hoof, J. Munoz, S. R. Braam, M. W. Pinkse, R. Linding, A. J. Heck, C. L. Mummery, J. Krijgsveld, Phosphorylation dynamics during early differentiation of human embryonic stem cells, Cell Stem Cell **5** (2009) 214-226.
[35] L. M. Brill, W. Xiong, K. B. Lee, S. B. Ficarro, A. Crain, Y. Xu, A. Ter-

skikh, E. Y. Snyder, S. Ding, Phosphoproteomic analysis of human embryonic stem cells, Cell Stem Cell **5** (2009) 204-213.
[36] W. Dormeyer, D. van Hoof, S. R. Braam, A. J. Heck, C. L. Mummery, J. Krijgsveld, Plasma membrane proteomics of human embryonic stem cells and human embryonal carcinoma cells, J Proteome Res **7** (2008) 2936-2951.
[37] A. J. Levine, W. Hu, Z. Feng, The P53 pathway: what questions remain to be explored?, Cell Death Differ **13** (2006) 1027-1036.
[38] K. Hughes, E. Nikolakaki, S. E. Plyte, N. F. Totty, J. R. Woodgett, Modulation of the glycogen synthase kinase-3 family by tyrosine phosphorylation, EMBO J **12** (1993) 803-808.
[39] I. Pucci-Minafra, P. Cancemi, S. Fontana, L. Minafra, S. Feo, M. Becchi, A. M. Freyria, S. Minafra, Expanding the protein catalogue in the proteome reference map of human breast cancer cells, Proteomics **6** (2006) 2609-2625.
[40] G. Wallon, J. Rappsilber, M. Mann, L. Serrano, Model for stathmin/OP18 binding to tubulin, EMBO J **19** (2000) 213-222.
[41] K. Abe, M. Hirai, K. Mizuno, N. Higashi, T. Sekimoto, T. Miki, T. Hirano, K. Nakajima, The YXXQ motif in gp 130 is crucial for STAT3 phosphorylation at Ser727 through an H7-sensitive kinase pathway, Oncogene **20** (2001) 3464-3474.
[42] S. Yuri, S. Fujimura, K. Nimura, N. Takeda, Y. Toyooka, Y. Fujimura, H. Aburatani, K. Ura, H. Koseki, H. Niwa, R. Nishinakamura, Sall4 is essential for stabilization, but not for pluripotency, of embryonic stem cells by repressing aberrant trophectoderm gene expression, Stem Cells **27** (2009) 796-805.
[43] J. Zhang, W. L. Tam, G. Q. Tong, Q. Wu, H. Y. Chan, B. S. Soh, Y. Lou, J. Yang, Y. Ma, L. Chai, H. H. Ng, T. Lufkin, P. Robson, B. Lim, Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1, Nat Cell Biol **8** (2006) 1114-1123.
[44] B. Madan, V. Madan, O. Weber, P. Tropel, C. Blum, E. Kieffer, S. Viville, H. J. Fehling, The pluripotency-associated gene Dppa4 is dispensable for embryonic stem cell identity and germ cell development but essential for embryogenesis, Mol Cell Biol **29** (2009) 3186-3203.
[45] M. Okano, D. W. Bell, D. A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, Cell **99** (1999) 247-257.

[46] D. M. Ouwens, N. D. de Ruiter, G. C. van der Zon, A. P. Carter, J. Schouten, C. van der Burgt, K. Kooistra, J. L. Bos, J. A. Maassen, H. van Dam, Growth factors can activate ATF2 via a two-step mechanism: phosphorylation of Thr71 through the Ras-MEK-ERK pathway and of Thr69 through RalGDS-Src-p38, EMBO J **21** (2002) 3782-3793.
[47] F. De Graeve, A. Bahr, K. T. Sabapathy, C. Hauss, E. F. Wagner, C. Kedinger, B. Chatton, Role of the ATFa/JNK2 complex in Jun activation, Oncogene **18** (1999) 3491-3500.
[48] L. Wang, T. C. Schulz, E. S. Sherrer, D. S. Dauphin, S. Shin, A. M. Nelson, C. B. Ware, M. Zhan, C. Z. Song, X. Chen, S. N. Brimble, A. McLean, M. J. Galeano, E. W. Uhl, K. A. D'Amour, J. D. Chesnut, M. S. Rao, C. A. Blau, A. J. Robins, Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling, Blood **110** (2007) 4111-4119.
[49] X. H. Feng, R. Derynck, Ligand-independent activation of transforming growth factor (TGF) beta signaling pathways by heteromeric cytoplasmic domains of TGF-beta receptors, J Biol Chem **271** (1996) 13123-13129.
[50] Y. Shi, J. Massague, Mechanisms of TGF-beta signaling from cell membrane to the nucleus, Cell **113** (2003) 685-700.
[51] P. Dvorak, D. Dvorakova, S. Koskova, M. Vodinska, M. Najvirtova, D. Krekac, A. Hampl, Expression and potential role of fibroblast growth factor 2 and its receptors in human embryonic stem cells, Stem Cells **23** (2005) 1200-1211.
[52] L. Armstrong, O. Hughes, S. Yung, L. Hyslop, R. Stewart, I. Wappler, H. Peters, T. Walter, P. Stojkovic, J. Evans, M. Stojkovic, M. Lako, The role of PI3K/AKT, MAPK/ERK and NFkappabeta signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis, Hum Mol Genet **15** (2006) 1894-1913.
[53] Z. Yao, G. Zhou, X. S. Wang, A. Brown, K. Diener, H. Gan, T. H. Tan, A novel human STE20-related protein kinase, HGK, that specifically activates the c-Jun N-terminal kinase signaling pathway, J Biol Chem **274** (1999) 2118-2125.
[54] L. Cai, Z. Ye, B. Y. Zhou, P. Mali, C. Zhou, L. Cheng, Promoting human embryonic stem cell renewal or differentiation by modulating Wnt signal and culture conditions, Cell Res **17** (2007) 62-72.
[55] C. Liu, Y. Li, M. Semenov, C. Han, G. H. Baeg, Y. Tan, Z. Zhang,

X. Lin, X. He, Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism, Cell **108** (2002) 837-847.
[56] S. Amit, A. Hatzubai, Y. Birman, J. S. Andersen, E. Ben-Shushan, M. Mann, Y. Ben-Neriah, I. Alkalay, Axin-mediated CKI phosphorylation of beta-catenin at Ser 45: a molecular switch for the Wnt pathway, Genes Dev **16** (2002) 1066-1076.
[57] S. Tsuruzoe, K. Ishihara, Y. Uchimura, S. Watanabe, Y. Sekita, T. Aoto, H. Saitoh, Y. Yuasa, H. Niwa, M. Kawasuji, H. Baba, M. Nakao, Inhibition of DNA binding of Sox2 by the SUMO conjugation, Biochem Biophys Res Commun **351** (2006) 920-926.

Chapter III

Targeted analysis of tyrosine phosphorylation by immuno-affinity enrichment of tyrosine phosphorylated peptides prior to mass spectrometric analysis

Adja D. Zoumaro-Djayoon, Albert J. R. Heck, Javier Muñoz

Biomolecular Mass Spectrometry and Proteomics. Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH, Utrecht (The Netherlands) and Netherlands Proteomics Center.

Based on Methods. 2012 Feb;56(2):268-274

Abstract

Tyrosine phosphorylation is a key process that regulates seminal biological functions, hence, deregulation of this mechanism is an underlying cause of several diseases including cancer and immunological disorders. Due to its low abundance, tyrosine phosphorylation is typically under-represented in most of the global MS-based phosphoproteomic studies. Here, we describe a selective approach based on immuno-affinity purification using specific antibodies to enrich tyrosine phosphorylated peptides from a complex proteolytic digest. LC-MS/MS analysis is subsequently used for peptide identification allowing the exact localization of the phosphorylated residue within the sequence. Using this approach more than 1,000 non-redundant phosphotyrosine peptides can be identified in less than 6 h of MS analysis, reflecting the high sensitivity and specificity of the technique. The identified tyrosine phosphorylated peptides can be used to study different biological aspects of tyrosine signaling and disease.

1. **Introduction**

Since its discovery in 1979 (*1*), reversible phosphorylation of tyrosines in proteins has emerged as an essential molecular mechanism that controls important protein functions. In eukaryotes, tyrosine phosphorylation plays a key role in regulating several processes such as cell growth, cell cycle, differentiation, cell motility and gene transcription (*2*). A fine balance of protein tyrosine kinases (PTKs), protein tyrosine phosphatases (PTPs), associated with numerous adapters and scaffolding proteins regulates the process of tyrosine phosphorylation and de-phosphorylation. The human genome encodes for 90 tyrosine kinases (*3*) and 107 tyrosine phosphatases (*4*). Mutation, overexpression, or functional alteration of these enzymes is involved in many diseases including cancer and immunodeficiency diseases (*2, 5*). Therefore, targeted analysis of tyrosine phosphorylated residues in proteins and elucidation of its biological role in determining protein functions is mandatory in order to understand their contribution to signaling networks and, consequently, to pathological processes.

In the last few years, mass spectrometry (MS) has routinely been used to study and characterize phosphorylation events. Due to the low stoichiometry of this modification, methods such as low-pH SCX (*6*), TiO_2 (*7, 8*) and IMAC (*9*) are used for the enrichment of phosphopeptides from complex proteolytic lysates. Tyrosine phosphorylation is estimated to represent ~0.5% of all human phosphorylation events (*10, 11*) with the majority occurring via serine (90%) or threonine (10%) residues. Therefore, these aforementioned approaches are not well suited for the study of tyrosine phosphorylation.

Several tools are available for the specific analysis of tyrosine phosphorylation. Profiling the global tyrosine phosphorylation state of cells can be done using Src homology 2 (SH-2) domains (*12*), which bind selectively to specific tyrosine phosphorylated sites. However, this approach is limited to those phosphotyrosine proteins that interact with the SH2-containing bait used in the assay. The development of specific and high affinity antibodies against phosphorylated tyrosines provides an interesting alternative for the global analysis of tyrosine phosphorylation (*2*). These antibodies are obtained by immunization of a suitable organism (e.g. mouse) with phosphotyramine or phosphotyrosine bound to carrier proteins such as bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH) (Table 1). Traditionally, an-

ti-phosphotyrosine antibodies have been used to detect tyrosine phosphorylated residues in a protein of interest by Western blotting, although, this approach can lack of specificity. Other applications include enzyme-linked immunosorbent assay (ELISA), immunohistochemisty, cell sorting by flow cytometry and immunofluorescence. In the last few years, these antibodies have been also used for the enrichment of tyrosine phosphorylated proteins followed by mass spectrometry analysis (*13-15*). However, detection of the actual phosphotyrosine residue is hampered by the high presence of unmodified peptides derived from the phosphotyrosine protein and its binding partners. More recently, anti-phosphotyrosine antibodies have been successfully used at the peptide level for the immune affinity purification (IAP) of phosphotyrosine peptides from complex samples such as whole cell lysate digests (*16-18*). Importantly, this approach allows the identification and localization of the exact phosphotyrosine residue in the protein. Here, we describe a robust and simple procedure that allows the identification of hundreds of tyrosine phosphorylated residues by using anti-phosphotyrosine antibodies within 6 hours of mass spectrometry analysis time. When coupled to labeling strategies such as iTRAQ (*19*), SILAC (*20*) or dimethylation (*17*), this technique can be used to extract biologically rich information on tyrosine phosphorylation driven signaling.

Table 1: *The different anti-phosphotyrosine antibodies with their characteristics and the reference of studies where each type is been used. * KLH= keyhole limpet haemocyanin,*

Antibody	Antigen	Produced in	Class	Type	Manufacturer	References
PY99		Mouse	IgG_{2b}	Monoclonal	Santa Cruz Biotechnology	(*19-22*)
PY100	Phosphotyrosine containing peptides (KLH*-conjugated)	Mouse	IgG_{1}	Monoclonal	Cell Signaling Technonolgy	(*13, 16, 18, 21, 23-34*)
4G10	Phosphotyramine (KLH*-conjugated)	Mouse	IgG_{2bk}	Monoclonal	Millipore	(*13, 21, 22, 30, 32, 34-39*)
13F9	Phosphotyrosine (KLH*-conjugated)	Mouse	IgG_{1k}	Monoclonal	Abcam	(*32*)
PT66	Phosphotyrosine (BSA/KLH*-conjugated)	Mouse	IgG_{1}	Monoclonal	Sigma	(*15, 23, 24, 40*)

2. **Materials and methods**

2.1. Cell lysis and sample processing

2.1.1. Cell lysis

The extremely low abundance of tyrosine phosphorylation requires the use of relatively large amounts of input material for IAP. This amount can vary between different studies and is often sample dependent (e.g. culture cell lines or tissues). Several groups have reported the use of 10^7 to 10^9 cells (*16*, *17*, *29*). However, these levels are usually difficult to obtain for *in vivo* materials (e.g. primary cells, and tissues). The experiments described here are performed using 5 x 10^7 HeLa cells that were treated with pervanadate, a generic tyrosine phosphatase inhibitor, as described earlier (*17*). Treating HeLa cells with pervanadate ensures maximum levels of protein tyrosine phosphorylation, therefore making the cells suitable for testing the efficiency of the IAP methodology described here. Cells are first lysed in 2 mL of a buffer containing 8 M urea in 50 mM ammonium bicarbonate with protease and phosphatase inhibitors (mini complete EDTA-free) (Roche Diagnostics, Germany) (Figure 1). Protease and phosphatase inhibitors are necessary at this stage to prevent, respectively, unspecific proteolysis and dephosphorylation of proteins during the lysis procedure. Furthermore, dephosphorylation of proteins during sample preparation can compromise the accuracy and precision in quantitative analyses when phosphatase inhibitors are not included during the lysis procedure. Due to the large amount of cells used in this technique, sonication of the sample in the lysis buffer is often required to increase the protein yield and fragment the DNA strands. The lysate is vortexed vigorously and centrifuged at 14,000 x *g* for 10 min at 4°C to separate the pellet of cellular debris, including insoluble membrane and DNA, from the soluble fraction. Protein amount is determined with a Bradford assay (Bio-Rad Laboratories, USA). Using this amount of cells and, under these lysis conditions, around 5 mg of total protein yield is expected.

2.1.2. Proteolysis of the sample

Prior to the enzymatic digestion, proteins are reduced with DTT (final concentration = 2 mM) for 30 minutes at 56°C. The reduced sulfhydryl groups are then alkylated with iodacetamide (final concentration = 4 mM) at room

temperature, in the darkness for 20 minutes to prevent reformation of disulfide bonds. Subsequently, proteins in the cell lysate are enzymatically digested (Figure 1). Efficient digestion is critical for the success of the IAP and, in our hands, sequential digestion with Lys-C and trypsin gives the best results. Lys-C has the advantage of maintaining its activity in 8M urea in which the vast majority of proteins becomes unfolded. Pre-digestion with Lys-C (Wako, USA) (1:85 enzyme:substrate ratio) for 4 h at 37^{O} C is usually enough. The drawback of Lys-C is that it only cleaves at the carboxyl side of lysines producing peptides that are too long for standard MS/MS sequencing. Therefore, a subsequent trypsin digestion step is performed. To this end, the urea concentration is diluted to 2 M with ammonium bicarbonate, and trypsin (Promega, USA) is added at 1:100 enzyme:substrate ratio. The mixture is incubated overnight at 37°C. After digestion, we recommend running an aliquot of the sample on SDS-PAGE and performing a routine LC-MS/MS analysis to ensure complete digestion. Trypsin enzymatic activity is finally quenched by acidification of the sample with formic acid (5% final concentration) to pH 2-3.

2.1.3. Peptide desalting

Prior to IAP, peptides must be desalted in order to remove any compound present in the lysate which could hamper the antigen-antibody affinity efficiency. Commercially available Sep-Pak tC18 3 cc/200 mg cartridges (Waters, USA) are suitable for this purpose as they have a capacity to bind several mg of peptides. The C18 material is conditioned with 5 mL (5 times the column bed volume) of 100% acetonitrile and equilibrated with 5 mL of 0.05% acetic acid. The sample is then loaded at a relative low flow to improve peptide binding and the column is washed twice with 5 mL of 0.05% acetic acid. Finally, the peptides are eluted with 5 mL of 80 % acetonitrile to ensure complete peptide release and the eluate lyophilized *in-vacuo*.

2.2. Immuno-affinity based enrichment of tyrosine phosphorylated peptides

2.2.1. Immuno-affinity purification

Antibody-antigen binding is more efficient in aqueous buffers at physio-

logical pH and ionic strength. Consequently, the lyophilized peptides are re-suspended in 800 μL of cold IAP buffer which consists of 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, protease inhibitors (Roche Diagnostics, Germany) and 1% n-octyl-β-D-glucopyranoside (NOG) (Sigma, Germany). The use of a non-ionic detergent, such as NOG, has been reported to increase the selectivity of tyrosine phosphorylated peptides (*40*). The concentration of NOG should be above its critical micelle concentration, which in this case is ~ 0.7% (*40*). The peptide mixture is agitated on a shaker for 30 minutes to dissolve the peptides thoroughly. We have observed that the pH of the sample is critical for an efficient binding, therefore, the pH of the peptide mixture is checked using pH-indicator strips and adjusted if necessary to pH 7.4 (Figure 1).

A variety of anti-phosphotyrosine antibodies are commercially available (Table 1). The three most commonly used are PY99 (Santa Cruz biotechnology; CA USA), 4G10 (Millipore; USA) and PY100 (Cell Signalling Technology; USA), which have all been successfully applied for the IAP of tyrosine phosphorylated peptides (Table 1). Most of these antibodies are supplied covalently linked to agarose beads. This is especially convenient as the elution conditions could release antibody molecules from the beads making the downstream MS analysis of the phosphotyrosine peptides difficult. In our laboratory, we use the PY99 antibody and we find it to be both sensitive and specific. The amount of PY99 antibody beads should be optimized for each application as this can significantly affect the outcome of the analysis. The experiment described here is performed with 45-50 μL slurry of beads which we found to be optimal when 5 mg of protein input are used. Before adding the peptide solution the beads must be washed thoroughly to remove glycerol and other chemical compounds present in the provided vial. This is done by adding 800 μL of cold IAP buffer and inverting the tube several times, followed by centrifugation at 1,500 x *g* for 1 minute at 4°C. Supernatant is then discarded and this step is repeated 2-3 times.

Peptides are then added to the beads and gently shaken to obtain a homogeneous suspension. The peptide-antibody mixture is incubated overnight at 4°C on a rotator. Shorter incubation times can be considered if performed at room temperature, nevertheless a higher background of non-phosphorylated peptides might be obtained. The next day, the mixture is centrifuged at 1,500 x *g* for 1 minute at 4°C and the supernatant containing the unbound peptides (mainly non-phosphorylated peptides) is collected. The beads are

washed three times with 800 μL of cold IAP buffer as described above. The washing buffer is cautiously removed after each washing step to prevent the loss of beads, and this is easily achieved using GELoader pipette tips (Eppendorf, Germany). The beads are washed twice with 500 μL of milli-Q water. The elution of the bound peptides is performed using a strong acid such as trifluoroacetic acid (TFA) (Sigma, Germany) which disrupts antigen-antibody interactions. 50 μL of 0.15% TFA are added to the beads and incubated for 10 minutes at room temperature followed by centrifugation at 1,500 x *g* for 1 minute. The supernatant is transferred to a fresh tube and a second elution is carried out as before. Both eluates are finally combined in a single tube.

2.2.2. Concentration of peptides

Peptides must be desalted and concentrated prior to the LC-MS/MS analysis. Due to the extremely low amount of peptides at this point of the procedure, the use of small C18 devices instead of large SepPak cartridges is advised. We usually prepare home-made columns with C18 material (Aqua™ C_{18}, 5 μm, Phenomenex, Torrance, CA) at the restricted end of a GELoader tip, as described elsewhere (*41*), but commercially available ZipTips (Waters, USA) can also be used for this purpose. This will also prevent any beads to be carried over into the LC/MS system as they will be retained on top of the column. The column is first pre-conditioned with 20 μL of 100% acetonitrile followed by equilibration with 20 μL of 0.05% acetic acid. The peptides are slowly loaded onto the column and washed twice with 20 μL of 0.05% acetic acid. Elution is accomplished with 20 μL of 80% acetonitrile/ 0.05% acetic acid, repeated twice and combined. Peptides are finally lyophilized to remove the organic component and re-suspended in 40 μL of 10% formic acid (Figure 1). At this point, we recommend the immediate analysis of the samples by LC-MS/MS to avoid any loss due to unspecific absorbance of peptides to the tube.

2.3. Mass Spectrometric analysis

2.3.1. *High resolution LC-MS/MS*

The enriched phosphotyrosine peptides are analyzed by nanoflow LC-

MS/MS (Figure 1) using an Agilent 1100 HPLC system (Agilent Technologies, Waldbronn, Germany) coupled to an LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). 20 μL of the sample are injected in the system and peptides are delivered to a trap column (Aqua™ C_{18}, 5 μm (Phenomenex, Torrance, CA); 20 mm x 100 μm inner diameter) at 5 μL/min in 100% solvent A (0.1 M acetic acid in water). Subsequently, the peptides are eluted from the trap column onto an analytical column (Repro-Sil-Pur C_{18}-AQ, 3 μm (Dr. Maisch GmbH, Ammerbuch, Germany); 40 cm x 50-μm inner diameter) at 100 nL/min in a 3 h gradient from 0 to 50% solvent B (0.1 M acetic acid in 8:2 (v/v) acetonitrile/water). All columns used in the HPLC system are packed in-house. The eluent is sprayed via distal coated emitter tips (New Objective, Cambridge, MA) (o.d., 360 μm; i.d., 20 μm; tip i.d., 10 μm) butt-connected to the analytical column. The tip is subjected to 1.7 kV. A 33 MΩ resistor is introduced between the high voltage supply and the electrospray needle to reduce ion current. The mass spectrometer is operated in data-dependent mode, automatically switching between MS and MS/MS. Full-scan MS spectra (from *m*/*z* 300 to 1,500) are acquired in the Orbitrap with a resolution of 60,000 at *m*/*z* 400 after accumulation to a target value of 500,000 in the linear ion trap. The ten most intense ions above the threshold of 500 counts are selected for collision-induced fragmentation in the linear ion trap at normalized collision energy of 35% after accumulation to a target value of 10,000. A technical replicate is performed with the remaining 20 μL of the sample under identical conditions.

2.3.2. Data Analysis

Raw files are processed with Proteome Discoverer (version 1.3.096, Thermo Fisher Scientific, Bremen, Germany). Database search is performed using a concatenated forward-decoy UniProt human database (v2010-12; 41,008 sequences) using Mascot search engine (version 2.3.02, Matrix Science, UK). Carbamidomethylation of cysteines is set as fixed modification. Oxidized methionine and phosphorylation of tyrosine, serine, and threonine are set as variable modifications. Trypsin is specified as the proteolytic enzyme, and up to two missed cleavages are allowed. The mass tolerance of the precursor ion is initially set to 50 ppm, and 0.6 Da for fragment ions. Mascot results are filtered afterwards with a 10 ppm precursor mass tolerance, a Mascot Ion Score > 20 and a minimum of 6 residues per peptide which results

in a peptide FDR < 1%. Functional classification is performed with PANTHER (www.pantherdb.org). Over-represented phosphotyrosine motifs are extracted from the data set using the Motif-X algorithm (http://motif-x.med.harvard.edu/motif-x.html) (*42*). Phosphotyrosine containing sequences are aligned around the phosphorylated tyrosine and, if necessary, peptide sequences are elongated according to the author's instructions. The human IPI database is used in this case as the background reference. Relationships between kinases and substrates are predicted with the NetworKIN algorithm as described earlier (http://networkin.info/version_2_0/search.php) (*43*, *44*). The MS data associated with this manuscript can be downloaded from ProteomeCommons.org under the following Tranche hash:u08L0Q2fyd05DsyQ3ug3Ua7rvoSVUn34Cke/1MMQ5mZ5zXnaqs4Grb/BO8ieIUEL46W3g6kpmyhCnHBPOUF5FiDagtQAAAAAAAAEGA==; with the passphrase: o0FYl75a7LInKvFhPCVy

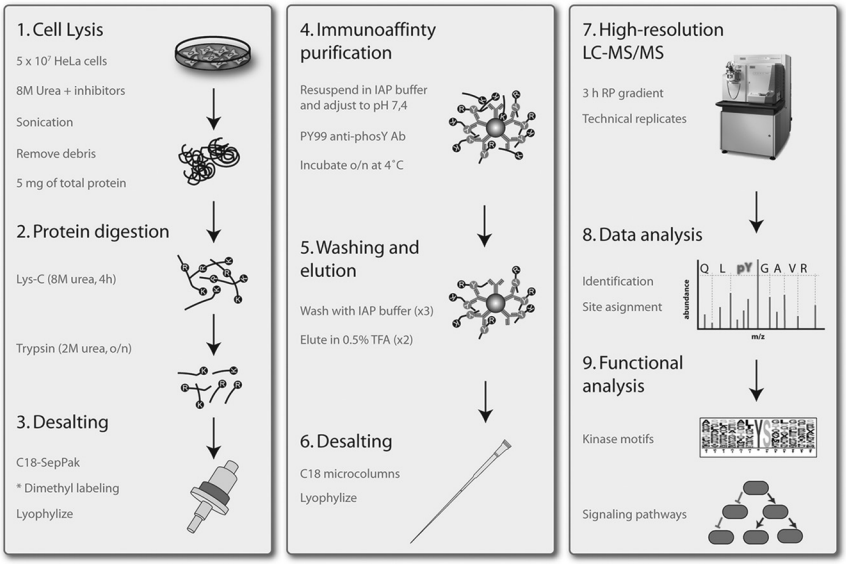


Figure 1: *Experimental design. (1) Cells are harvested (metabolic labeling (e.g. SILAC (20)) can be applied before this stage), and lysed in a lysis buffer containing 8 M urea and protease and phosphatase inhibitors. Protein concentration is measured by Bradford assay and (2) ~5 mg proteins are digested with Lys-C and trypsin. (3) Peptides are desalted and concentrated. At this point, chemical labeling (e.g. dimethylation (17), iTRAQ (30, 45)) can be applied. (4) Peptides are re-suspended in the IAP buffer and tyrosine phosphorylated peptides enrichment is performed using PY99 antibodies conjugated to agarose beads. After washing, tyrosine phosphorylated peptides are eluted in the presence of TFA (5) and finally, the eluate is desalted and concentrated (6). Samples are analyzed by high resolution MS on a three hours gradient (7). Peptides and proteins are identified by database search (8), and functional analysis (9) is performed.*

3. **Results and discussion**

Building on recent developments (*16*, *17*, *19*, *46*), we describe here a technique that enables the profiling of several hundred phosphotyrosine peptides from complex samples. As a case study, we used HeLa cells and, after MS analysis, 26,053 MS/MS spectra were collected from which we identified 4,387 peptide spectrum matches (PSMs) resulting in a global peptide false discovery rate of 0.2%. The peptide sequences matched a total of 1,072 unique protein groups in the UniProt human database. Importantly, the identified proteins included 735 tyrosine phosphorylated proteins representing 1,258 unique tyrosine phosphorylated peptides (Table S1). We further processed the data to localize the phosphorylation sites within the peptide sequence using the Phospho-RS algorithm implemented in Proteome Discoverer software (Thermo Fisher Scientific, Bremen, Germany) (see Methods section). We could assign 1,000 phosphotyrosine sites with a probability higher than 90% (Table S1). These results clearly show the high level of enrichment and specificity of the immune affinity enrichment technique: 68% of the identified proteins were found to be tyrosine phosphorylated in two LC-MS/MS runs. This is a significant reduction in the mass spectrometric analysis time when compared with global phosphoproteomic studies (e.g. SCX-TiO_2) where hundreds of runs are usually performed and tyrosine phosphorylated proteins rarely represent 1-5% of the total identified phosphoproteins (*47*). To further characterize the identified tyrosine phosphorylated proteins, we classified our data set by molecular function, biological process and pathways (data not shown). When compared to the whole set of human genes (background set), as anticipated, we found that the tyrosine phosphorylated proteins are enriched in GO terms related to intracellular signaling cascade (p=1.65E-16, Chi-square test), cell cycle (p=2.23E-15), cell communication (p=1.55E-11) and mitosis (p=4.58E-11). In addition, we found 64 kinases to be tyrosine phosphorylated (p=1.81E-15) pointing to a complex regulatory mechanism of kinase activity which is in agreement with previous reports (*48*). Taken together, these results indicate that the here described IAP protocol is a valuable approach for those interested in the areas of phosphotyrosine signaling.

To gain further insights into the upstream kinases of the detected tyrosine phosphorylated substrates, we used several bioinformatics approaches. Only the unambiguously localized phosphotyrosine sites are used for this

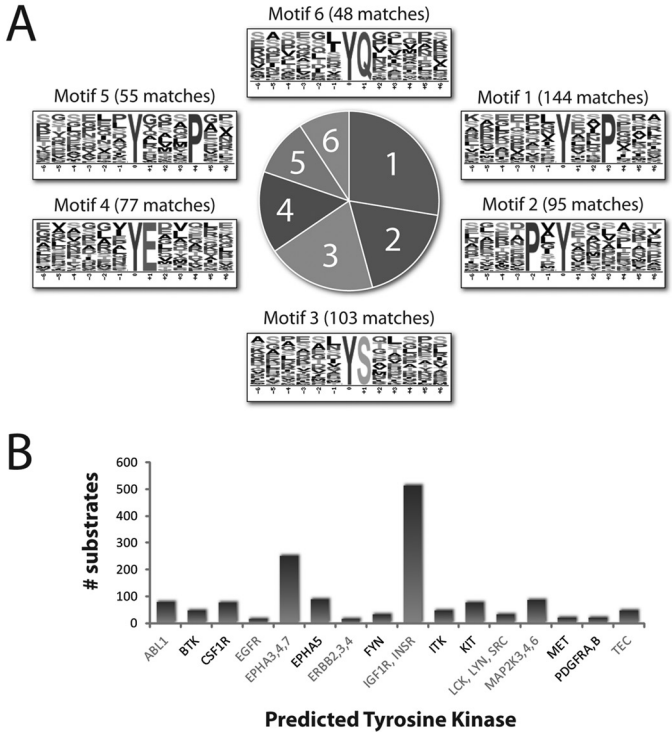


Figure 2: *A) Motif-X (42) analysis identified six over-represented motifs in our HeLa phosphotyrosine dataset. The number of peptides matching a particular motif in the data set is also shown. B) The NetworKIN algorithm (43) was used to predict upstream kinases. The bar chart represents the number of substrates for each predicted kinase. Those kinases found in this study to be tyrosine phosphorylated are displayed in light grey.*

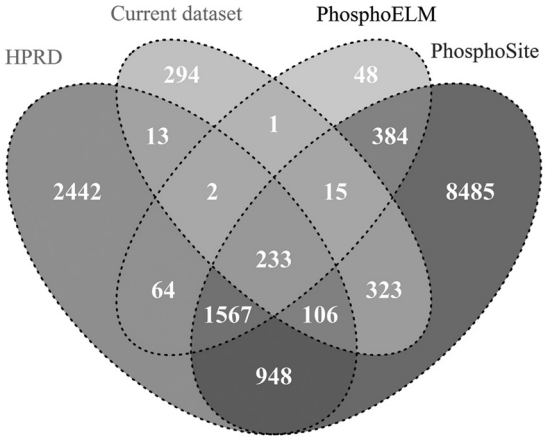


Figure 3: *Venn diagram representing the overlap between the three major databases on human tyrosine phosphorylation sites: HRPD, PhosphoSitePlus and phospho.ELM and our own pervanadate treated HeLa dataset.*

purpose (Phospho-RS probability > 90%, Table S1). First, over-represented motifs present in our data set are extracted using the MOTIF-X algorithm (*42*). We found 6 phosphotyrosine motifs significantly enriched (Figure 2A). Three of which (Motif 1, 3 and 4) have been reported previously (*17, 42*), suggesting predominant tyrosine kinase activities. Half of the motifs contain a proline in positions +2, -3 and -4 whereas the other three contain serine, glutamic acid or glutamine in position +1. We then used a more extensive tool, NetworKIN (*43*) to reconstruct kinase-substrate relationships in our data set. In this approach, upstream kinases are predicted for experimentally identified phosphosites using known linear motifs in combination with contextual information retrieved from databases such as String (*49*). Overall, 16 tyrosine kinase groups are predicted to regulate the phosphotyrosine substrates identified in our study (Figure 2B). Interestingly, 50% of these predicted tyrosine kinases are experimentally identified by MS in our data (Table S1). Amongst them, two groups of RTKs involved in insulin (INSR and IGF1R) and epidermal growth factor (EFGR and ERBBs) signaling. In addition, several non-receptor tyrosine kinases such as Src family (LCK, LYN, and SRC) and ABL1 are also predicted for several substrates. In many cases, we found that the algorithm assigns multiple kinases to the same site. This ambiguity is caused by a lack of knowledge in either the linear motifs and/or contextual information. The results derived from these computational analyses indicate that our IAP protocol can be used as a valuable tool to further investigate tyrosine kinase activities which, ultimately, will serve to design more accurate and precise predictive algorithms.

The described method represents the standard procedure used in many laboratories (*16, 17, 19, 32*) to enrich in tyrosine phosphorylated peptides from complex samples, although some of the key aspects in the workflow should be carefully considered when performing such experiments. We have found important differences when analyzing samples from different biological origins (e.g. human tissues, platelets, cell lines). Here, we have used pervanadate treated HeLa cells and we identified more than 1,200 phosphotyrosine peptides. For instance, Chronical Myelogenous Leukemia (CML) cell lines are known to have higher tyrosine phosphorylation levels than other cell lines in culture (*32*). Furthermore, activation of signaling pathways may enhance the amount of phosphorylated peptides due to the rapid activation of kinases, including tyrosine kinases, which will phosphorylate hundreds of sites in a time-dependent manner. For instance, we have seen that stim-

ulation of human embryonic stem cells (hESCs) with FGF-2 (a growth factor necessary to maintain the undifferentiated state) increases the number of detected phosphotyrosine peptides by a factor of two when compared to hESCs deprived of FGF-2 (*46*). When studying signaling networks, a quantitative approach that allows accurate measurements of relative changes in phosphorylation must be introduced in the workflow scheme. Although SILAC (*20*) and iTRAQ (*45*) can be applied, we use routinely a chemical labeling strategy that employs stable isotope dimethyl labels and allows up to three conditions to be compared (*17*). Using dimethyl labeling in combination with phosphotyrosine IAP, we have recently profiled changes in tyrosine phosphorylation upon FGF-2 stimulation of hESCs (*46*). We observed trans-activation of multiple receptors as well as numerous members of PI3-K, MAPK and Src family members pointing to cytoskeletal remodeling processes in the maintenance of the undifferentiated and pluripotent state (*46*). The dimethyl labeling is especially convenient since metabolic labeling strategies like SILAC (*20*) or isobaric tags (e.g. iTRAQ and TMT) (*45, 50*) are cost-prohibitive due to the large amount of protein input necessary for the technique. In addition, dimethyl labeling could also be applied in the analysis of tissue samples such as tumor biopsies of cancer patients (*18*) where metabolic labeling is not easily achievable.

The protocol discussed here uses a double enzymatic digestion with Lys-C and trypsin which enables a more efficient digestion of plasma membrane proteins such us RTKs (*51*). Alternatively, other proteases could be introduced in the experimental set up which is especially beneficial in phosphoproteomics studies due to their complementarity. For instance, Rush *et al.* reported the use of chymotrypsin, Glu-C and elastase, resulting in a significant increase in the number of identified phosphotyrosine peptides when compared with trypsin alone (*16*). Recently, we have shown that the metalloendopeptidease Lys-N (*52*) enables the identification of a different set of phosphopeptides in global phosphoproteomics studies (*52*) demonstrating its complementarity to other enzymes and making Lys-N suitable for IAP studies as well.

In Table 1, some of the most commonly used antibodies to profile tyrosine phosphorylation are listed. Although we obtained the best results with PY99, a combination of different antibodies could be considered to increase the number of identified phosphotyrosine peptides (*13, 30*). Besides the monoclonal nature of these antibodies, occasionally, we have found differ-

ent efficiencies between the supplied batches. To control for this batch-to-batch variation, we prepared several synthetic phosphotyrosine peptides that contain some of the motifs discussed above. These peptides can be used as a positive control to evaluate the quality of the antibodies and to normalize and compare samples of different biological nature. Normally, ~40 fmol of the synthetic phosphotyrosine peptides are spiked in the sample prior the IAP procedure, and the efficiency of the technique is evaluated by comparing the extracted ion chromatograms with a standard containing an equal amount of synthetic phosphotyrosine peptides.

The large increase in the number of phosphotyrosine residues identified when performing two technical replicates (~30%) reflects the high complexity of the sample. The flow-through fraction (unbound peptides) can be re-processed in the same manner with fresh antibodies in a second re-IAP (data not shown). This double-IAP significantly increases the total number of phosphotyrosine peptides suggesting that many sites remain to be identified and also, highlighting the large dynamic range of this modification in mammalian cells. To overcome this large dynamic range, additional strategies could be incorporated into the workflow before and after the IAP, with the aim of improving sensitivity and specificity of the technique. These strategies are mainly designed to reduce the number of background (i.e. unmodified) peptides present in the eluates. Prior to the IAP, samples could be precleared with beads containing another unspecific antibody (e.g. anti-IgG from the same species i.e. mouse) and subsequently incubated with the anti-phosphotyrosine beads. Another option is to reduce the complexity of the sample by pre-fractionation strategies. Methods such as reverse-phase solid-phase extraction have been used in combination with phosphorylated tyrosine IAP (*16*). With a similar reasoning, we proposed the use of low-pH SCX (*6*) to pre-enrich in fractions containing phosphopeptides which can be further enriched in tyrosine phosphorylated peptides with the IAP-based procedures described here. After the IAP, metal affinity-based chromatography such as IMAC (*29, 45*) has been used as an efficient way to remove the non-phosphorylated peptides from the IAP eluates prior to MS analysis. The availability of alternative activation techniques such as electron transfer dissociation (ETD) (*53*) and higher collision C-trap dissociation (HCD) (*54*), in modern mass spectrometers allows now the use of optimized fragmentation designs to improve peptide identification (*55*). In our data set we have seen that the identified tyrosine phosphorylated peptides contain more

missed cleavages and consequently, a higher charge state than regular unmodified tryptic peptides (data not shown), making phosphotyrosine peptides suitable to ETD fragmentation (*55*). In addition, the high resolution at the MS/MS level of HCD will also improve site localization over CID spectra (*56*).

In the last decade, we have witnessed an unprecedented explosion in the amount of phosphorylation sites identified by MS-based phosphoproteomic studies (*11*), which has made necessary the generation of dedicated databases to manage, share and annotate these results. Human Protein Reference Database (HPRD) (*57*), PhosphoSitePlus, and Phopsho.ELM are the three repositories containing the largest number of tyrosine phosphorylated residues. For human proteins, HPRD (latest update 2010) includes 5,375 tyrosine phosphorylation sites, phospho.ELM around 2,314 sites (latest update 2010) and the largest number of phosphotyrosine sites (12,061) are hosted by PhosphoSitePlus (Figure 3). We have compared the identified phosphotyrosines in our HeLa data set with these databases and, remarkably, found that 294 (30% of our data) are novel sites suggesting that many phosphotyrosine sites remain to be identified (Figure 3). Overall, 14,925 unique phosphotyrosine residues are included in this comparison which belong to 5,866 proteins illustrating the large landscape of this modification. Besides the fact that many of these sites could be functionally irrelevant (*58*) as a consequence of promiscuous kinase activities in evolutionary mechanisms (*59*), we are still far from comprehensively understanding the complex nature of the phosphoproteome as the largest study to date has reported more than 30,000 unique phosphosites in mouse tissues (*60*).

4. **Concluding remarks**

Here, we describe in detail an IAP method that allows the identification of hundreds of phosphotyrosine sites from complex cell lysates in just a few hours of LC-MS/MS analysis. The workflow is flexible and several modifications to this scheme can be introduced. In conjunction with a quantitative strategy such as dimethyl labeling, this technique can be used to profile temporal dynamics of tyrosine phosphorylation in response to a biological stimulus. It is our hope that the knowledge acquired from tyrosine phosphorylation studies will gradually facilitate the assimilations in cellular signaling

networks and the characterization of biological responses of the cell to specific signals in physiological conditions and diseases.

5. Acknowledgements

We thank Pieter Neerincx for bioinformatics support and Christian Preisinger and Onno Bleijerveld and others members from the Heck Lab for fruitful discussions. We also thank Vanessa Ding and Andre Choo from The Stem Cells Group at the Bioprocessing Technology Institute in Singapore for sharing experiences on the protocol. This work was supported by the Netherlands Proteomics Centre, which is part of the Netherlands Genomics Initiative. This work was in parts supported by the PRIME-XS project, grant agreement number 262067, funded by the European Union 7th Framework Programme.

Supplementary data is available at **http://tinyurl.com/azdchapter3**

6. References

[1] W. Eckhart, M. A. Hutchinson, T. Hunter, An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates, Cell **18** (1979) 925-933.

[2] T. Hunter, 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** (1998) 583-605.

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

[4] A. Alonso, J. Sasin, N. Bottini, I. Friedberg, A. Osterman, A. Godzik, T. Hunter, J. Dixon, T. Mustelin, Protein tyrosine phosphatases in the human genome, Cell **117** (2004) 699-711.

[5] T. Hunter, Signaling--2000 and beyond, Cell **100** (2000) 113-127.

[6] S. A. Beausoleil, M. Jedrychowski, D. Schwartz, J. E. Elias, J. Villen, J. Li, M. A. Cohn, L. C. Cantley, S. P. Gygi, Large-scale characterization of HeLa cell nuclear phosphoproteins, Proc Natl Acad Sci U S A **101** (2004) 12130-12135.

[7] M. R. Larsen, T. E. Thingholm, O. N. Jensen, P. Roepstorff, T. J. Jor-

gensen, Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns, Molecular & cellular proteomics : MCP **4** (2005) 873-886.

[8] M. W. Pinkse, P. M. Uitto, M. J. Hilhorst, B. Ooms, A. J. Heck, Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns, Anal Chem **76** (2004) 3935-3943.

[9] L. Andersson, J. Porath, Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography, Anal Biochem **154** (1986) 250-254.

[10] T. Hunter, B. M. Sefton, Transforming gene product of Rous sarcoma virus phosphorylates tyrosine, Proceedings of the National Academy of Sciences of the United States of America **77** (1980) 1311-1315.

[11] S. Lemeer, A. J. Heck, The phosphoproteomics data explosion, Curr Opin Chem Biol **13** (2009) 414-420.

[12] K. Machida, C. M. Thompson, K. Dierck, K. Jablonowski, S. Karkkainen, B. Liu, H. Zhang, P. D. Nash, D. K. Newman, P. Nollau, T. Pawson, G. H. Renkema, K. Saksela, M. R. Schiller, D. G. Shin, B. J. Mayer, High-throughput phosphotyrosine profiling using SH2 domains, Mol Cell **26** (2007) 899-915.

[13] B. Blagoev, S. E. Ong, I. Kratchmarova, M. Mann, Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics, Nat Biotechnol **22** (2004) 1139-1145.

[14] A. Pandey, A. V. Podtelejnikov, B. Blagoev, X. R. Bustelo, M. Mann, H. F. Lodish, Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors, Proc Natl Acad Sci U S A **97** (2000) 179-184.

[15] A. R. Salomon, S. B. Ficarro, L. M. Brill, A. Brinker, Q. T. Phung, C. Ericson, K. Sauer, A. Brock, D. M. Horn, P. G. Schultz, E. C. Peters, Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry, Proc Natl Acad Sci U S A **100** (2003) 443-448.

[16] J. Rush, A. Moritz, K. A. Lee, A. Guo, V. L. Goss, E. J. Spek, H. Zhang, X. M. Zha, R. D. Polakiewicz, M. J. Comb, Immunoaffinity profiling of tyrosine phosphorylation in cancer cells, Nat Biotechnol **23** (2005) 94-101.

[17] P. J. Boersema, L. Y. Foong, V. M. Ding, S. Lemeer, B. van Breukelen, R. Philp, J. Boekhorst, B. Snel, J. den Hertog, A. B. Choo, A. J. Heck, In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling, Mol Cell Proteomics **9** 84-99.

[18] K. Rikova, A. Guo, Q. Zeng, A. Possemato, J. Yu, H. Haack, J. Nardone, K. Lee, C. Reeves, Y. Li, Y. Hu, Z. Tan, M. Stokes, L. Sullivan, J. Mitchell, R. Wetzel, J. Macneill, J. M. Ren, J. Yuan, C. E. Bakalarski, J. Villen, J. M. Kornhauser, B. Smith, D. Li, X. Zhou, S. P. Gygi, T. L. Gu, R. D. Polakiewicz, J. Rush, M. J. Comb, Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer, Cell **131** (2007) 1190-1203.

[19] Y. Zhang, A. Wolf-Yadlin, P. L. Ross, D. J. Pappin, J. Rush, D. A. Lauffenburger, F. M. White, Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules, Mol Cell Proteomics **4** (2005) 1240-1250.

[20] G. Zhang, D. S. Spellman, E. Y. Skolnik, T. A. Neubert, Quantitative phosphotyrosine proteomics of EphB2 signaling by stable isotope labeling with amino acids in cell culture (SILAC), J Proteome Res **5** (2006) 581-588.

[21] A. Wolf-Yadlin, N. Kumar, Y. Zhang, S. Hautaniemi, M. Zaman, H. D. Kim, V. Grantcharova, D. A. Lauffenburger, F. M. White, Effects of HER2 overexpression on cell signaling networks governing proliferation and migration, Mol Syst Biol **2** (2006) 54.

[22] A. M. Hinsby, J. V. Olsen, K. L. Bennett, M. Mann, Signaling initiated by overexpression of the fibroblast growth factor receptor-1 investigated by mass spectrometry, Mol Cell Proteomics **2** (2003) 29-36.

[23] K. Schmelzle, S. Kane, S. Gridley, G. E. Lienhard, F. M. White, Temporal dynamics of tyrosine phosphorylation in insulin signaling, Diabetes **55** (2006) 2171-2179.

[24] J. E. Kim, F. M. White, Quantitative analysis of phosphotyrosine signaling networks triggered by CD3 and CD28 costimulation in Jurkat cells, J Immunol **176** (2006) 2833-2843.

[25] T. L. Gu, V. L. Goss, C. Reeves, L. Popova, J. Nardone, J. Macneill,

D. K. Walters, Y. Wang, J. Rush, M. J. Comb, B. J. Druker, R. D. Polakiewicz, Phosphotyrosine profiling identifies the KG-1 cell line as a model for the study of FGFR1 fusions in acute myeloid leukemia, Blood **108** (2006) 4202-4204.

[26] P. H. Huang, A. Mukasa, R. Bonavia, R. A. Flynn, Z. E. Brewer, W. K. Cavenee, F. B. Furnari, F. M. White, Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma, Proc Natl Acad Sci U S A **104** (2007) 12867-12872.

[27] A. Guo, J. Villen, J. Kornhauser, K. A. Lee, M. P. Stokes, K. Rikova, A. Possemato, J. Nardone, G. Innocenti, R. Wetzel, Y. Wang, J. MacNeill, J. Mitchell, S. P. Gygi, J. Rush, R. D. Polakiewicz, M. J. Comb, Signaling networks assembled by oncogenic EGFR and c-Met, Proc Natl Acad Sci U S A **105** (2008) 692-697.

[28] F. E. Boccalatte, C. Voena, C. Riganti, A. Bosia, L. D'Amico, L. Riera, M. Cheng, B. Ruggeri, O. N. Jensen, V. L. Goss, K. Lee, J. Nardone, J. Rush, R. D. Polakiewicz, M. J. Comb, R. Chiarle, G. Inghirami, The enzymatic activity of 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase is enhanced by NPM-ALK: new insights in ALK-mediated pathogenesis and the treatment of ALCL, Blood **113** (2009) 2776-2790.

[29] L. K. Iwai, C. Benoist, D. Mathis, F. M. White, Quantitative phosphoproteomic analysis of T cell receptor signaling in diabetes prone and resistant mice, J Proteome Res **9** 3135-3145.

[30] P. H. Huang, E. R. Miraldi, A. M. Xu, V. A. Kundukulam, A. M. Del Rosario, R. A. Flynn, W. K. Cavenee, F. B. Furnari, F. M. White, Phosphotyrosine signaling analysis of site-specific mutations on EGFRvIII identifies determinants governing glioblastoma cell growth, Mol Biosyst **6** 1227-1237.

[31] Y. Zhang, Y. Yoshida, M. Nameta, B. Xu, I. Taguchi, T. Ikeda, H. Fujinaka, S. M. Mohamed, H. Tsukaguchi, Y. Harita, E. Yaoita, T. Yamamoto, Glomerular proteins related to slit diaphragm and matrix adhesion in the foot processes are highly tyrosine phosphorylated in the normal rat kidney, Nephrol Dial Transplant **25** 1785-1795.

[32] S. Bergstrom Lind, M. Molin, M. M. Savitski, L. Emilsson, J. Astrom, L. Hedberg, C. Adams, M. L. Nielsen, A. Engstrom, L. Elfineh, E. Andersson, R. A. Zubarev, U. Pettersson, Immunoaffinity enrich-

ments followed by mass spectrometric detection for studying global protein tyrosine phosphorylation, J Proteome Res **7** (2008) 2897-2910.

[33] J. Tong, P. Taylor, E. Jovceva, J. R. St-Germain, L. L. Jin, A. Nikolic, X. Gu, Z. H. Li, S. Trudel, M. F. Moran, Tandem immunoprecipitation of phosphotyrosine-mass spectrometry (TIPY-MS) indicates C19ORF19 becomes tyrosine-phosphorylated and associated with activated epidermal growth factor receptor, J Proteome Res **7** (2008) 1067-1077.

[34] M. Kruger, I. Kratchmarova, B. Blagoev, Y. H. Tseng, C. R. Kahn, M. Mann, Dissection of the insulin signaling pathway via quantitative phosphoproteomics, Proc Natl Acad Sci U S A **105** (2008) 2451-2456.

[35] W. X. Schulze, L. Deng, M. Mann, Phosphotyrosine interactome of the ErbB-receptor kinase family, Mol Syst Biol **1** (2005) 2005 0008.

[36] H. Steen, A. Pandey, J. S. Andersen, M. Mann, Analysis of tyrosine phosphorylation sites in signaling molecules by a phosphotyrosine-specific immonium ion scanning method, Sci STKE **2002** (2002) pl16.

[37] H. Steen, B. Kuster, M. Fernandez, A. Pandey, M. Mann, Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode, Anal Chem **73** (2001) 1440-1448.

[38] P. Mertins, H. C. Eberl, J. Renkawitz, J. V. Olsen, M. L. Tremblay, M. Mann, A. Ullrich, H. Daub, Investigation of protein-tyrosine phosphatase 1B function by quantitative proteomics, Mol Cell Proteomics **7** (2008) 1763-1777.

[39] J. R. Wisniewski, N. Nagaraj, A. Zougman, F. Gnad, M. Mann, Brain phosphoproteome obtained by a FASP-based method reveals plasma membrane protein topology, J Proteome Res **9** 3280-3289.

[40] G. Zhang, T. A. Neubert, Use of detergents to increase selectivity of immunoprecipitation of tyrosine phosphorylated peptides prior to identification by MALDI quadrupole-TOF MS, Proteomics **6** (2006) 571-578.

[41] J. Gobom, E. Nordhoff, E. Mirgorodskaya, R. Ekman, P. Roepstorff, Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass

spectrometry, J Mass Spectrom **34** (1999) 105-116.

[42] D. Schwartz, S. P. Gygi, An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets, Nature biotechnology **23** (2005) 1391-1398.

[43] R. Linding, L. J. Jensen, A. Pasculescu, M. Olhovsky, K. Colwill, P. Bork, M. B. Yaffe, T. Pawson, NetworKIN: a resource for exploring cellular phosphorylation networks, Nucleic Acids Res **36** (2008) D695-699.

[44] R. Linding, L. J. Jensen, G. J. Ostheimer, M. A. van Vugt, C. Jorgensen, I. M. Miron, F. Diella, K. Colwill, L. Taylor, K. Elder, P. Metalnikov, V. Nguyen, A. Pasculescu, J. Jin, J. G. Park, L. D. Samson, J. R. Woodgett, R. B. Russell, P. Bork, M. B. Yaffe, T. Pawson, Systematic discovery of in vivo phosphorylation networks, Cell **129** (2007) 1415-1426.

[45] A. M. Del Rosario, F. M. White, Quantifying oncogenic phosphotyrosine signaling networks through systems biology, Curr Opin Genet Dev **20** 23-30.

[46] V. M. Ding, P. J. Boersema, L. Y. Foong, C. Preisinger, G. Koh, S. Natarajan, D. Y. Lee, J. Boekhorst, B. Snel, S. Lemeer, A. J. Heck, A. Choo, Tyrosine phosphorylation profiling in FGF-2 stimulated human embryonic stem cells, PLoS One **6** (2011) e17538.

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

[48] H. Daub, J. V. Olsen, M. Bairlein, F. Gnad, F. S. Oppermann, R. Korner, Z. Greff, G. Keri, O. Stemmann, M. Mann, Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle, Mol Cell **31** (2008) 438-448.

[49] L. J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork, C. von Mering, STRING 8--a global view on proteins and their functional interactions in 630 organisms, Nucleic Acids Res **37** (2009) D412-416.

[50] A. Thompson, J. Schafer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, R. Johnstone, A. K. Mohammed, C. Hamon, Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS, Analytical chemistry **75**

(2003) 1895-1904.

[51] W. Dormeyer, D. van Hoof, S. R. Braam, A. J. Heck, C. L. Mummery, J. Krijgsveld, Plasma membrane proteomics of human embryonic stem cells and human embryonal carcinoma cells, Journal of proteome research **7** (2008) 2936-2951.

[52] S. Gauci, A. O. Helbig, M. Slijper, J. Krijgsveld, A. J. Heck, S. Mohammed, Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach, Analytical chemistry **81** (2009) 4493-4501.

[53] D. M. Good, M. Wirtala, G. C. McAlister, J. J. Coon, Performance characteristics of electron transfer dissociation mass spectrometry, Molecular & cellular proteomics : MCP **6** (2007) 1942-1951.

[54] J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, M. Mann, Higher-energy C-trap dissociation for peptide modification analysis, Nat Methods **4** (2007) 709-712.

[55] C. K. Frese, A. F. Altelaar, M. L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A. J. Heck, S. Mohammed, Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos, Journal of proteome research **10** (2011) 2377-2388.

[56] C. D. Kelstrup, O. Hekmat, C. Francavilla, J. V. Olsen, Pinpointing Phosphorylation Sites: Quantitative Filtering and a Novel Site-specific x-Ion Fragment, Journal of proteome research (2011).

[57] S. Mathivanan, B. Periaswamy, T. K. Gandhi, K. Kandasamy, S. Suresh, R. Mohmood, Y. L. Ramachandra, A. Pandey, An evaluation of human protein-protein interaction data in the public domain, BMC Bioinformatics **7 Suppl 5** (2006) S19.

[58] P. Cohen, The role of protein phosphorylation in neural and hormonal control of cellular activity, Nature **296** (1982) 613-620.

[59] C. S. Tan, A. Pasculescu, W. A. Lim, T. Pawson, G. D. Bader, R. Linding, Positive selection of tyrosine loss in metazoan evolution, Science **325** (2009) 1686-1688.

[60] E. L. Huttlin, M. P. Jedrychowski, J. E. Elias, T. Goswami, R. Rad, S. A. Beausoleil, J. Villen, W. Haas, M. E. Sowa, S. P. Gygi, A tissue-specific atlas of mouse protein phosphorylation and expression, Cell **143** (2010) 1174-1189.

Chapter IV

Quantitative comparison of tyrosine phosphorylation levels between hiPSCs and hESCs

Adja D. Zoumaro-Djayoon[1], Yee Jiun Kok[2], Javier Muñoz[1], Andre Choo[2], Albert J. R. Heck[1]

1- Biomolecular Mass Spectrometry and Proteomics. Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH, Utrecht (The Netherlands) and Netherlands Proteomics Center.

2-Stem Cells Group. Bioprocessing Technology Institute, 20 Biopolis Way. Singapore.

Abstract

Human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) are both pluripotent cells that possess the ability to differentiate into multiple specialized cells. However, the exact level of similarity between these pluripotent cell lines at the molecular level remains controversial. Growth factors such as FGF-2 are employed in the culture of these cells to keep them undifferentiated. Often, these growth factors signal through receptor tyrosine kinases, however, little is known about the downstream tyrosine phosphorylation signaling in both cell lines. Here, we used a targeted approach to monitor tyrosine phosphorylated proteins in hiPSCs, and compared them to those in hESCs, using a combination of an immuno-affinity enrichment approach for the enrichment of phosphotyrosine peptides and stable isotope dimethyl labeling for quantitative mass spectrometric analysis. With this highly specific methodology, 31 proteins out of a total of 93 tyrosine phosphorylated proteins showed differential levels of tyrosine phosphorylation between hiPSCs and hESCs, whereby these results could be confirmed in two different hiPS cell lines. These differential tyrosine phosphorylated proteins include proteins involved in the regulation of pluripotency in hESCs such as the SRC family of kinases (e.g. FYN, LYN, LCK), receptor tyrosine kinases (e.g. EPHA2, FGFR1) and cytoskeletal associated proteins (e.g. PKP3, PTK2, paxillin). These results provide a basis for future investigations on the molecular mechanisms driven by tyrosine phosphorylation to control pluripotency in hiPSC and hESC.

1. **Introduction**

HESCs are derived from pre-implantation embryos and have the ability to be cultured indefinitely and to form derivatives of all three germ layers (i.e. ectoderm, mesoderm and endoderm) (*1*). These unique properties of self-renewal and pluripotency made hESCs an attractive option in regenerative medicine as a source of tissue replacement to treat degenerative diseases and other medical conditions. However, the use of embryos to generate hESCs cell lines for clinical purposes is of ethical debate within many communities (*2, 3*). In addition, the use of hESCs in cell therapy has encountered several problems such as the immune rejection of foreign cells and teratoma formation by residual undifferentiated cells (*3*).

The groundbreaking discovery that somatic cells can be reprogrammed to a state of pluripotency has opened new doors for the field of regenerative medicine. The reprogrammed somatic cells, termed induced pluripotent stem cells (iPSCs), are derived by the ectopic expression of specific transcription factors e.g. OCT3/4, SOX2, KLF4, cMyc (*1, 4*). iPSCs are morphologically similar to ESCs, they express genes that hallmark ESCs and can differentiate into the three primary germ layers (*5-7*). Direct reprogramming of patient-specific cells into pluripotent cells will circumvent many of the ethical issues and will allow autologous transplants thereby potentially preventing immune rejection (*8*). Nevertheless, assessing molecular differences between iPSCs and their ESCs counterparts is essential, given that such differences may impact their potential clinical use. To this end, Chin *et al.* analyzed gene expression profiles of ESCs and iPSCs and found that a recurrent gene expression signature appears in iPSCs regardless of their origin or the method by which they are generated (*9*). However, upon extended culture, hiPSCs seem to adopt a gene expression profile that is more similar to hESCs (*9*). Guenther *et al.* compared the global chromatin structure and gene expression profiles of hESCs and hiPSCs (*10*). Their results indicated that nucleosomes with histone H3K4me3 and H3K27me3 modifications, which are involved in gene expression regulation, have little differences between ESCs and iPSCs (*10*). Nevertheless, these differences were not sufficient to distinguish iPSCs and ESCs. In another report, analysis on DNA methylation patterns revealed differentially methylated regions in iPSCs which were enriched in tissue-specific genes suggesting an aberrant DNA epigenetic state (*10*) and the presence of an epigenetic memory from the parental cell donors

(*11*). Additionally, miRNA profiles of hESCs and hiPSCs (*12*) indicated significant differences in the expression of miR-371/372/373 cluster (*12*). More recently, two independent proteomic analyses have been undertaken one being from our group, wherein the proteomes of hiPSCs and hESCs were compared (*13, 14*). Although both groups found minor differences in the expression of some proteins, these changes were not alike between these reports suggesting the presence of lab-specific variations rather than a recurrent molecular signature between hiPSCs and hESCs.

Despite all these genome-wide analyses, the molecular similarity of iPSCs and ESCs is still under debate (*9, 10*). In that respect, protein phosphorylation, which has an important role in signal transduction, represents another level of regulation. Several signaling pathways have been shown to be key for the maintenance of undifferentiated hESCs including Wnt, phosphoinositide 3-kinase/protein kinase B (PI3K/AKT) and mitogen activated kinase (MAPK) pathways (*15-17*). Bearing this in mind, Phanstiel *et al.* compared the phosphoproteomes of hESCs and hiPSCs and found that 292 phosphoproteins differed significantly between these pluripotent cell lines (*13*). Although this data set is the largest analysis of protein phosphorylation conducted in ESCs and iPSCs to date, only 2% of the 4,564 quantified sites were tyrosine phosphorylated residues. Tyrosine signaling is key for self-renewal and pluripotency since many growth factors normally used in the maintenance of hESCs signal through receptor tyrosine kinases (e.g. IGF, FGF-2) (*17-19*). However, the large-scale identification of tyrosine phosphorylation events by mass spectrometry (MS) is usually hindered by the low frequency of these events and, often, low abundance of these proteins (*20*). Therefore, targeted enrichment of tyrosine phosphorylated proteins or peptides prior to mass spectrometric analysis is essential (*20*). Here, we have quantified tyrosine phosphorylation levels in hESCs and hiPSCs using an immuno-affinity purification technique to enrich phosphotyrosine containing peptides (*16*). Using this approach, 190 unique tyrosine phosphorylated peptides could be identified in two independent experiments. Most importantly, 26% of these phosphotyrosine residues showed more than two-fold difference between hiPSCs and hESCs. Our results complement existing phosphorylation analyses on pluripotent cells and may suggest differences in tyrosine signaling between hESCs and hiPSCs.

2. **Materials and methods**

Cell lines and culture conditions

Human ESC line, HES-3 (46,XX) was obtained from ES Cell International (ESI, Singapore, http://escellinternational.com). Briefly, the cells were cultured on Matrigel-coated (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) culture dishes and supplemented with conditioned medium from immortalised mouse embryonic fibroblast (Δ-MEF) (*21*). The medium was supplemented with 10 ng/mL of Fibroblast Growth Factor 2 (FGF-2) (Invitrogen, Carlsbad, CA, USA), and changed daily (*22, 23*).
Human Induced pluripotent stem cells iPS-IMR90 (lung fibroblasts) and iPS-Newborn Foreskin Fibroblasts (NFF), were obtained from James Thomson (*6*). iPS cells were cultured as the hESC cultures, with the exception that 100 ng/mL of FGF-2 was supplemented to the conditioned medium (*6*).

Cell lysis and sample preparation

Cells were harvested and lysed on ice in 8 M urea, 1 mM sodium orthovanadate and 1x *PhosSTOP* (Roche Diagnostics, Switzerland) in the presence of protease inhibitors. Protein concentration was determined using Bradford Assay. Total protein lysate of ~1.5 mg per cell line was reduced with dithiothreitol (DTT) at a final concentration of 10 mM at 56°C and subsequently alkylated with 55 mM iodoacetamide. Lysates were pre-digested with Lys-C (1:75) for 4 hours at 37°C. The samples were diluted 4-fold with 100 mM ammonium bicarbonate and digested overnight with trypsin (1:100) at 37°C.

Dimethyl labeling

Tryptic peptides were first desalted using a Sep-Pak C18 column (Waters, USA, Massachusetts). The eluted peptides were lyophilized, and re-suspended in 100 μL of 100 mM ammonium bicarbonate. Triplex stable isotope dimethyl labeling was performed on column using Sep-pak cartridges as previously described (*24, 25*).

Immuno-affinity purification of tyrosine phosphorylated peptides

Peptides were re-suspended in 800 μL of cold immuno-affinity purification buffer which consists of 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, protease inhibitors (Roche Diagnostics, Germany) and 1% n-octyl-β-D-glucopyranoside (NOG) (Sigma, Germany). The peptide mixture was agitated on a shaker for 30 minutes to dissolve the peptides thoroughly and the pH was adjusted, if necessary, to pH 7.4. Labeled peptides were mixed in 1:1:1 ratio based on total peptide amount, determined by analyzing an aliquot of the labeled samples on a regular LC-MS/MS run and comparing overall peptide signal intensities. The immuno-affinity purification was performed as previously described (*26*). In summary, the peptides were added to 45-50 μL slurry of beads and the peptide-antibody mixture was incubated overnight at 4°C on a rotator. The beads were washed and bound peptides were eluted twice using 0.15% trifluoroacetic acid (TFA) (Sigma, Germany). Home-made columns with C18 material (Aqua™ C_{18}, 5 μm, Phenomenex, Torrance, CA) at the restricted end of a GELoader tip were used for desalting, as described elsewhere (*27*). Peptides were finally lyophilized and re-suspended in 40 μL of 10% formic acid for LC-MS/MS analysis (*26*).

High resolution mass spectrometric analysis

The enriched phosphotyrosine peptides were analyzed by nanoflow LC-MS/MS using an Agilent 1100 HPLC system (Agilent Technologies, Waldbronn, Germany) coupled to an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). 20 μL of the sample were injected in the system and peptides were delivered to a trap column (Aqua™ C_{18}, 5 μm (Phenomenex, Torrance, CA); 20 mm x 100 μm inner diameter) at 5 μL/min in 100% solvent A (0.1 M acetic acid in water). Subsequently, the peptides were eluted from the trap column onto an analytical column (ReproSil-Pur C_{18}-AQ, 3 μm (Dr. Maisch GmbH, Ammerbuch, Germany); 40 cm x 50-μm inner diameter) at 100 nL/min in a 3 h gradient from 0 to 50% solvent B (0.1 M acetic acid in 8:2 (v/v) acetonitrile/water). All columns used in the HPLC system were packed in-house. The eluent was sprayed via distal coated emitter tips (New Objective, Cambridge, MA) (o.d., 360 μm; i.d., 20 μm; tip i.d., 10 μm) butt-connected to the analytical column. The tip was subjected to 1.7 kV. A 33 MΩ resistor was introduced between the high

voltage supply and the electrospray needle to reduce ion current. The mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS. Full-scan MS spectra (from *m/z* 350 to 1,500) were acquired in the Orbitrap with a resolution of 30,000 at *m/z* 400 after accumulation to a target value of 500,000. The ten most intense ions above the threshold of 500 counts were selected for fragmentation. For the fragmentation, a decision tree method was used as previously described (*28*). Briefly, HCD was used for doubly charged peptides and ETD was performed for peptides with more than two charge states. Fragment ions were analyzed in the Orbitrap when the precursor ion has a mass to charge ratio (m/z) less than 1000 and charge state of 4. A technical replicate was performed with the remaining 20 μL of the sample under identical conditions.

Data processing & analysis

Raw files were processed with Proteome Discoverer (version 1.3.096, Thermo Fisher Scientific, Bremen, Germany). The non-fragment filter was used for ETD spectra to filter the precursor peaks from the peak list. Database search was performed using a concatenated forward-decoy IPI human database (v3.52; 148,408 sequences) using Mascot (version 2.3.02, Matrix Science, UK) as the search engine. Carbamidomethylation of cysteines was set as fixed modification. Oxidized methionine, phosphorylation of tyrosine, serine, and threonine and dimethyl labels light, intermediate and heavy on the N-termini and on lysine residues were all set as variable modifications. Trypsin was specified as the proteolytic enzyme, and up to two missed cleavages were allowed. The mass tolerance of the precursor ion was initially set to 50 ppm, and 0.6 Da for fragment ions in ion trap readout and 0.05 Da for fragment ions in Orbitrap readout. Mascot results were filtered afterwards with a 10 ppm precursor mass tolerance, a Mascot ion score > 20 and a minimum of 6 residues per peptide which results in a peptide FDR < 1%. For phosphorylation site localization, phosphoRS node in Proteome Discoverer was used (*29*). For peptide quantification, the precursor ion quantifier node in Proteome Discoverer was used, briefly, peptide ratios were determined using the extracted ion chromatograms (XICs) of the monoisotopic peaks of the “light”, “intermediate” and “heavy” labelled peptides. The retention time tolerance of isotope pattern multiplets was set at 0.2 min and the use of channels with single peaks was allowed for quantification. Further, peptides

missing one of the three quantification channels were considered. Phosphopeptides ratios (log_2 scale) were normalized by subtracting the median (log_2 scale) of non-phosphopeptides ratios to correct for errors introduced during the sample mixing. Manual quantification was performed on peptides that were identified in two biological replicates but only quantified in one replicate. Functional analyses were carried out with STRING database (version 9.0) (*30*).

3. **Results**

Protein identification and quantification

Before careful applications of hESCs and hiPSCs in cell therapy and research, it is important to determine potential molecular differences and similarities in between both cell lines. Previously, we and others concluded that at the global protein level, these two cell lines were much alike (*13, 14*), but still difference may occur at the PTM levels, e.g. protein phosphorylation. Protein tyrosine phosphorylation plays an important role in several signaling pathways, including those regulating self-renewal and pluripotency of ESCs. It is therefore interesting to examine the global tyrosine phosphorylation state in hiPSCs in comparison to hESCs. For this comparison, a quantitative mass spectrometry study was carried out where two different hiPS cell lines, named hiPS-NFF and hiPS-IMR90, and one hES cell line, HES-3, were used. For each cell line two biological replicates were analyzed to assess the reproducibility of the quantitative measurements. Cells were lysed and proteins digested using trypsin. The resulting peptides were labeled using triplex stable isotope dimethyl labeling and combined as depicted in Figure 1, followed by tyrosine phosphorylation enrichment. The enrichment was performed twice, a so-called re-IP) on the same peptide mixture, as we have previously observed an increase in the overall identification rate. The four phosphotyrosine enriched samples (two samples for each biological replicate) were desalted and subsequently analyzed in two technical replicates by LC-MS/MS.

The aforementioned experimental workflow led to the identification of 114 unique tyrosine phosphorylated peptides on 74 proteins in the first biological replicate. In the second replicate, 122 unique phosphotyrosine peptides were found on 95 proteins. From those, 79 and 97 peptides could be quanti-

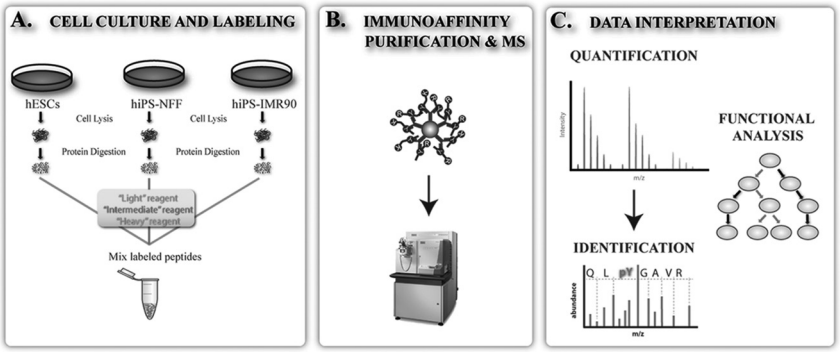


Figure 1: *Schematic workflow of the experiment. (A) Cells were lysed and proteins were digested. The resulting peptides were labeled and mixed. (B) Enrichment of tyrosine phosphorylated peptides was performed. The samples were analyzed by high resolution LC-MS/MS. (C) Peptides were identified and quantified and functional analysis was performed.*

Table 1: *Number of phosphotyrosine peptides and proteins identified and quantified in the data set.*

Classes	Biological replicate 1	Biological replicate 2	Total
	pY peptides	*pY peptides*	*pY peptides*
# Identified	114	122	190
# Quantified	79	97	131
# Peptides in both hiPSCs with $\log_2$ ratio > 1	22	25	47
# Peptides in both hiPSCs with $\log_2$ ratio <1	5	8	13
# Peptides in both hiPSCs with differential fold change	30	35	45
	pY proteins	*pY proteins*	*pY proteins*
# Identified	74	95	123
# Quantified	60	70	93
# Proteins in both hiPSCs with $\log_2$ ratio > 1	22	15	25
# Proteins in both hiPSCs with $\log_2$ ratio <1	5	7	10
# Peptides in both hiPSCs with differential fold change	27	22	35

fied, in the first and second biological replicate respectively (Table 1). In total, a cumulative number of 190 unique tyrosine phosphorylated peptides were identified (131 quantified) on 123 proteins (Table 1). 46 phosphotyrosine peptides could be quantified in common in the two biological replicates. A positive correlation was observed between the quantified phosphotyrosine peptides in hiPS-IMR90 and hiPS-NFF compared to HES3 (Figure 2). This trend was observed for both biological replicates indicating that

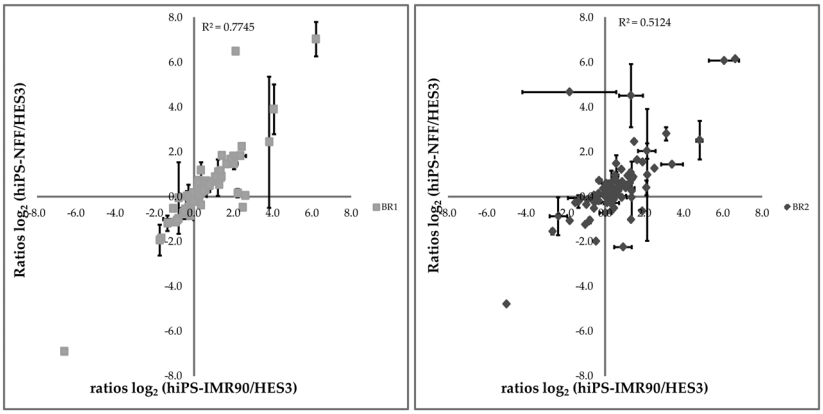


Figure 2: *Scatter plot of log_2 ratios of tyrosine phosphorylated peptides showing the reproducibility in quantitation in hiPS-NFF (y-axis) and hiPS-IMR90 (x-axis) when compared to HES3. a) Quantified phosphotyrosine peptides in biological replicate 1 and b) Quantified phosphotyrosine peptides in biological replicate 2. Error bars represent the standard deviation of the average values for technical replicates. Horizontal and vertical error bars are standard deviations in log_2 ratios for hiPS-IMR90 and hiPS-NFF, respectively, when compared to HES3.*

the phosphotyrosine peptides were reproducibly quantified in both hiPS cell lines in two biological replicates.

Global protein levels of both hiPS cell lines in comparison to the used hES cell lines were recently reported (*13, 14*). These studies showed a high proteome similarity in these cell lines. Of note, the same cell lines and culture conditions used in our analysis were also used in the study by Munoz *et al*. (*14*). We therefore quantified non-phosphorylated peptides identified in our LC/MS analyses to confirm the high degree of proteome similarity found by Munoz *et al.* (*14*) and to evaluate that the changes we observed in phosphorylation levels were not a reflection of changes in protein abundance (Figure 3). As observed in the box plot (Figure 3i) and in the MA-plot (Figure 3ii), the non-phosphorylated peptides have a less dispersed distribution compared to the phosphotyrosine peptides. A similar interpretation can be made from the MA-plot representation of the quantified tyrosine phosphorylated peptides distribution and the protein abundance in Figure 3iii. In accordance with Figure 3, the majority of the proteins quantified in our analyses have a ratio of less than two-fold, in agreement with previously published data where 97.8% of the quantified proteins have very similar protein levels (*14*). Likewise, the majority of phosphotyrosine peptides have a ratio less than two-fold. In contrast to the unphosphorylated peptides, phosphotyrosine peptides display larger variability in abundance. With an arbi-

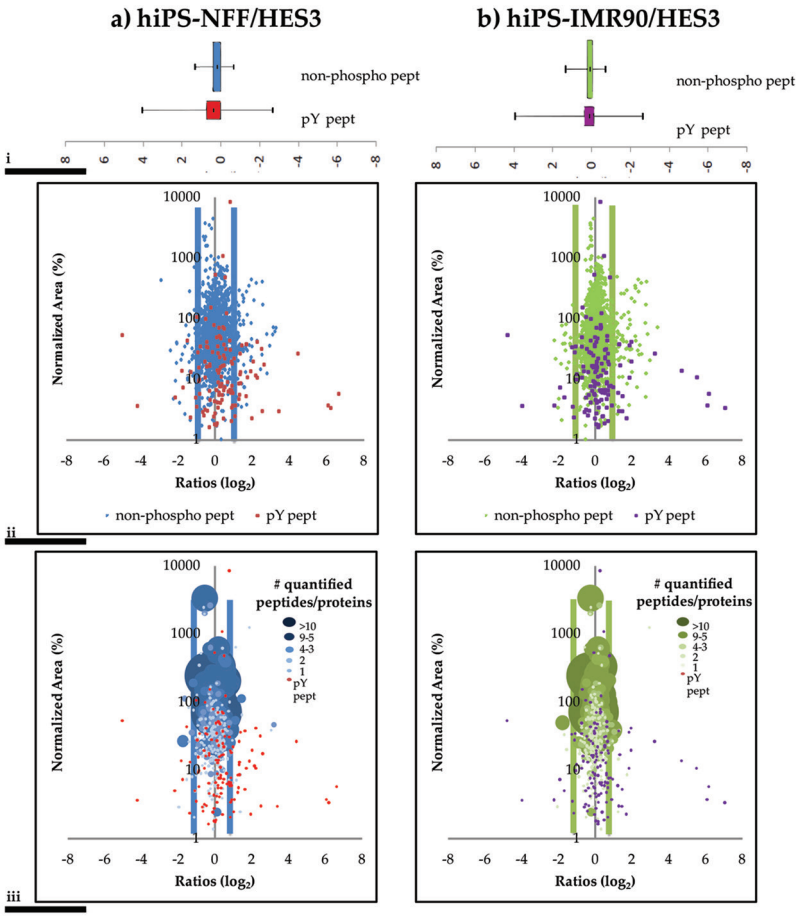


Figure 3: *Ratio distribution of proteins, tyrosine phosphorylated (pY pept) and non-phosphorylated peptides (non-phospho pept). i) In the box plots, mean values are shown as vertical dashes and the 75% confidence intervals are shown as colored boxes. The lowest and highest log_2 ratios values of the data are shown as error bars. ii) & iii) The colored lines in the MA plots show the arbitrary log_2 ratio cut off of 1. iii) In the MA-plots legend, the size of the dots represents the number of unique quantified peptides per protein.*

trary cut-off of two fold (ratio of 1 in log_2 scale), no less than 24% (31 phosphotyrosine peptides) of the tyrosine phosphorylated peptides have a different phosphorylation level in iPS-IMR90 along with 31% (40 phosphotyrosine peptides) in iPS-NFF when compared to HES3. In total, 45 unique phosphotyrosine peptides (belonging to 35 proteins) showed differences in tyrosine phosphorylation levels in hiPS-IMR90 and in hiPS-NFF cell lines (Table1). Amongst them, 35 peptides (31 proteins) showed similar phosphory-

lation levels in both hiPSCs, but being differential when compared to hESCs. The list of all the phosphotyrosine peptides quantified is displayed in Table 2.

SRC family kinases and RTKs show different levels of tyrosine phosphorylation in hiPSCs compared to hESCs

Pathways involving FGF-2 and PI3K are important in the maintenance of hESCs pluripotency (*31, 32*). We therefore examined the presence of proteins belonging to these pathways. Amongst the proteins with different levels of tyrosine phosphorylation in hiPS-NFF and hiPS-IMR90 compared to HES3, we found several kinases of the SRC family (Figure 4a) and also several receptors tyrosine kinases (RTKs) (Figure 4b). Activation of SRC family kinases is thought to be important in the maintenance of ES cells in an undifferentiated state (*33*). In this study, SRC family kinase FYN was found to be phosphorylated at multiple sites (Y185, Y420, and Y213) and was detected with an elevated tyrosine phosphorylation level in both hiPSCs cell lines at its activation site Y420. Additionally, another SRC kinase, LCK ,was identified to be phosphorylated at three different sites (Y192, Y505, and Y394) with a lower level of phosphorylation of at least 2.5 fold in its inhibitory site Y192 in both hiPSCs (Figure 4a, Table 2). This result reflects potentially the site-specific regulation of these SRC family kinases in hiPSCs (*34*).
We reported previously, early signaling events in FGF-2 stimulated hESCs showed that all four FGFRs are activated upon FGF-2 stimulation and multiple receptor tyrosine kinases (RTKs) are transactivated (*16*). Interestingly, in the current study, tyrosine phosphorylation was detected on FGFR1 (Y653) and FGFR4 (Y642), with a higher phosphorylation level on FGFR1 (Y653) in hiPSCs (Figure 4b). Furthermore, another RTK, ephrin receptor EPHA2 was found with an elevated phosphorylation level on two tyrosine phosphorylation sites (Y575 and Y772) in hiPSCs (Figure 4b).

Protein-protein interactions are important processes regulating signal transduction pathways in the cell. To gain more insight into the links and functional roles among the proteins showing different tyrosine phosphorylation levels between hiPSCs and hESCs, the STRING database was used (Figure 5). Interestingly, the majority of the proteins in the resulting network included proteins involved in neuronal and cytoskeletal processes. Cytoskeletal

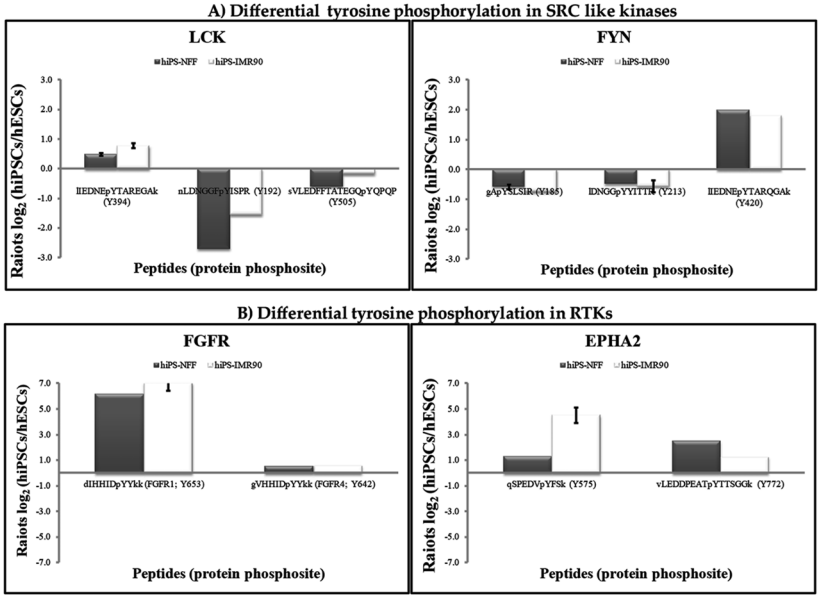


Figure 4: *Bar charts representing the tyrosine phosphorylation levels in hiPSCs when compared to hESCs of several tyrosine kinases.*

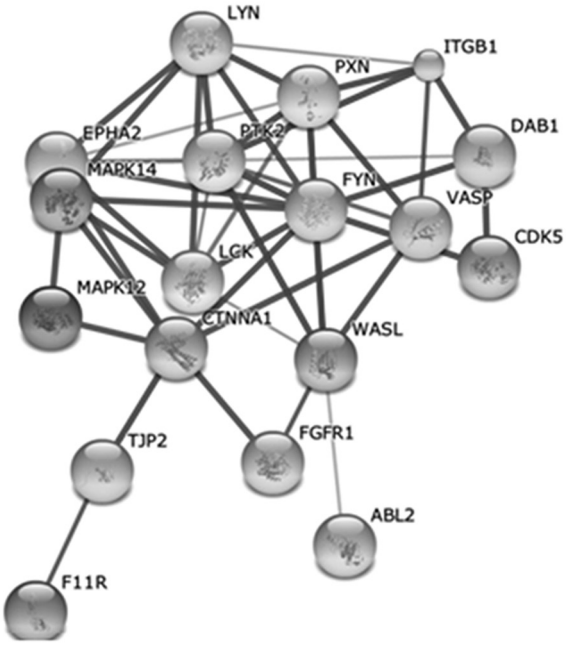


Figure 5: *STRING networks of the 31 tyrosine phosphorylated proteins showing consistent difference in phosphorylation levels in hiPSCs when compared to hESCs. 19 proteins out of 31 are linked in the same network. Proteins interacting in the network correspond to proteins in cytoskeletal processes (WASL, PXN, PTK2) and neuronal processes (CDK5, FYN, DAB1, ITGB1).*

proteins have been implicated in the maintenance of hESCs pluripotency (*15*). Moreover, Src family kinases LYN, FYN, and LCK and many Src kinase substrates are part of these cytoskeletal associated proteins included in the String interaction network (TJP2, F11R, PXN, WASL, CDK5, and PTK2). Src family kinases are involved in many cellular events such as cell proliferation, cytoskeletal rearrangement, differentiation, survival, cell adhesion and cell migration. Their involvement in these important cellular mechanisms is due to their ability to interact with many classes of receptors that regulate the mentioned processes. These receptors include G protein coupled receptors (GPCRs), cadherins, integrins, immunoglobulin super family adhesion molecules (CAMs) and RTKs (*35*). The interaction of Src kinases with receptors facilitate cross-talk between the different receptors (*35*). As mentioned earlier, we found several Src family kinases and many of their substrates with different levels of tyrosine phosphorylation in hiPSCs and hESCs suggesting that the regulation of several cellular processes involving Src kinases might differ between both cell lines.

Comapring hIPSC and hESCs phosphoproteomics datasets

Recently, Phanstiel *et al*. (*13*) reported a proteomic and phosphoproteomic comparison of hESCs and hiPSCs. This large dataset of 4,564 phosphosites included only 208 tyrosine phosphorylated peptides. We compared our dataset with the phosphotyrosine peptides identified by Phanstiel *et al*. (*13*) and observed a low overlap between both datasets. A plausible explanation for this finding may reside in the different experimental workflows used to generate both datasets indicating their complementarity. Additionally, different types of hiPS cell lines and hES cell lines were used in both studies, suggesting that possibly also biological variability contributes to the detection of different population of phosphotyorine peptides. Nine tyrosine phosphorylated peptides were identified and quantified in both datasets (Figure 6). Four peptides that were found in common in both datasets show more than two-fold difference between hiPSCs and hESCs in our dataset (Table 2), however, these difference in phosphorylation levels were not observed in the Phanstiel dataset (*13*). The discrepancy may pinpoint to earlier reported lab-specific signatures observed in iPSCs and ESCs (*36*), whereby microenvironmental conditions of distinct laboratories might have an effect on differences observed in iPSCs and ESCs at the molecular level.

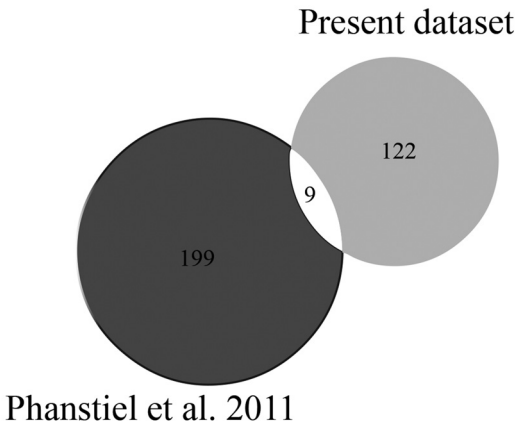


Figure 6: *Comparison of datasets describing (tyrosine) phosphorylation levels in hESCs and hiPSCs. The Venn diagram shows the overlap of tyrosine phosphorylated peptides between our dataset and the recently published data. Phanstiel et al (13) compared in their study hESCs and hiPSCs on proteomic and phosphoproteomic levels and found subtle and consistent differences between multiple hiPSCs and hESCs. Four out of the nine peptides found in common in both datasets have different tyrosine phosphorylation levels in hiPSCs when compared to hESCs in our data. These peptides can be found in Table 2 (peptide sequences in orange).*

4. **Discussion**

Reversible protein phosphorylation is a mechanism required in transferring signals from growth factors into the interior of the cell. To use hESCs and hiPSCs for clinical applications and research purposes it is important to investigate the phosphorylation-mediated signaling networks maintaining stem cell self-renewal and pluripotency. In the past few years, several studies have shown that multiple signaling factors such as FGF-2 and activin/nodal (*32, 37*), play a role in the maintenance of hESCs *in vitro*. Many of these growth factors signal through tyrosine phosphorylation. Therefore, it is essential to find out if these factors and signaling pathways controlling the pluripotency and self-renewal properties are similar in hiPSCs. In this study we aim to assess the phosphorylation levels of phosphotyrosine proteins in two different hiPS cell lines, hiPS-NFF and hiPS-IMR90 and compared them to those detected in hESCs (HES-3) by employing a stable isotope triple dimethyl labeling, immuno-affinity enrichment of tyrosine phosphorylated peptides and LC MS/MS analysis. We detected 190 phosphotyrosine peptides from two biological replicates. From the quantified proteins, we observed a consistent difference of two-fold or more in 31 phosphotyrosine proteins in both studied hiPSCs when compared to the hESCs. Analy-

ses of two hiPS cell lines ensure that the differences observed are likely not cell line specific.

Most of the observed differences were not detected in a recently reported phosphoproteomic dataset (*13*) on hESCs and hiPSCs, which anyway had little overlap with the tyrosine phosphorylated peptides in our work. The poor overlap between both datasets might be due to the different phosphopeptide enrichment techniques used. Phanstiel *et al*. (*13*) used Fe^{3+} coated magnetic beads after a pre-fractionation of the samples by SCX to enrich for phosphorylated peptides including serine, threonine and tyrosine phosphorylated peptides while our method is specifically used for the enrichment of tyrosine phosphorylated peptides..

In our analysis, several proteins showing a difference in tyrosine phosphorylation levels have been demonstrated in other studies to be involved in the regulation of pluripotency of hESCs. Multiple receptor tyrosine kinases (RTKs) and their downstream pathway members are shown to be necessary for the maintenance of the undifferentiated state (*16, 38*). In this study, we identify several RTKs and other tyrosine phosphorylated proteins with an involvement in the regulation of hESCs self-renewal and pluripotency such as SRC family kinases, and cytoskeletal associated proteins showing a difference in tyrosine phosphorylation level in hiPSCs (Table 2). Although, mouse and human iPSCs are not cultured under the same conditions and their unique properties of self-renewal and pluripotency are not controlled by the same pathways, numerous studies on miPSCs show that some (tyrosine) phosphorylation signaling proteins involved in the maintenance of self-renewal and pluripotency seem to play a role in the reprogramming of somatic cells (*39-41*). SRC family kinases are proteins that are suggested to have an inhibitory effect on reprogramming (*39*). In the absence of exogenous Sox2, inhibition of Src family kinases allows the reprogramming of somatic cells (*39*). Moreover, several growth factors or the inhibition of growth factors (e.g. Alk5 inhibitor, Wnt3A) have been found to improve reprogramming (*40*). Silva *et al.* (*41*) demonstrated that inhibition of MAPK and GSK3 signaling pathway and stimulation by LIF promoted more rapidly the reprogramming of mouse brain derived stem cells in iPSCs. Therefore, a suggestion can be made from the results of the aforementioned studies that a number of pathways in hiPSCs might as well effect the reprogramming of human somatic cells. In agreement with this proposition, several proteins (Table 2) in our current analysis showing different levels of tyrosine phos-

phorylation in hiPSCs might have an influence on the success rate of reprogramming and the maintenance of stem cell properties in reprogrammed cells. These proteins include SRC family kinases which are shown to have a negative effect on reprogramming of mouse somatic cells. In addition to Src family kinases, follow up studies on proteins showing differential phosphorylation levels in our dataset will be necessary to find candidates involved in the reprogramming of hiPSCs. Modifying the activity of tyrosine kinases responsible for the phosphorylation of these proteins might lead to efficient reprogramming of human somatic cells.

Table 2: *list of quantified tyrosine phosphorylated peptides with details about peptide sequences with the indication of the phosphorylated residues (pY), protein accession number, protein description, phosphosites, log_2 ratios for each biological replicate (BR1, BR2) and log_2 protein ratios (proteins).*

Cells in green and red: increased (red) and decreased (green) in phosphorylation and proteins levels in hiPSCs when compared to hESCs.

Cells in blue: peptides which have more than two-fold change in phosphorylation in hiPSCs compared to hESCs. White peptides: peptides with contradictory phosphorylation levels in hiPS-NFF and hiPS-IMR90.

Orange peptides: peptides found in common in our dataset and in the dataset from Phanstiel et a.l. (13) showing different phosphorylation levels in our dataset.

				hiPS-NFF/HES3			hiPS-IMR90/HES3		
SEQUENCES	**PROTEIN ACCESSIONS**	**DESCRIPTION**	**pY SITES**	**BR1**	**BR2**	**PROTEINS**	**BR1**	**BR2**	**PROTEINS**
aAASTDpYYk	IPI00909951.1	PDHA1/LOC79064 protein (Fragment)	Y242		1.3	-0.6		-1.0	-0.5
acpYRDmSSFPETk	IPI00745182.1	Putative uncharacterized protein DKFZp434N101 (Fragment)	Y262		0.1			-0.2	
aDGGAEpYATYQTk	IPI00479283.1	Isoform 3 of Membrane cofactor protein	Y384		-0.1	-0.1		0.4	0.1
aLDpYYmLR	IPI00642355.1	Talin 1	Y70	0.6		1.0	0.3		0.2
aLDYpYmLR	IPI00642355.1	Talin 1	Y71	0.4		1.0	0.3		0.2
apSSPAESSPEDSGpYmR	IPI00464978.1	Insulin receptor substrate 2 insertion mutant (Fragment)	Y730, Y742		0.8	-0.9		0.5	-0.2
aPYTcGGDSDQpYVLmSSPVGR	IPI00464978.1	Insulin receptor substrate 2 insertion mutant (Fragment)	Y823		-0.3	-0.9		-0.2	-0.2
aSpSPAESSPEDSGpYmR	IPI00464978.1	Insulin receptor substrate 2 insertion mutant (Fragment)	Y731, Y742		0.8	-0.9		0.5	-0.2
aSSPAEpSSPEDSGpYmR	IPI00464978.1	Insulin receptor substrate 2 insertion mutant (Fragment)	Y735, Y742		0.9	-0.9		0.6	-0.2
aSSPAESSPEDSGpYmR	IPI00464978.1	Insulin receptor substrate 2 insertion mutant (Fragment)	Y742		0.2	-0.9		0.5	-0.2
aTEQpYYAmk	IPI00889577.1	Protein kinase, cAMP-dependent, catalytic, beta	Y69		-0.9	-0.6		-0.1	0.0

aTEQYpYAmk	IPI00889577.1	Protein kinase, cAMP-dependent, catalytic, beta	Y70		-0.9	-0.6		-0.1	0.0
aVcSTpYLQSR	IPI00215949.5	Isoform 2 of Homeodomain-interacting protein kinase 2	Y361	0.0	0.0		-0.2	0.0	
dDpYFAk	IPI00009032.1	Lupus La protein	Y188		0.0	0.1		0.4	-0.1
dIpYETDYYR	IPI00027232.3	Insulin-like growth factor 1 receptor	Y1161	0.3		0.7	-0.4		0.4
dIYETDpYYR	IPI00027232.3	Insulin-like growth factor 1 receptor	Y1165	0.2		0.7	0.0		0.4
dIYETDYpYR	IPI00027232.3	Insulin-like growth factor 1 receptor	Y1166	0.2		0.7	0.0		0.4
dNEVDGQDpYHFVVSR	IPI00647338.1	synapse-associated protein 102 isoform b	Y673	-0.2		0.0	-0.3		0.2
dPTNGpYYNVR	IPI00844319.1	Kin of IRRE like	Y637		0.4			-0.1	
dSLpYAQGk	IPI00738898.2	similar to large subunit ribosomal protein L36a	-		0.0			0.0	
dSLpYAQGR	IPI00056494.5	60S ribosomal protein L36a-like	Y34		0.4	0.3		0.0	-0.3
dSNPpYATLPR	IPI00893194.2	cDNA FLJ42612 fis, clone BRACE3013780	Y300		0.5	0.7		1.0	0.8
dSpYVGDEAQSk	IPI00894365.2	cDNA FLJ52842, highly similar to Actin, cytoplasmic 1	Y53	-0.3	-1.0		-0.5	-0.3	
dSYSSRDpYPSSR	IPI00816796.1	cDNA FLJ38696 fis, clone KIDNE2001931, highly similar to HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN G	Y212		-0.6	0.0		-0.5	0.3
eHALLApYTLGVk	IPI00396485.3	Elongation factor 1-alpha 1	Y141	0.1		0.3	0.0		0.0
ePPPPpYLPA	IPI00014172.1	Lysosomal-associated transmembrane protein 4A	Y230		1.0			0.4	
ePPPPpYVSA	IPI00903298.1	Isoform 2 of Lysosomal-associated transmembrane protein 4B	Y367		0.9	1.3		0.7	
eVSTpYIk	IPI00396485.3	Elongation factor 1-alpha 1	Y177		-0.4	0.3		0.1	0.0
eYDQLpYEEYTR	IPI00796770.1	PIK3R2 protein	Y464	0.3		0.3	0.1		
fDTQYPpYGEk	IPI00024911.1	Endoplasmic reticulum protein ERp29	Y66		-0.3	0.2		0.7	0.3
gApYSLSIR	IPI00166845.2	Isoform 3 of Proto-oncogene tyrosine-protein kinase Fyn	Y185	-0.6		0.2	-0.7		0.3
gEPNVSpYIcSR	IPI00880060.1	Putative uncharacterized protein GSK3A	Y279	0.4	0.4	0.0	0.2	0.4	0.3
gHEpYTNIk	IPI00298347.1	Isoform 2 of Tyrosine-protein phosphatase non-receptor type 11	Y546		0.9	0.8		-2.2	-0.6
gPSEpYPTkNpYV	IPI00011084.2	Claudin-6	Y214, Y219	0.2		0.0	0.7		0.0
gSGDpYmPmSPk	IPI00019471.1	Insulin receptor substrate 1	Y989		0.1			0.5	1.5
gVHHIDpYYkk	IPI00910521.1	cDNA FLJ60908, highly similar to Homo sapiens fibroblast growth factor receptor 4 (FGFR4), transcript variant 2, mRNA	Y642	0.5			0.6		
gYPGDpYTk	IPI00218408.1	Isoform Short of Trifunctional purine biosynthetic protein adenosine-3	Y348		0.5	0.3		0.1	0.0
iAIpYELLFk	IPI00008438.1	40S ribosomal protein S10	Y12	0.2		0.0	0.1		0.1
iGEGTpYGVVYk	IPI00073536.2	Isoform 2 of Cell division control protein 2 homolog	Y15	0.4	0.3	0.2	0.3	0.6	0.6

iGEGTYGVVpYk	IPI00073536.2	Isoform 2 of Cell division control protein 2 homolog	Y19	0.2	0.1	0.2	0.0	0.3	0.6
iIQLLDDpYPk	IPI00794884.1	24 kDa protein	-	0.2	0.6		0.0	0.4	
imEpYYEk	IPI00719366.1	vacuolar H+ ATPase E1 isoform c	Y34		-0.5	0.2		0.0	0.1
imEYpYEk	IPI00719366.1	vacuolar H+ ATPase E1 isoform c	Y35		-0.6	0.2		0.1	0.1
iQNTGDpYYDLYGGEk	IPI00658002.1	Isoform 3 of Tyrosine-protein phosphatase non-receptor type 11	Y62	0.8		0.8	0.7		-0.6
iQNTGDYpYDLYGGEk	IPI00658002.1	Isoform 3 of Tyrosine-protein phosphatase non-receptor type 11	Y63	0.8		0.8	0.7		-0.6
iWHHTFpYNELR	IPI00894365.2	cDNA FLJ52842, highly similar to Actin, cytoplasmic 1	Y91	-1.0			-1.1		
iYQpYIQSR	IPI00219251.1	Isoform 2 of Dual specificity tyrosine-phosphorylation-regulated kinase 1A	Y321	0.2	0.3	0.1	0.1	0.1	-1.3
lcDFGSASHVADNDITPpYLVSR	IPI00552169.1	43 kDa protein	-	-0.2	0.0		0.0	0.6	
lDNGGpYYITTR	IPI00166845.2	Isoform 3 of Proto-oncogene tyrosine-protein kinase Fyn	Y213	-0.5	-0.4	0.2	-0.6	-0.1	0.3
lDNGGYpYITTR	IPI00166845.2	Isoform 3 of Proto-oncogene tyrosine-protein kinase Fyn	Y214	-0.8	-0.6	0.2	-1.0	-0.2	0.3
lIEDNEpYTAREGAk	IPI00555672.1	Proto-oncogene tyrosine-protein kinase LCK	Y394		0.5	-1.9		0.8	-0.5
lmpYQELk	IPI00328243.2	Phospholipase D3	Y7	0.2		0.9	0.6		0.9
mQNHGpYENPTYk	IPI00334946.4	Isoform 5 of Amyloid-like protein 2	Y750	0.6	1.1	0.1	0.7	0.9	1.4
nAGNEQDLGIQpYk	IPI00215948.4	Isoform 1 of Catenin alpha-1	Y177	-0.4	-0.5	0.2	-0.6	-2.0	0.2
nEEENIpYSVPHDSTQGk	IPI00718985.1	Isoform 2 of Glucocorticoid receptor DNA-binding factor 1	Y1105	0.0	0.2	0.4	-0.1	0.1	0.1
nLpYAGDYpYR	IPI00219996.1	Isoform 2 of Epithelial discoidin domain-containing receptor 1	Y792, Y797		0.3			-0.1	0.7
nLpYAGDYYR	IPI00219996.1	Isoform 2 of Epithelial discoidin domain-containing receptor 1	Y792	0.3	0.5		1.2	1.5	0.7
nNASTDpYDLSDk	IPI00889059.1	similar to ribosomal protein L3, partial	-		0.1			0.6	
nVpYYELNDVR	IPI00908793.1	cDNA FLJ56199, weakly similar to p130Cas-associated protein	Y165	-0.5			-0.7		
pYGLFkEENPpYAR	IPI00911013.1	cDNA FLJ57175, moderately similar to Pituitary tumor-transforming gene 1 protein-interacting protein	Y111, Y120	0.8		0.0	0.5		0.0
pYmEDSTYYk	IPI00216218.1	Isoform 4 of Focal adhesion kinase 1	Y570		0.3	0.2		-0.3	0.1
qASEQNWANpYSAEQNR	IPI00910868.1	cDNA FLJ50955, highly similar to Gap junction alpha-1 protein	-	0.7	0.8		0.5	0.0	
qEDHAEAALpYk	IPI00848233.1	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog isoform 2 precursor	Y703		0.6			0.6	2.1
qGpYVIYR	IPI00910417.1	Ribosomal protein L15	Y59		0.2	-0.1		0.4	0.0
qGTpYGGYFR	IPI00873286.2	cDNA FLJ55750, highly similar to Eukaryotic translation initiation factor 3 subunit 8	Y703	0.1		0.2	0.1		0.2

qVAYIpYk	IPI00639931.1	Adenylyl cyclase-associated protein	Y351		0.7	0.9		0.3	0.0
rSDSASSEPVGIpY-QGFEk	IPI00887373.1	Protein kinase C delta VIII (Fragment)	Y90	-0.2	0.3		-0.8	-0.6	
sDPpYHATSGALp-SPAk	IPI00910868.1	cDNA FLJ50955, highly similar to Gap junction alpha-1 protein	-		-0.8			-1.1	
sDPpYHATSGALSPAk	IPI00910868.1	cDNA FLJ50955, highly similar to Gap junction alpha-1 protein	-	-0.2	-0.8		-0.4	-1.0	
sESVVpYADIR	IPI00893201.1	Myelin protein zero-like 1	Y263	0.2	0.3	1.3	0.0	-0.1	0.2
sFLDSGpYR	IPI00291175.7	Isoform 1 of Vinculin	Y822	0.7	1.1	0.2	0.5	0.5	0.0
sGDLPpYDGR	IPI00908558.1	cDNA FLJ55699, highly similar to Homo sapiens liver-specific bHLH-Zip transcription factor (LISCH7), transcript variant 3, mRNA	Y430	-0.1			0.2		-0.9
spYSPDGkESPsDkk	IPI00017297.1	Matrin-3	Y597	-0.9		-0.4	-1.0		0.2
sRDPHpYDDFR	IPI00908558.1	cDNA FLJ55699, highly similar to Homo sapiens liver-specific bHLH-Zip transcription factor (LISCH7), transcript variant 3, mRNA	Y395	0.2			0.4		-0.9
sSGPpYGGGGQY-FAkPR	IPI00797148.1	Isoform 2 of Heterogeneous nuclear ribonucleoprotein A1	Y341	0.3		0.1	0.1		0.6
sSGPYGGGGQpY-FAkPR	IPI00797148.1	Isoform 2 of Heterogeneous nuclear ribonucleoprotein A1	Y347	0.3		0.1	0.1		0.6
sTcIpYGGAPk	IPI00798375.2	cDNA FLJ59357, highly similar to Probable ATP-dependent RNA helicase DDX5	Y202		-0.5	-0.5		-0.3	0.2
sTTTGHLIpYk	IPI00396485.3	Elongation factor 1-alpha 1	Y29	0.5	0.2	0.3	0.3	0.3	0.0
sVLEDFFTATEGQpY-QPQP	IPI00555672.1	Proto-oncogene tyrosine-protein kinase LCK	Y505	-0.1	-0.6	-1.9	-0.2	-0.2	-0.5
tVcSTpYLQSR	IPI00554546.1	Isoform 2 of Homeodomain-interacting protein kinase 3	Y359		0.4			0.0	
vLLPEpYGGTk	IPI00220362.5	10 kDa heat shock protein, mitochondrial	Y76	-0.5	0.0	-0.6	-0.7	-0.7	-0.2
vLpYYmEk	IPI00908558.1	cDNA FLJ55699, highly similar to Homo sapiens liver-specific bHLH-Zip transcription factor (LISCH7), transcript variant 3, mRNA	-		0.1			-0.3	
vLYpYmEk	IPI00908558.1	cDNA FLJ55699, highly similar to Homo sapiens liver-specific bHLH-Zip transcription factor (LISCH7), transcript variant 3, mRNA	-	0.2	0.1		0.4	-0.3	
vQPNEAVpYTk	IPI00884082.2	cDNA FLJ56794, highly similar to Glucose-6-phosphate 1-dehydrogenase	-	-0.8	0.2		0.0	0.1	
vVQDTpYQImk	IPI00006176.3	Hepatocyte growth factor-regulated tyrosine kinase substrate	Y132		0.0	0.2		0.2	0.3
vYTpYIQSR	IPI00555676.1	Isoform 2 of Dual specificity tyrosine-phosphorylation-regulated kinase 4	Y264	-0.1	-0.1		0.0	0.1	

aNRDLIpYk	IPI00215948.4	Isoform 1 of Catenin alpha-1	Y245		-1.0	0.2		-1.2	0.2
dAVpYSEYk	IPI00514780.5	Isoform 2 of Disks large homolog 5	Y429	-1.4		0.4	-1.2		
dIHHIDpYYkk	IPI00455176.1	Isoform 12 of Basic fibroblast growth factor receptor 1	Y653	6.2		-0.8	7.0		1.9
eDIpYSGGGGGGSR	IPI00011913.1	Heterogeneous nuclear ribonucleoprotein A0	Y180	1.8	1.4	-0.2	1.7	0.9	0.3
eFEVpYGPIk	IPI00219484.1	Isoform 3 of U1 small nuclear ribonucleoprotein 70 kDa	Y126	2.2	2.1	0.1	0.2	0.4	0.0
eGVpYDVPk	IPI00433142.2	Isoform DAB469 of Disabled homolog 1	Y232	3.8	2.1	2.5	2.5	1.0	
eSYSVpYVYk	IPI00018534.4	Histone H2B type 1-L	Y41	-0.3	-2.4		0.1	-0.9	
eSYSVYVpYk	IPI00018534.4	Histone H2B type 1-L	Y43	-0.8	-1.2		-1.0	-0.1	
gIVVpYTGDR	IPI00414005.2	Isoform Short of Sodium/potassium-transporting ATPase subunit alpha-1	Y260	1.4	3.1	0.2	0.9	2.8	0.3
gQmPENPpYSEVGk	IPI00788637.1	121 kDa protein	-	1.4			1.8		
hTDDEmpTGpYVATR	IPI00221142.3	Isoform Mxi2 of Mitogen-activated protein kinase 14	T180, Y182		3.4	1.0		1.5	-0.7
hTDDEmTGpYVATR	IPI00221142.3	Isoform Mxi2 of Mitogen-activated protein kinase 14	Y182	4.0	4.8	1.0	3.9	2.5	-0.7
iDYGEpYmDk	IPI00029744.1	Single-stranded DNA-binding protein, mitochondrial	Y116		6.6	-0.5		6.2	0.0
iGEGpTpYGVVYk	IPI00910650.1	cDNA FLJ54979, highly similar to Homo sapiens cyclin-dependent kinase 2 (CDK2), transcript variant 2, mRNA	T14, Y15	-0.4	-1.4		0.0	-0.2	
iGEGTpYGTVFk	IPI00827700.1	Protein kinase CDK5 splicing variant	Y15	1.1	0.7	0.3	0.8	0.4	-0.2
iNPIpYDALSYSSPSDSYQGk	IPI00887539.1	similar to Polio virus receptor protein	-		1.5			2.5	
lGEGpTpYATVYk	IPI00442129.1	cDNA FLJ16665 fis, clone THYMU2031249, highly similar to SERINE/THREONINE-PROTEIN KINASE PCTAIRE-1	T196, Y197	-6.6	-1.8		-6.9	-1.1	
lGEGTpYATVYk	IPI00442129.1	cDNA FLJ16665 fis, clone THYMU2031249, highly similar to SERINE/THREONINE-PROTEIN KINASE PCTAIRE-1	Y197	1.2	0.3		0.9	0.4	
lIEDNEpYTARQGAk	IPI00166845.2	Isoform 3 of Proto-oncogene tyrosine-protein kinase Fyn	Y420	2.0	1.9	0.2	1.8	1.6	0.3
lmTGDTpYTAHAGAk	IPI00914546.1	v-abl Abelson murine leukemia viral oncogene homolog 2 isoform e	Y439	2.0	1.3		1.5	1.1	
lSQDPpYDLPk	IPI00064607.3	Isoform 1 of Multiple epidermal growth factor-like domains 10	Y1099		-1.8			4.7	
lVpYDGIR	IPI00215948.4	Isoform 1 of Catenin alpha-1	Y619	-1.1		0.2	-0.5		0.2
mQNHGpYENPpTYk	IPI00334946.4	Isoform 5 of Amyloid-like protein 2	Y750, T754		1.2	0.1		0.6	1.4
mQNHGpYENPTpYk	IPI00334946.4	Isoform 5 of Amyloid-like protein 2	Y750, Y755		1.2	0.1		0.6	1.4
mQNHGYENPTpYk	IPI00334946.4	Isoform 5 of Amyloid-like protein 2	Y755	1.3	1.3	0.1	0.5	0.8	1.4

mQQNGpYENPpTYk	IPI00909502.1	cDNA FLJ54261, highly similar to Homo sapiens amyloid beta (A4) protein, transcript variant 3, mRNA	Y609, T613		1.9			-0.6	
mQQNGpYENPTpYk	IPI00909502.1	cDNA FLJ54261, highly similar to Homo sapiens amyloid beta (A4) protein, transcript variant 3, mRNA	Y609, Y614		1.3			0.0	
mQQNGpYENPTYk	IPI00909502.1	cDNA FLJ54261, highly similar to Homo sapiens amyloid beta (A4) protein, transcript variant 3, mRNA	Y609	1.2	1.3		1.1	0.3	
mQQNGYENPTpYk	IPI00909502.1	cDNA FLJ54261, highly similar to Homo sapiens amyloid beta (A4) protein, transcript variant 3, mRNA	Y614		1.3			0.0	
nLDNGGFpYISPR	IPI00555672.1	Proto-oncogene tyrosine-protein kinase LCK	Y192	-1.7	-2.7	-1.9	-1.8	-1.5	-0.5
nLIpYDNADNk	IPI00026952.1	Plakophilin-3	Y390		-5.0	0.4		-4.8	
pYHGHpSmSDP-GVSpYR	IPI00642820.1	Pyruvate dehydrogenase (Lipoamide) alpha 1	Y289, S293, Y301	2.6		-0.6	0.1		-0.5
pYmEDSTpYYk	IPI00216217.1	Isoform 4 of Focal adhesion kinase 1	Y570, Y576	1.3	0.4	0.2	1.2	-0.5	0.1
qADSEmTGpYVVTR	IPI00853109.1	Putative uncharacterized protein MAPK12	Y137		0.8			1.2	
qmNpYSSLPEk	IPI00009607.1	Ras-related protein Rap-2c	Y166		-1.5			-0.3	0.0
qSPEDVpYFSk	IPI00021267.1	Ephrin type-A receptor 2	Y575	2.1	1.3	2.2	6.5	4.5	0.7
sEDIpYADPAAYVmR	IPI00853477.1	Phosphoinositol 3-phosphate-binding protein 3 (Fragment)	Y492	-1.8			-1.9		
vGEEEHVpYSFPNk	IPI00220030.1	Isoform Alpha of Paxillin	Y118	2.5		1.3	-0.5		
vIEDNEpYTAREGAk	IPI00432416.4	Isoform LYN B of Tyrosine-protein kinase Lyn	Y397	2.4	1.6	0.0	2.2	1.6	0.3
vIpYDFIEk	IPI00011676.2	Neural Wiskott-Aldrich syndrome protein	Y256	2.3	2.1	-0.2	1.8	2.1	0.2
vIpYSQPSAR	IPI00069985.3	Junction adhesion molecule	Y280	1.4	0.6	0.2	0.9	0.1	-0.3
vLEDDPEATpYTTSGGk	IPI00021267.1	Ephrin type-A receptor 2	Y772		2.5	2.2		1.3	0.7
vQIpYHNPTANSFR	IPI00301058.5	Vasodilator-stimulated phosphoprotein	Y39	1.1		0.6	0.9		-0.1
vVDTLpYDGk	IPI00291668.6	Tight junction protein 2	Y638		6.0	0.0		6.1	-0.1
wDTGENPIpYk	IPI00217563.3	Isoform Beta-1A of Integrin beta-1	Y783	1.6		1.2	1.5		0.3

5. References

[1] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, J. M. Jones, Embryonic stem cell lines derived from human blastocysts, Science **282** (1998) 1145-1147.

[2] D. Dey, G. R. Evans, Generation of Induced Pluripotent Stem (iPS) Cells by Nuclear Reprogramming, Stem Cells Int **2011** (2011) 619583.

[3] R. Passier, L. W. van Laake, C. L. Mummery, Stem-cell-based therapy and lessons from the heart, Nature **453** (2008) 322-329.

[4] K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors, Cell **126** (2006) 663-676.

[5] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, Induction of pluripotent stem cells from adult human fibroblasts by defined factors, Cell **131** (2007) 861-872.

[6] J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, Slukvin, II, J. A. Thomson, Induced pluripotent stem cell lines derived from human somatic cells, Science **318** (2007) 1917-1920.

[7] J. Yu, K. Hu, K. Smuga-Otto, S. Tian, R. Stewart, Slukvin, II, J. A. Thomson, Human induced pluripotent stem cells free of vector and transgene sequences, Science **324** (2009) 797-801.

[8] J. W. Han, Y. S. Yoon, Induced pluripotent stem cells: emerging techniques for nuclear reprogramming, Antioxid Redox Signal **15** (2011) 1799-1820.

[9] M. H. Chin, M. J. Mason, W. Xie, S. Volinia, M. Singer, C. Peterson, G. Ambartsumyan, O. Aimiuwu, L. Richter, J. Zhang, I. Khvorostov, V. Ott, M. Grunstein, N. Lavon, N. Benvenisty, C. M. Croce, A. T. Clark, T. Baxter, A. D. Pyle, M. A. Teitell, M. Pelegrini, K. Plath, W. E. Lowry, Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures, Cell Stem Cell **5** (2009) 111-123.

[10] M. G. Guenther, G. M. Frampton, F. Soldner, D. Hockemeyer, M. Mitalipova, R. Jaenisch, R. A. Young, Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells, Cell Stem Cell **7** (2010) 249-257.

[11] Y. Ohi, H. Qin, C. Hong, L. Blouin, J. M. Polo, T. Guo, Z. Qi, S. L. Downey, P. D. Manos, D. J. Rossi, J. Yu, M. Hebrok, K. Hochedlinger, J. F. Costello, J. S. Song, M. Ramalho-Santos, Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells, Nat Cell Biol **13** (2011) 541-549.

[12] K. D. Wilson, S. Venkatasubrahmanyam, F. Jia, N. Sun, A. J. Butte, J. C. Wu, MicroRNA profiling of human-induced pluripotent stem cells, Stem Cells Dev **18** (2009) 749-758.

[13] D. H. Phanstiel, J. Brumbaugh, C. D. Wenger, S. Tian, M. D. Probasco, D. J. Bailey, D. L. Swaney, M. A. Tervo, J. M. Bolin, V. Ruotti, R. Stewart, J. A. Thomson, J. J. Coon, Proteomic and phosphoproteomic comparison of human ES and iPS cells, Nat Methods **8** (2011) 821-827.

[14] J. Munoz, T. Y. Low, Y. J. Kok, A. Chin, C. K. Frese, V. Ding, A. Choo, A. J. Heck, The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells, Molecular systems biology **7** (2011) 550.

[15] A. D. Zoumaro-Djayoon, V. Ding, L. Y. Foong, A. Choo, A. J. Heck, J. Munoz, Investigating the role of FGF-2 in stem cell maintenance by global phosphoproteomics profiling, Proteomics **11** (2011) 3962-3971.

[16] V. M. Ding, P. J. Boersema, L. Y. Foong, C. Preisinger, G. Koh, S. Natarajan, D. Y. Lee, J. Boekhorst, B. Snel, S. Lemeer, A. J. Heck, A. Choo, Tyrosine phosphorylation profiling in FGF-2 stimulated human embryonic stem cells, PLoS One **6** (2011) e17538.

[17] S. Avery, K. Inniss, H. Moore, The regulation of self-renewal in human embryonic stem cells, Stem Cells Dev **15** (2006) 729-740.

[18] A. Pebay, R. C. Wong, S. M. Pitson, E. J. Wolvetang, G. S. Peh, A. Filipczyk, K. L. Koh, I. Tellis, L. T. Nguyen, M. F. Pera, Essential roles of sphingosine-1-phosphate and platelet-derived growth factor in the maintenance of human embryonic stem cells, Stem Cells **23** (2005) 1541-1548.

[19] S. C. Bendall, M. H. Stewart, P. Menendez, D. George, K. Vijayaragavan, T. Werbowetski-Ogilvie, V. Ramos-Mejia, A. Rouleau, J. Yang, M. Bosse, G. Lajoie, M. Bhatia, IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro, Nature **448** (2007) 1015-1021.

[20] P. J. Boersema, L. Y. Foong, V. M. Ding, S. Lemeer, B. van Breukelen, R. Philp, J. Boekhorst, B. Snel, J. den Hertog, A. B. Choo, A. J. Heck, In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling, Molecular & cellular proteomics : MCP **9** (2010) 84-99.

[21] A. Choo, J. Padmanabhan, A. Chin, W. J. Fong, S. K. Oh, Immortalized feeders for the scale-up of human embryonic stem cells in feed-

er and feeder-free conditions, J Biotechnol **122** (2006) 130-141.

[22] A. Choo, J. Padmanabhan, A. Chin, W. J. Fong, S. K. Oh, Immortalized feeders for the scale-up of human embryonic stem cells in feeder and feeder-free conditions, J Biotechnol **122** (2006) 130-141.

[23] V. Ding, A. B. Choo, S. K. Oh, Deciphering the importance of three key media components in human embryonic stem cell cultures, Biotechnol Lett **28** (2006) 491-495.

[24] P. J. Boersema, L. Y. Foong, V. M. Ding, S. Lemeer, B. van Breukelen, R. Philp, J. Boekhorst, B. Snel, J. den Hertog, A. B. Choo, A. J. Heck, In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling, Mol Cell Proteomics **9** 84-99.

[25] P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A. J. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics, Nat Protoc **4** (2009) 484-494.

[26] A. D. Zoumaro-Djayoon, A. J. Heck, J. Munoz, Targeted analysis of tyrosine phosphorylation by immuno-affinity enrichment of tyrosine phosphorylated peptides prior to mass spectrometric analysis, Methods (2011).

[27] J. Gobom, E. Nordhoff, E. Mirgorodskaya, R. Ekman, P. Roepstorff, Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry, J Mass Spectrom **34** (1999) 105-116.

[28] C. K. Frese, A. F. Altelaar, M. L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A. J. Heck, S. Mohammed, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos, Journal of proteome research **10** (2011) 2377-2388.

[29] T. Taus, T. Kocher, P. Pichler, C. Paschke, A. Schmidt, C. Henrich, K. Mechtler, Universal and confident phosphorylation site localization using phosphoRS, Journal of proteome research **10** (2011) 5354-5362.

[30] B. Snel, G. Lehmann, P. Bork, M. A. Huynen, STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene, Nucleic Acids Res **28** (2000) 3442-3444.

[31] P. Dvorak, D. Dvorakova, S. Koskova, M. Vodinska, M. Najvirtova, D. Krekac, A. Hampl, Expression and potential role of fibroblast growth factor 2 and its receptors in human embryonic stem cells, Stem Cells **23** (2005) 1200-1211.

[32] L. Vallier, M. Alexander, R. A. Pedersen, Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells, J Cell Sci **118** (2005) 4495-4509.

[33] C. Anneren, C. A. Cowan, D. A. Melton, The Src family of tyrosine kinases is important for embryonic stem cell self-renewal, The Journal of biological chemistry **279** (2004) 31590-31598.

[34] J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, M. Mann, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks, Cell **127** (2006) 635-648.

[35] S. M. Thomas, J. S. Brugge, CELLULAR FUNCTIONS REGULATED BY SRC FAMILY KINASES, Annual Review of Cell and Developmental Biology **13** (1997) 513-609.

[36] A. M. Newman, J. B. Cooper, Lab-specific gene expression signatures in pluripotent stem cells, Cell Stem Cell **7** (2010) 258-262.

[37] R. H. Xu, R. M. Peck, D. S. Li, X. Feng, T. Ludwig, J. A. Thomson, Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells, Nat Methods **2** (2005) 185-190.

[38] L. M. Brill, W. Xiong, K. B. Lee, S. B. Ficarro, A. Crain, Y. Xu, A. Terskikh, E. Y. Snyder, S. Ding, Phosphoproteomic analysis of human embryonic stem cells, Cell Stem Cell **5** (2009) 204-213.

[39] J. Staerk, C. A. Lyssiotis, L. A. Medeiro, M. Bollong, R. K. Foreman, S. Zhu, M. Garcia, Q. Gao, L. C. Bouchez, L. L. Lairson, B. D. Charette, L. Supekova, J. Janes, A. Brinker, C. Y. Cho, R. Jaenisch, P. G. Schultz, Pan-Src family kinase inhibitors replace Sox2 during the direct reprogramming of somatic cells, Angew Chem Int Ed Engl **50** (2011) 5734-5736.

[40] Y. Yoshida, S. Yamanaka, Recent stem cell advances: induced pluripotent stem cells for disease modeling and stem cell-based regeneration, Circulation **122** (2010) 80-87.

[41] J. Silva, O. Barrandon, J. Nichols, J. Kawaguchi, T. W. Theunissen, A. Smith, Promotion of reprogramming to ground state pluripotency by signal inhibition, PLoS Biol **6** (2008) e253.

Chapter V

A comparative analysis of phosphotyrosine peptides enriched by means of immuno-affinity precipitation and metal-based affinity chromatography

Adja Zoumaro-Djayoon, Serena Di Palma, Mao Peng, Harm Post, Christian Preisinger, Shabaz Mohammed, Javier Munoz, Albert J.R. Heck

Biomolecular Mass Spectrometry and Proteomics. Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH, Utrecht (The Netherlands) and Netherlands Proteomics Center.

Manuscript in preparation

Abstract

Analysis of tyrosine phosphorylation by mass spectrometry (MS)-based proteomics remains challenging, due to the low occurrence of this post-translational modification representing about 1% of all phosphorylation events in mammalian systems. Conventional metal affinity based chromatography methods used to enrich phosphopeptides can produce over 10,000 phosphopeptides, but they are not suitable for the selective enrichment of Tyr phosphorylated peptides as they then become obscured by the co-enriched 99% Ser and Thr phosphorylated peptides. Therefore, a more targeted approach based on immuno-affinity chromatography has been introduced for the specific analysis of Tyr phosphorylated peptides. This method typically leads only to the detection of a few hundreds of phosphopeptides, albeit over 70% of those can be Tyr phosphorylated peptides. In this study, we evaluated and compared the pools of phosphotyrosine peptides enriched by phosphoTyr immuno-affinity enrichment and a multidimensional approach consisting of metal affinity based enrichment (Ti^{4+}-IMAC) followed by hydrophilic interaction liquid chromatography (HILIC) fractionation. The goal of this comparison is to see how much overlap there is in these pools, and to characterize peptide properties determining the selectivity of both methods. Our analysis suggests that both strategies are highly complementary and should both be used for the comprehensive study of tyrosine phosphorylation.

1. **Introduction**

Protein phosphorylation is a ubiquitous post-translational modification (PTM) involved in several key intracellular processes including metabolism, secretion, homeostasis, transcriptional and translational regulation, cellular signaling and cell-cell communication (*1, 2*). It is a reversible and dynamic modification that typically induces changes in conformation, activity and interaction partners of a protein (*3*). Phosphorylation is catalyzed by different protein kinases and mostly occurs on serine and threonine residues and, to a lesser extent, on tyrosine residues. The human genome is estimated to comprise 518 kinases genes. Tyrosine kinases form a distinct group within these enzymes and 90 genes enconding tyrosine kinases are found in the human genome (*4*). Mass spectrometry (MS)-based phosphoproteomics is to-date probably the most powerful tool to analyze large-scale protein phosphorylation events in a variety of biological samples (*5, 6*). However, significant analytical barriers still hamper the routine application of phosphoproteomics. Since protein phosphorylation is typically present at substoichiometric levels, the detection of phosphopeptides by MS can be impaired by low ionization efficiency and signal suppression in the presence of non-phosphorylated species (*7*). Therefore, the success of phosphoproteomic experiments greatly relies on the use of selective enrichment strategies, which decrease the number of unphosphorylated peptides in the sample, improving phosphopeptide identification by MS/MS sequencing.

There are several phosphoproteomic enrichment strategies which are typically applied following proteolytic digestion.The most widely applied method nowadays is based on chemical coordination by affinity chromatography including immobilized metal ion affinity chromatography (IMAC) (*8*) and metal oxide affinity chromatography (MOAC) such as porous titanium dioxide microspheres (TiO_2) (*9*). The mechansism of action of TiO_2 seems to work through its amphoteric ion exchange properties (*9*). With IMAC, metal ions (Fe^{3+}, Al^{3+}, Co^{2+}, Ga^{3+}, Ti^{4+}) are chelated to coated beads, forming a stationary phase. With these approaches negatively charged phosphopeptides can selectively bind to the positively charged metal ions in an acidified buffer, leaving the vast majority of non-phosphorylated peptides unbound (*10*). Phosphopeptides are then eluted at alkaline pH. One of the issues associated with these techniques is the unspecific binding of non-phosphorylated peptides in highly complex peptide mixtures (e.g. cell lysate digest), reduc-

ing the selectivity of the method.

Enrichment techniques are more efficient when combined with fractionation methods to decrease sample complexity prior to the enrichment. Several methods based on liquid chromatography (LC) are especially suited for this purpose, such as ion exchange chromatography or hydrophilic interaction liquid chromatography (HILIC). In large-scale phosphoproteomic studies, strong cation exchange (SCX) is by far the most common fractionation technique applied prior to either IMAC or TiO_2 chromatography (*11-15*). SCX separates peptides by their solution charge state and isolate distinct groups of phosphopeptides from the bulk of non-phosphorylated peptides since both peptide groups have different net charge states (*11*). At pH 2.7, most tryptic peptides (without missed-cleavages) carry a positive net charge state of +2 in solution and most phosphorylated peptides carry a positive net charge state of one (without missed-cleavages), as the phosphate group maintains a negative charge. Recently, HILIC has been employed in phosphoproteomic studies, either before or after the enrichment step (*16*). HILIC chromatography has proven to be well suited for polar compounds, such as several PTMs (e.g. phosphorylation, glycosylation, acetylation, and methylation) (*16-19*). More recently, HILIC has been used to fractionate phosphopeptide enriched samples in order to improve protein phosphoproteome coverage (*20*).

Usually, these methods mainly enrich for phospho-serine and -threonine peptides, as these are by far the most frequent phosphorylated peptide species. In fact, tyrosine phosphorylation occurs in mammalian system at a lower frequency and its investigation therefore relies almost exclusively on the availability of specific antibodies for targeted immunoaffinity purification. Selective anti-phosphotyrosine antibodies initially have been successfully employed for the enrichment of phosphotyrosine proteins from whole cell lysate digests (*21, 22*). Nowadays, immunoprecipitation (IP) is more often performed at the peptide level as this seems to be more efficient (*23, 24*). However, some of the major disadvantages of using IP strategies at this stage are the large amounts of protein starting material required, namely mg of protein sample, for an efficient enrichment and the batch-to-batch variability of the commercially available antibodies (*25, 26*).

In this study, we set out to evaluate and compare the pools of tyrosine phosphorylated peptides enriched by either immunoaffinity enrichment or with an extensive multidimensional approach based on Ti^{4+}-IMAC affinity en-

richment in combination with HILIC fractionation (2D strategy). The latter approach was recently used by our group in a phosphoproteomic analysis of the human chronic myelogenous leukemia cell line (K562) that resulted in the identification of more than 10,000 phosphorylation sites (*20*). The objective of our comparison is to determine if a method dependent preference exists towards the enrichment of specific types of tyrosine phosphorylated peptides. To this end, we performed several immuno-affinity enrichments on two independent cell lines (pervanadate treated HeLa and K562) to generate relatively large pools of enriched phosphotyrosine peptides. To systematically assess differences in phosphotyrosine peptide sequences between this approach and the aforementioned 2D strategy, we used the previously reported very large phosphopeptide datasets from Zhou et al. (2012) (*20*) and generated an additional large phosphopeptide dataset from pervanadate treated HeLa cells.

2. **Material and Methods**

Cell culture & treatment

HeLa cells and K562 cells were grown in either Dulbecco's modified Eagle's medium (HeLa) or RPMI medium (K562) supplemented with 10% fetal bovine serum, 10 mM Glutamine and 5% Penicillin/Streptomycin (Lonza, Belgium) at 37°C in the presence of 5% CO2. HeLa cells were treated with 1 mM pervanadate (prepared by incubating 1 mM orthovanadate with 1 mM hydrogen peroxide for 5 min) for 10 min immediately prior to harvesting. Cells were washed twice with ice-cold phosphate-buffered saline (Lonza, Belgium) after harvesting and freezed at -80°C until lysis.

Cell lysis & sample preparation

Cells were lysed on ice in 8M urea, 1 mM sodium orthovanadate and 1x *PhosSTOP* (Roche Diagnostics, Switzerland) in the presence of protease inhibitors (mini complete EDTA-free) (Roche Diagnostics, Germany). Protein concentration was determined using Bradford Assay. Total protein lysate of each cell line was reduced with dithiothreitol (DTT) at a final concentration of 10 mM at 56°C for 30 min and subsequently alkylated with 55 mM iodoacetamide for 20 min. Lysates were pre-digested with Lys-C (1:75) for 4 hours at 37°C. The samples were diluted 4-fold with 100 mM ammonium bicarbon-

ate and digested overnight with trypsin (1:100; Promega), at 37°C. The peptides were desalted using Sep-Pak C18 3 cc/200 mg cartridges (Waters, USA) as described previously (*27*). Briefly, the C18 material was conditioned with 100% acetonitrile and equilibrated with 0.05% acetic acid. The samples were then loaded on the column and washed with 0.05% acetic acid. The peptides were eluted using 80 % acetonitrile and the eluate lyophilized *in vacuo*.

Immuno-affinity purification of phosphotyrosine peptides

For the immuno-affinity enrichment, 5-8 mg of protein digest were re-suspended in 800 μL of cold immuno-affinity purification buffer consisting of 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, protease inhibitors (Roche Diagnostics, Germany) and 1% n-octyl-β-D-glucopyranoside (NOG) (Sigma, Germany) (*25*). The peptide mixture was agitated on a shaker for 30 minutes to dissolve the peptides thoroughly and the pH was adjusted to pH 7.4. The immuno-affinity purification was performed as described previously (*27*). In summary, the peptides were added to 45-50 μL slurry of pY99 antibodies beads (Santa Cruz biotechnology, USA) and the peptide-antibody mixture was incubated overnight at 4°C on a rotator. The beads were washed and bound peptides were eluted twice using 0.15% trifluoroacetic acid (TFA) (Sigma, Germany). Home-made columns with C18 material (Aqua™ C_{18}, 5 μm, Phenomenex, Torrance, CA) at the restricted end of a GELoader tip were used for desalting, as described elsewhere (*28*). Peptides were finally lyophilized and re-suspended in 40 μL of 10% formic acid for LC-MS/MS analysis (*27*).

Ti^{4+}-IMAC enrichment of phosphopeptides

Ti^{4+}-IMAC material was prepared and used essentially as described previously (*29*). Briefly, Ti^{4+}-IMAC material was loaded onto Gel-loader tip microcolumns using a C8 plug. The columns were pre-equilibrated with 2 × 30 μL of Ti^{4+}-IMAC loading buffer consisting of 80% ACN/6% TFA. Next, 6 mg of peptides were re-suspended in loading buffer and loaded onto the equilibrated gel-loader tip microcolumns (250μg protein digests per column). Columns were sequentially washed with 60 μL of loading buffer, followed by washing with 60 μL of 50% ACN/0.5% TFA containing 200 mM NaCl and an additional wash with 60 μL of 50% ACN/0.1% TFA. Bound peptides were

eluted with 20 μL of 5% ammonia into 20 μL of 10% formic acid. Oasis C-18 columns (Waters, Millford) were used for desalting.

HILIC separation & fractionation

ZIC-cHILIC separation was performed on a Dionex Ultimate LC system, using a vented column setup, as described previously (*30*) The HPLC was equipped with a 100 μm × 20 mm, 5 μm, 200 Å ZIC-cHILIC trap column and a ZIC-cHILIC 75 μm × 200 mm, 5 μm, 200 Å analytical column. All columns were packed in-house. The Ti^{4+}IMAC enriched phosphopeptides were desalted using the OASIS system (Waters Corporation, Milford, MA) and eluted directly with the HILIC loading buffer. Subsequently, the samples were injected on the HILIC trap column. Trapping was performed at 15 μL/min for 10 min at 100% buffer A (95% acetonitrile, 0.5% acetic acid and 5 mM ammonium acetate); elution was achieved with a gradient of 0–55% buffer B (5 mM ammonium acetate) over 40 min at a flow rate of 0.35 mL/min passively split to 300 nL/min. One-min fractions were collected during the elution in a 96-well plate, with each well containing 40 μL of 10% formic acid. Half of each of HILIC fractions was subjected to RP-LC-MS/MS analysis.

Mass spectrometry analysis

Briefly, 10-20 μL of the samples were injected in the LC system and peptides were delivered to a trap column (Aqua™ C_{18}, 5 μm (Phenomenex, Torrance, CA); 20 mm x 100 μm inner diameter) at 5 μL/min with the Agilent 1100 HPLC (Agilent Technologies, Waldbronn, Germany) and at a maximum pressure of 800 bar with the EASY-nanoLC 1000 (Thermo Fisher Scientific, Bremen, Germany) in 100% solvent A (0.1 M acetic acid in water). Subsequently, the peptides were eluted from the trap column onto an analytical column (ReproSil-Pur C_{18}-AQ, 3 μm (Dr. Maisch GmbH, Ammerbuch, Germany); 40 cm x 50-μm inner diameter) at 100 or 150 nL/min in a 3 h gradient from 0 to 50% solvent B (0.1 M acetic acid in 8:2 (v/v) acetonitrile/water). All columns used in the LC systems were packed in-house. The eluent was sprayed via distal coated emitter tips (New Objective, Cambridge, MA) (o.d., 360 μm; i.d., 20 μm; tip i.d., 10 μm) butt-connected to the analytical column. The tip was subjected to 1.7 kV. A 33 MΩ resistor was introduced between the high voltage supply and the electrospray needle to

reduce ion current. The mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS. Full-scan MS spectra (from *m*/*z* 350 to 1,500) were acquired in the Orbitrap with a resolution of 60,000 for the Orbitrap XL and classic, 30,000 for the Orbitrap velos and Elite or 35,000 for the Q-exactive. Up to ten most intense ions above the threshold of 500 counts were selected for fragmentation. CID fragmentation was performed when using the Orbitrap XL or classic instruments and HCD fragmentation was performed when using the Q-exactive. For the fragmentation using the Velos instrument, a decision tree method was used as described previously (*31*). Briefly, HCD was used for doubly charged peptides and ETD was performed for peptides with higher charge states. Fragment ions were analyzed in the Orbitrap when the precursor ion has a mass to charge ratio (*m*/*z*) less than 1000 and charge state of 4 and higher. For the fragmentation using the Elite instrument, a CID/ETD decision tree was applied. Briefly, CID was used for doubly charged peptides and ETD was performed for peptides with higher two charge states.

Data processing & analysis

Raw files were processed with Proteome Discoverer (version 1.3.0.339, Thermo Fisher Scientific, Bremen, Germany). The non-fragment filter was used for ETD spectra to filter the precursor ion mass from the MS/MS peak list. Database search was performed using a concatenated forward-decoy swissprot human database (v2010_12; 41,008 sequences; 22,806,364 residues) using Mascot (version 2.3.02, Matrix Science, UK) as the search engine. Carbamidomethylation of cysteines was set as fixed modification. Oxidized methionine and phosphorylation of tyrosine, serine, and threonine were all set as variable modifications. Trypsin was specified as the proteolytic enzyme, and up to two missed cleavages were allowed. The mass tolerance of the precursor ion was initially set to 50 ppm, and 0.6 Da for fragment ions in ion trap readout and 0.05 Da for fragment ions in Orbitrap readout. Mascot results were filtered afterwards with a 10 ppm precursor mass tolerance, a Mascot ion score > 20 and a minimum of 7 residues per peptide which results in a peptide FDR < 1%. For phosphorylation site localization, phosphoRS node in Proteome Discoverer was used (*32*). The phosphorylation sites with a phosphoRS probability score above 75% were used.

IceLogo sequence pattern analysis

For the sequence pattern analysis with IceLogo (*33*), the peptide sequences were centered at the tyrosine phosphorylation site and extended with 9 residues (±4 residues). The sequences were submitted to the IceLogo algorithm for sequence pattern extraction from the datasets. The antibody based enrichment datasets were used as the foreground sets and the 2D strategy enriched datasets were used as background sets for the comparison. The p-value for a sequence pattern was set as 0.01. Sequence logos were automatically generated by the IceLogo software.

3. **Results and discussion**

Datasets of phosphotyrosine peptides

The objective of this study was to compare two phosphopeptide enrichment methods and characterize peptide properties that determine the selectivity of these methods towards phosphotyrosine peptides. The schematic workflow used for this comparison is depicted in Figure 1. Basically, we generated four large pools of phosphotyrosine datasets; 1) a pTyr IP based dataset from Hela, 2) a very large phosphopeptide dataset containing a few pTyr peptides from Hela using Ti^{4+}-IMAC in combination with HILIC, 3) a pTyr IP based dataset from K562, and 4) a very large phosphopeptide dataset including a small number of pTyr peptides from K562 using Ti^{4+}-IMAC in combination with HILIC.

To generate a large phosphotyrosine dataset, we used pervanadate to block a large set of tyrosine phosphatases in HeLa cells. Proteins extracted from the cells were digested and we performed subsequently multiple phosphotyrosine peptide immuno-affinity enrichments using pY99 antibodies which were then analyzed by LC-MS/MS. We identified in total 2123 unique tyrosine phosphorylated peptides in these Hela cells (suppl. Table 1). We observed that 95% of these phosphotyrosine peptides were already identified in the first three IPs (Figure 2). Three additional experiments only added a 5% extra phosphotyrosine peptides to the dataset (Figure 2), indicating that a threshold is being reached with this method. A similar saturation effect was observed in analyzing the tyrosine phosphoproteome of the K562 cells (data not shown) but also earlier in a reported study on human embryon-

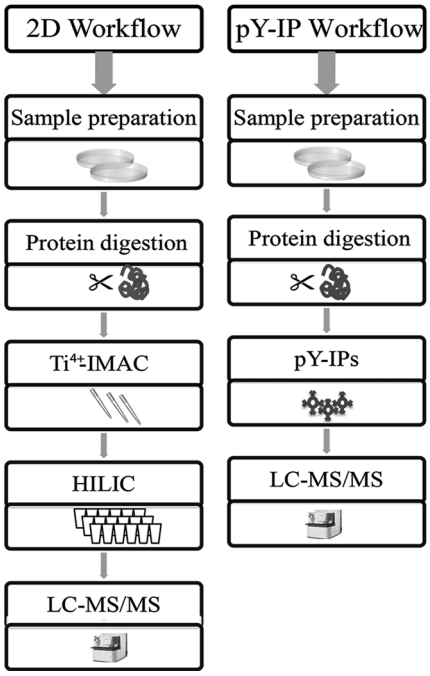


Figure 1: Experimental procedure. Pervanadate stimulated HeLa cells or K562 cells were cultured and harvested. The cells were lysed and extracted proteins were digested. For the 2D workflow, the cell lysate digests were used to enrich for phosphopeptides by Ti^{4+}-IMAC followed by a fractionation using HILIC prior to LC-MS/MS. For the generation of the phosphotyrosine datasets with the pY-IP workflow, pY99 anitbodies were used for immnunoprecipitations of phosphotyrosine peptides followed by LC-MS/MS analyses.

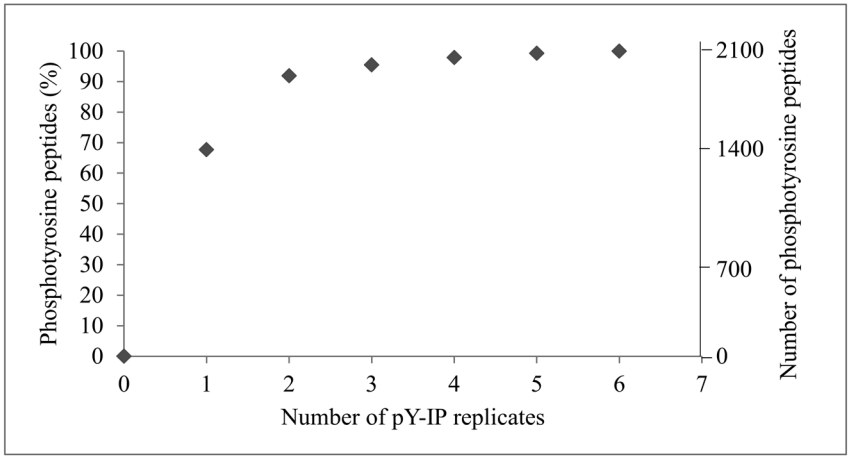
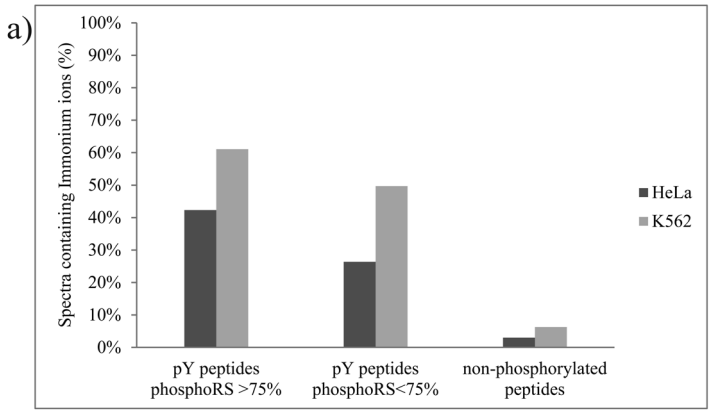


Figure 2: Relative and absolute number of unique phosphotyrosine peptides identified in each antibody based enrichment analysis. The HeLa phosphotyrosine peptides dataset was covered for 95% by 3 IPs, the number of unique phosphotyrosine peptides identified in the other enrichments was less than 5% of the total dataset, indicating that within this method a threshold has been reached.

ic stem cells (*34*). For the K562 cells, we were able to generate a dataset containing in total 1811 unique tyrosine phosphorylated peptides, using pTyr immuno-affinity enrichment method (suppl. Table 1). The saturation effects observed could also imply that certain anti-phosphotyrosine antibodies may have a bias toward particular types of phosphotyrosine peptides. The antibodies might selectively enrich for phosphotyrosine peptides possessing specific physicochemical properties that facilitate the peptide-antibody interaction.

We only considered peptides with phosphoRS site probabilities above 75% as confidently localized phosphorylation sites. In addition, to improve the confidence in the dataset, we determined the presence of phosphotyrosine immonium ion (m/z=216.043), which is a characteristic marker ion for the presence of a phosphorylated tyrosine residue in a collisional induced fragmentation spectrum (suppl. Table 2). Fragmentation spectra with high resolution and high mass accuracy such as those produced by HCD fragmentation were used to determine the presence of phosphotyrosine immnonium ion. We were able to detect this marker immonium ion in the HCD spectra of 40-60% of the phosphopeptides with a phosphoRS probability score higher than 75% for phosphorylation on the tyrosine residues (Figure 3a and 3b). Interestingly, 25-50% of phosphopeptides with a phosphoRS probability score below 75% for phosphorylation on the tyrosine residues had also the characteristic immonium ion in their spectra (Figure 3a and 3c). This is significant as less than 3-6% of the non-modified peptides might have these phosphotyrosine immonium ions in their HCD spectra. The latter results suggest that by applying a phosphoRS cut-off of 75%, many tyrosine phosphorylated peptides may be wrongly excluded from the datasets. Therefore,



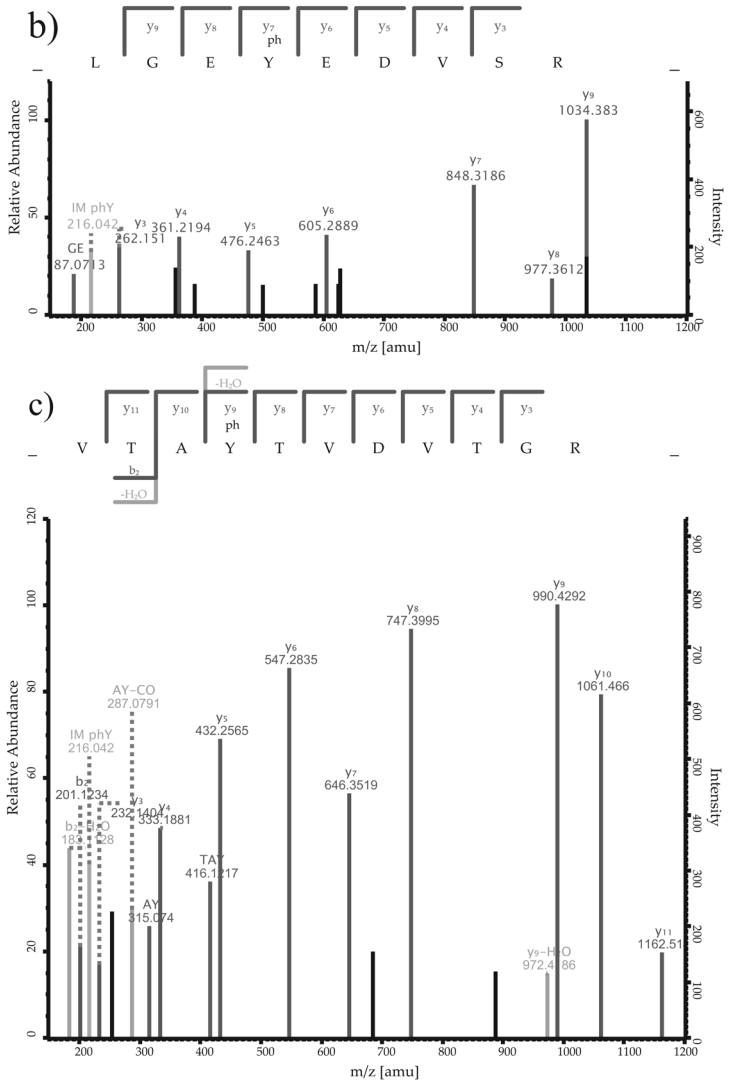


Figure 3: *Detection of phosphotyrosine signature immonium ions in MS/MS spectra of enriched phosphopeptides. a) Bar charts showing the proportion of spectra containing immonium ions (m/z= 216.043 ± 50ppm) in the antibody based enrichment datasets. Around 50% of all pTyr peptides display the characteristic immonium ion in peptides with a phosphoRS > 50, whereas this drops to < 5% for non-phosphorylated peptides b) and c) Illustrative examples of HCD spectra of peptides containing an immonium ion of phosphotyrosine (m/z= 216.043). The phosphosites were assigned to the phosphotyrosine residues however the peptide matching spectra displayed in (b) had a phosphoRS probability score above 75%; whereas the peptide matching spectra (c) had a phosphoRS probability score below 75%.*

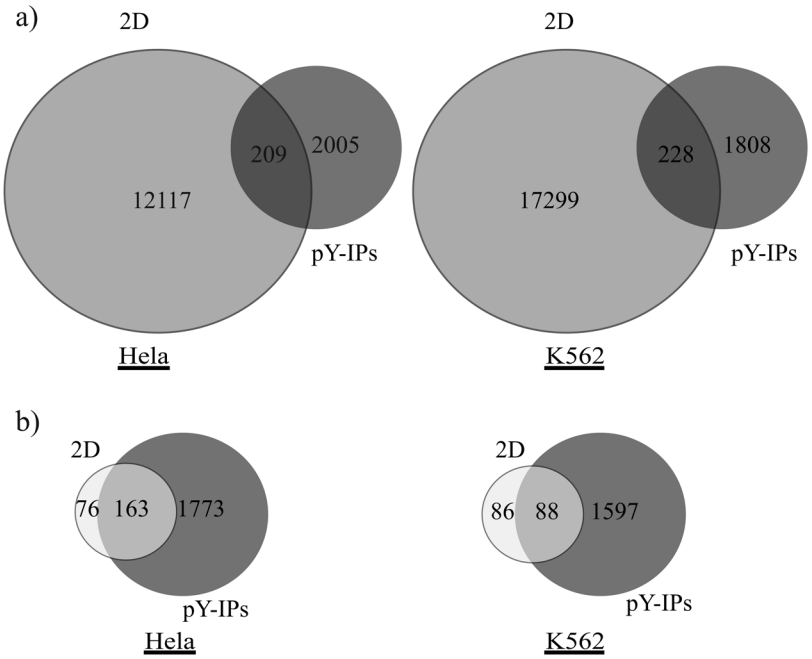


Figure 4: *Overlap in a) phosphopeptides and b) phosphotyrosine peptides in the datasets. a) Comparisons of the 2D based enrichment datasets and the antibody based enrichment datasets display a minimal overlap between a) the phosphopeptides and b) the phosphotyrosine peptides. This is apparent in the datasets detected both in Hela (left-side) and K562 (right-side).*

the presence of phosphotyrosine immonium ions in MS/MS spectra should be considered when assigning phosphorylation sites, although we also do not observe the immonium ion in all HCD spectra of bona fide pTyr peptides. Clearly, accounting phosphotyrosine immonium ions in algorithms for phosphorylation site assignments could improve the phosphosite assigments in peptides containing next to tyrosine, serine or threonine residues.

For our comparative studies, we also generated very large phosphopeptide datasets performing Ti^{4+}-IMAC-HILIC (2D) enrichment prior to the LC-MS/MS analysis (Figure 1). With this approach, we were able to identify 14,991 unique phosphopeptides from pervanadate treated HeLa cells (suppl. Table 1). Evidently, when compared to the antibody based enrichment dataset, this large dataset included only a low number of phosphotyrosine peptides, i.e. 330 (2.20%). Using a similar approach we gathered also a previously re-

ported pool of 19,733 unique phosphopeptides from K562 cells (suppl. Table 1). This dataset included 248 (1.26%) phosphotyrosine peptides.

For each cell line, we next determined the number of phosphopeptides (Figure 4a) and phosphotyrosine peptides (Figure 4b) that were identified in common by the two approaches (IMAC-based and IP-based). For the data comparison we apply a phosphoRS cut-off of 75% to all four datasets. We only found 1.5% and 1.2% of the phosphopeptides in common in K562 datasets and in HeLa datasets, respectively (Figure 4a). We further observed a very small overlap in phosphotyrosine peptides of 8% in Hela and 5% in K562 cells (Figure 4b). Although both the IP datasets in Hela and K562 cells generated a substantially larger number of pTyr peptides, the IMAC-based approach still is able to add a significant number of pTyr peptides, possibly representing peptides that will never be detected by using IPs only. This data indicates that there indeed may be a different specificity in each enrichment method toward certain classes of phosphopeptides.

Physicochemical properties of the phosphotyrosine peptides in the two different enrichment approaches

In order to evaluate the physicochemical properties of phosphotyrosine peptides in each dataset, we selected five parameters that are relevant for antibody-antigen interaction: hydrophobicity, isoelectric point, flexibility, antigenicity and the position of the phosphotyrosine residue within the sequence (suppl. Table 3). The results from these analyses are shown in Figure 5. We observed subtle but significant differences in the position of phosphotyrosine residues within the peptide sequences and in their hydrophobicity profiles ($p<0.01$, Fisher exact test). In our analyses, phosphotyrosine peptides enriched by the 2D approach had often tyrosine phosphorylated residues positioned toward the N-terminus of the peptide sequences when compared to phosphotyrosine peptides identified using specific antibodies (Figure 5a). In contrast, peptides detected using immuno-affinity enrichment had phosphotyrosine residues located around the center of the sequences. We also noticed that the dispersion in the data was much larger in the 2D enrichment datasets. In other words, in the 2D datasets, the position

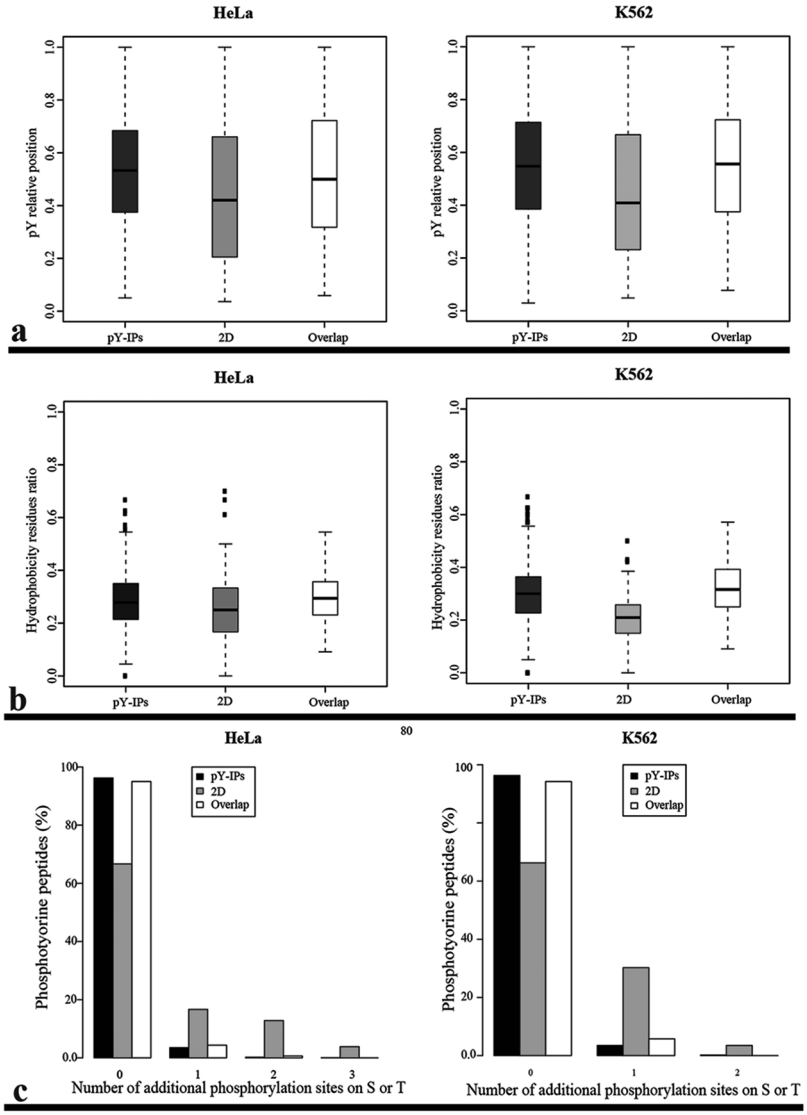


Figure 5: *Analysis of potential differences in physicochemical properties of enriched pTyr peptides (with phosphoRS above 75%). Each graph represents one property. Although the differences observed are subtle, they are in accordance when analyzing phosphotyrosine peptides from two different sources; Hela on the left and K562 on the right.*

a) Position of the phosphotyrosine residue in a peptide sequence; the pY position ratio on the y-axis is calculated by dividing the position of the phosphotyrosine residue by the number of residues in a peptide sequence. 0 represents the N-terminus and 1 represents the C-terminus of the peptide.

b) Hydrophobicity; the ratio on the y-axis is calculated by dividing the number of hydrophobic residues (A,V,I,L,M,F,Y,W) by the total number of residues.

c) Phosphorylation site distribution in the phosphotyrosine peptide sequences; except the phosphotyrosine peptides several peptides in the 2D enrichment based datasets contain additional phosphosites.

of the phosphotyrosine residues greatly varies in the peptide sequences suggesting that the interaction between the Ti^{4+} and the phosphate group on the peptides does not require a centered position of the phosphotyrosine residue. However, the position of the phosphotyrosine residue in the sequence might be important for the peptide-antibody interaction and consequently for the enrichment and detection of peptides.
Furthermore, we found that phosphotyrosine peptides enriched by antibodies exhibit a significantly higher hydrophobicity than their 2D enriched counterparts ($p<0.01$, Fisher exact test) (Figure 5b). Antigen hydrophobicity is known to affect the binding affinity towards antibodies (*35*). The variance observed in hydrophobicity suggests that anti-phosphotyrosine antibodies preferentially bind to more hydrophobic peptides whereas Ti^{4+}-IMAC interacts with more hydrophilic phosphotyrosine peptides. The hydrophilic phosphate group renders a peptide hydrophilic, indicating that the more phosphate groups a peptide possesses the more hydrophilic the peptide will become. To confirm our hypothesis, we estimated the number of phosphotyrosine peptides containing multiple phosphorylated residues and observed a great number of multiply phosphorylated peptides in the 2D datasets (Figure 5c).

IceLogo sequence patterns

To evaluate the occurrence of specific amino acids in peptide sequences of the distinct datasets, we used the well-established IceLogo algorithm (*33*). This software is used to analyze and visualize consensus patterns in aligned peptide sequences. We compared phosphotyrosine peptides obtained by the antibody based approach with phosphotyrosine peptides detected through the 2D based strategy using the latter as the reference set. The sequence patterns obtained with IceLogo revealed that hydrophobic residues (e.g. (A,V,I,L,M,Y)) were enriched in the phosphotyrosine peptides present in the antibody-based datasets confirming the hydrophobic characteristic of these peptides mentioned earlier. Furthermore, in the IceLogo sequence patterns, an enrichment of acidic residues was observed in the peptides sequences of the 2D based enrichment datasets (Figure 6). To support the outcome of the sequence patterns, we evaluated the presence of acidic residues (D or E) in the peptide sequences. A clear enrichment of tyrosine phosphorylated peptides containing more than 7 acidic residues was observed in the

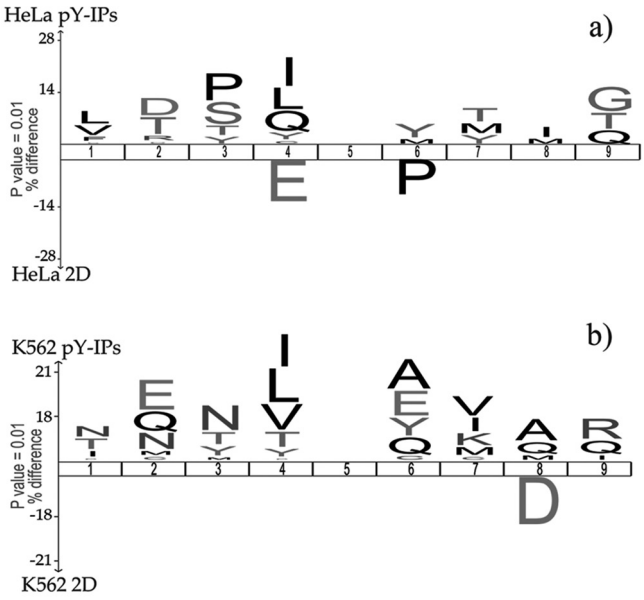


Figure 6: *Potential bias in motifs enriched by pTyr-antibodies and Ti-IMAC as displayed from IceLogo sequence patterns. Clearly acidic residues were enriched in the peptide sequences of the 2D enrichment based datasets. In the antibody based enrichment datasets, we mainly observed hydrophobic residues enriched in the peptides sequences (p-value =0.01), especially on position -1.*

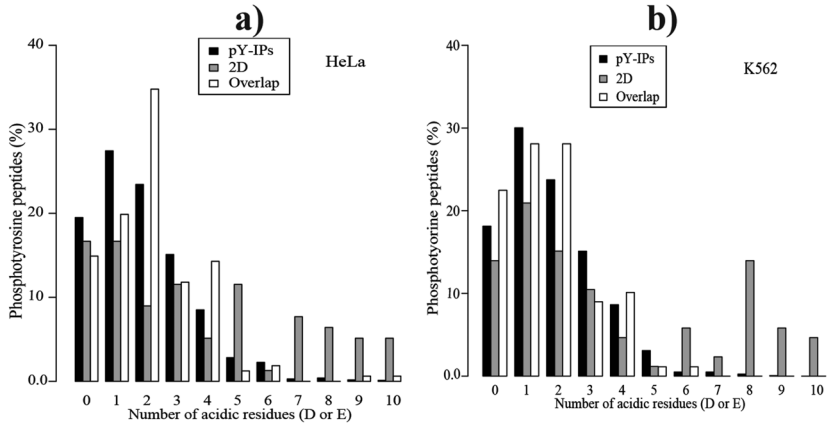


Figure 7: *Distribution of acidic residues in the pools of detected phosphotyrosine peptides. Highly acidic peptides containing more than 7 acidic residues were clearly enriched in the 2D enrichment based datasets.*

datasets generated with the 2D enrichment method (Figure 7).
Protein kinases select serine, threonine or tyrosine residues based, in part,

on the primary sequence surrounding the target residue (*36*). It has been shown that acidic residues in the vicinity of tyrosine residues may be important for the recognition of the tyrosine residues by kinases (*36*). More recently, a novel phosphotyrosine binding domain named HYB domain, present in the E3 ubiquitin ligase Hakai was shown to bind phosphotyrosine residues surrounded by acidic amino acids (*37*). These phosphotyrosine peptide sequences were found in several Src substrates (e.g. E-cadherin, cortactin and DOK1) (*37*). The fact that acidic amino acids flanking tyrosine phosphorylated residues are important for protein-protein interaction highlights the importance of the identification of related acidic peptides containing a tyrosine phosphorylated residue. For the identification of this category of phosphotyrosine peptides, Ti^{4+}-IMAC enrichment strategy might be a more suitable approach.

4. **Conclusion**

The use of antibodies is still to date the most suitable approach for the enrichment of large number of phosphotyrosine peptides and remains essential for the analysis of tyrosine phosphorylation. However, the data from our comparative study reveals that phosphotyrosine antibodies might have lower affinity for phosphotyrosine peptides with low hydrophobicity. Moreover, tyrosine phosphorylated peptides containing additional phosphorylation sites and a high number of acidic residues were under-represented when using the immunoaffinity enrichment method. Therefore, comprehensive phosphotyrosine peptide analysis by LC-MS/MS might require a combination of at least both these enrichment methods.

5. **Acknowledgements**

We want to thank Houjiang Zhou, Marco Hennrich and Shabaz Mohammed for generating part of the data used for analysis here. We thank members of the Biomolecular Mass Spectrometry and Proteomics Group, particularly Christian Frese and Maarten Altelaar for their help. This work was in part supported by the PRIME-XS project, Grant Agreement Number 262067, funded by the European Union Seventh Framework Program. The Netherlands Proteomics Centre, embedded in The Netherlands Genomics Initiative is acknowledged for funding.

Supplementary data is available at **http://tinyurl.com/azdchapter5**

6. References

[1] J. D. Graves, E. G. Krebs, Protein phosphorylation and signal transduction, Pharmacol Ther **82** (1999) 111-121.

[2] R. Patarca, Protein phosphorylation and dephosphorylation in physiologic and oncologic processes, Crit Rev Oncog **7** (1996) 343-432.

[3] T. Hunter, 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** (1998) 583-605.

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

[5] D. T. McLachlin, B. T. Chait, Analysis of phosphorylated proteins and peptides by mass spectrometry, Current Opinion in Chemical Biology **5** (2001) 591-602.

[6] R. Aebersold, D. R. Goodlett, Mass Spectrometry in Proteomics, Chemical Reviews **101** (2001) 269-296.

[7] M. Mann, S. E. Ong, M. Gronborg, H. Steen, O. N. Jensen, A. Pandey, Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome, Trends Biotechnol **20** (2002) 261-268.

[8] L. Andersson, J. Porath, Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography, Analytical biochemistry **154** (1986) 250-254.

[9] M. W. Pinkse, S. Mohammed, J. W. Gouw, B. van Breukelen, H. R. Vos, A. J. Heck, Highly robust, automated, and sensitive online TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster, J Proteome Res **7** (2008) 687-697.

[10] J. Fila, D. Honys, Enrichment techniques employed in phosphoproteomics, Amino Acids (2011).

[11] S. A. Beausoleil, M. Jedrychowski, D. Schwartz, J. E. Elias, J. Villen, J. Li, M. A. Cohn, L. C. Cantley, S. P. Gygi, Large-scale characterization of HeLa cell nuclear phosphoproteins, Proc Natl Acad Sci U S A **101** (2004) 12130-12135.

[12] J. C. Trinidad, C. G. Specht, A. Thalhammer, R. Schoepfer, A. L. Bur-

lingame, Comprehensive Identification of Phosphorylation Sites in Postsynaptic Density Preparations, Molecular & Cellular Proteomics **5** (2006) 914-922.
[13] J. Villén, S. A. Beausoleil, S. A. Gerber, S. P. Gygi, Large-scale phosphorylation analysis of mouse liver, Proceedings of the National Academy of Sciences **104** (2007) 1488-1493.
[14] J. Villen, S. P. Gygi, The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry, Nat. Protocols **3** (2008) 1630-1638.
[15] B. Zhai, J. Villén, S. A. Beausoleil, J. Mintseris, S. P. Gygi, Phosphoproteome Analysis of Drosophila melanogaster Embryos, Journal of proteome research **7** (2008) 1675-1682.
[16] D. E. McNulty, R. S. Annan, Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection, Mol Cell Proteomics **7** (2008) 971-980.
[17] D. Kolarich, P. H. Jensen, F. Altmann, N. H. Packer, Determination of site-specific glycan heterogeneity on glycoproteins, Nat. Protocols **7** (2012) 1285-1298.
[18] H. Lindner, B. Sarg, C. Meraner, W. Helliger, Separation of acetylated core histones by hydrophilic-interaction liquid chromatography, J Chromatogr A **743** (1996) 137-144.
[19] B. Sarg, E. Koutzamani, W. Helliger, I. Rundquist, H. H. Lindner, Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging, J Biol Chem **277** (2002) 39195-39201.
[20] H. Zhou, S. Di Palma, C. Preisinger, M. Peng, A. N. Polat, A. J. R. Heck, S. Mohammed, Towards a comprehensive characterization of a human cancer cell phosphoproteome, Journal of proteome research (2012).
[21] B. Blagoev, I. Kratchmarova, S. E. Ong, M. Nielsen, L. J. Foster, M. Mann, A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling, Nat Biotechnol **21** (2003) 315-318.
[22] A. Pandey, A. V. Podtelejnikov, B. Blagoev, X. R. Bustelo, M. Mann, H. F. Lodish, Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors, Proc Natl Acad Sci U S A **97** (2000) 179-184.
[23] S. Bergstrom Lind, M. Molin, M. M. Savitski, L. Emilsson, J. Astrom, L. Hedberg, C. Adams, M. L. Nielsen, A. Engstrom, L. Elfineh, E. Andersson,

R. A. Zubarev, U. Pettersson, Immunoaffinity enrichments followed by mass spectrometric detection for studying global protein tyrosine phosphorylation, J Proteome Res **7** (2008) 2897-2910.
[24] K. Rikova, A. Guo, Q. Zeng, A. Possemato, J. Yu, H. Haack, J. Nardone, K. Lee, C. Reeves, Y. Li, Y. Hu, Z. Tan, M. Stokes, L. Sullivan, J. Mitchell, R. Wetzel, J. Macneill, J. M. Ren, J. Yuan, C. E. Bakalarski, J. Villen, J. M. Kornhauser, B. Smith, D. Li, X. Zhou, S. P. Gygi, T. L. Gu, R. D. Polakiewicz, J. Rush, M. J. Comb, Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer, Cell **131** (2007) 1190-1203.
[25] P. J. Boersema, L. Y. Foong, V. M. Ding, S. Lemeer, B. van Breukelen, R. Philp, J. Boekhorst, B. Snel, J. den Hertog, A. B. Choo, A. J. Heck, In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling, Mol Cell Proteomics **9** (2010) 84-99.
[26] H. Johnson, F. M. White, Toward quantitative phosphotyrosine profiling in vivo, Seminars in Cell & Developmental Biology **23** (2012) 854-862.
[27] A. D. Zoumaro-Djayoon, A. J. Heck, J. Munoz, Targeted analysis of tyrosine phosphorylation by immuno-affinity enrichment of tyrosine phosphorylated peptides prior to mass spectrometric analysis, Methods (2011).
[28] J. Gobom, E. Nordhoff, E. Mirgorodskaya, R. Ekman, P. Roepstorff, Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry, J Mass Spectrom **34** (1999) 105-116.
[29] H. Zhou, M. Ye, J. Dong, G. Han, X. Jiang, R. Wu, H. Zou, Specific Phosphopeptide Enrichment with Immobilized Titanium Ion Affinity Chromatography Adsorbent for Phosphoproteome Analysis, Journal of proteome research **7** (2008) 3957-3967.
[30] S. Di Palma, S. Mohammed, A. J. R. Heck, ZIC-cHILIC as a fractionation method for sensitive and powerful shotgun proteomics, Nat. Protocols **7** (2012) 2041-2055.
[31] C. K. Frese, A. F. Altelaar, M. L. Hennrich, D. Nolting, M. Zeller, J. Griep-Raming, A. J. Heck, S. Mohammed, Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos, J Proteome Res **10** (2011) 2377-2388.
[32] T. Taus, T. Köcher, P. Pichler, C. Paschke, A. Schmidt, C. Henrich, K.

Mechtler, Universal and Confident Phosphorylation Site Localization Using phosphoRS, Journal of proteome research **10** (2011) 5354-5362.
[33] N. Colaert, K. Helsens, L. Martens, J. Vandekerckhove, K. Gevaert, Improved visualization of protein consensus sequences by iceLogo, Nat Meth **6** (2009) 786-787.
[34] J. Boekhorst, P. J. Boersema, B. B. Tops, B. van Breukelen, A. J. Heck, B. Snel, Evaluating experimental bias and completeness in comparative phosphoproteomics analysis, PLoS One **6** (2011) e23276.
[35] C. F. Barbas, 3rd, A. Heine, G. Zhong, T. Hoffmann, S. Gramatikova, R. Bjornestedt, B. List, J. Anderson, E. A. Stura, I. A. Wilson, R. A. Lerner, Immune versus natural selection: antibody aldolases with enzymic rates but broader scope, Science **278** (1997) 2085-2092.
[36] T. Hunter, J. A. Cooper, Protein-tyrosine kinases, Annu Rev Biochem **54** (1985) 897-930.
[37] M. Mukherjee, S. Y. Chow, P. Yusoff, J. Seetharaman, C. Ng, S. Sinniah, X. W. Koh, N. F. M. Asgar, D. Li, D. Yim, R. A. Jackson, J. Yew, J. Qian, A. Iyu, Y. P. Lim, X. Zhou, S. K. Sze, G. R. Guy, J. Sivaraman, Structure of a novel phosphotyrosine-binding domain in Hakai that targets E-cadherin, EMBO J **31** (2012) 1308-1319.

Summary & Outlook

Summary

Protein phosphorylation plays a key role in cellular signaling. Mass spectrometry (MS)-based proteomics is an effective strategy to analyze global protein phosphorylation events in cells. The main objective of this thesis is to increase our knowledge of the phosphorylation networks regulating pluripotent stem cells. To achieve this goal, we introduced new and formerly applied developed methods for phosphoproteomic analysis. In C**hapter 1,** general aspects of pluripotent stem cells are described focusing on our current understanding of the molecular mechanisms underlying the pluripotent state. In addition, an overview of the proteomic tools and methods used to analyze the proteome (i.e. all proteins present in a cell or tissue at a specific time) are given. We used several of these mass spectrometric tools and techniques in the phosphoproteomic studies described in this thesis.

In **chapter 2,** we analyzed the phosphorylation events induced by fibroblast growth factor-2 (FGF-2) signaling in human embryonic stem cells (hESCs). In this chapter, we combined strong cation exchange (SCX) chromatography and titanium dioxide (TiO_2) for the enrichment of phosphopeptides. Dimethyl labeling was then used for accurate relative quantification of phosphopeptides. Since FGF-2 signaling is essential for the culture of undifferentiated hESCs, a more in-depth study of this process would enable an assessment of the influence of this growth factor on hESCs. We identified and quantified 3,261 unique proteins from which 1,064 proteins were found to be phosphorylated at one or more residues. Around 40% of the proteins showed differential phosphorylation upon FGF-2 treatment. Of these, several protein members belong to the canonical pathways involved in pluripotency and self-renewal (e.g. Wnt, PI3K/AKT), hESC-associated proteins such as SOX2, RIF1, SALL4, DPPA4, DNMT3B and a number of proteins that are target genes of the pluripotency transcription factors SOX2, OCT4 and NANOG. This result provides new insights into how FGF-2 assists in maintaining the undifferentiated state of hESCs. Using the phosphoproteomic strategy mentioned above, we predominantly detected serine and threonine phosphorylated peptides. The low number of tyrosine phosphorylated peptides identified in global phosphoproteomic studies probably reflects the relative low abundance and low occurrence of this modification. Therefore, to investigate tyrosine phosphorylation, a more targeted approach based

on immuno-affinity purification is usually preferred. In **chapter 3,** we describe an approach that uses phospho-specific antibodies to enrich tyrosine phosphorylated peptides from a complex proteolytic digest followed by LC-MS/MS analysis. By using this method, a dataset was generated from pervanadate treated HeLa cells containing around 1000 unique phosphotyrosine peptides. The comparison of this dataset with three popular repositories of tyrosine phosphorylation sites, i.e Human Protein Reference Database (HRPD), PhosphoSitePlus and phospho.ELM, showed that novel tyrosine phosphorylation sites were still identified in our dataset indicating the large landscape of this modification. FGF-2 triggers the activation of receptor tyrosine kinases (i.e. FGFR) to maintain hESCs and human induced pluripotent stem cells (hiPSCs) undifferentiated, implying a key role for tyrosine phosphorylation in pluripotency Thereby, we used the method described in chapter 3 in combination with stable isotope dimethyl labeling to compare the tyrosine phosphorylation levels in hiPSCs and hESCs. Comparative studies on hiPSCs and hESCs are necessary to determine the similarities and differences of these cells at the molecular level, including tyrosine phosphorylation. These studies aim to improve our understanding of the molecular mechanisms underlying pluripotent stem cell fate, ultimately facilitating their application in research and therapy. Results derived from our comparisons of hiPSCs and hESCs are described in **chapter 4**. Briefly, we found that 31 proteins out of 93 showed different levels of tyrosine phosphorylation in hiPSCs and hESCs. These proteins include tyrosine-directed kinases such as FYN, LYN, LCK, EPHA2, FGFR1 and cytoskeletal associated proteins, which are implicated in the maintenance of the pluripotent state (e.g. TJP2, F11R, CDK5, and PTK2). This result suggests differences in tyrosine signaling between hiPSCs and hESCs. Follow-up experiments will be required to understand the specific role of the different phosphorylation levels of these proteins in hiPSCs and hESCs.

Comparison of these results with a recent global phosphoproteomic analysis of hESCs and hiPSCs by Phanstiel *et al*. (2011) showed poor overlap in the tyrosine phosphorylated peptides. To further investigate this observation, we performed a systematic comparison of tyrosine phosphorylation using metal affinity-based enrichment (typically used in global phosphoproteomic approaches as in Phanstiel *et al*.) and immuno-affinity purification. The preliminary results of these investigations are described in **Chapter 5**. In gener-

al, the overlap between the two methods was minor. Phosphotyrosine peptides identified by the metal affinity approach (i.e. Ti^{4+}-IMAC followed by HILIC) consisted of a significant number of peptides containing multiply phosphorylated sites. Comparison with the phosphotyrosine peptide datasets derived from the immuno-affinity enrichment, showed a bias towards highly acidic phosphotyrosine peptides with fewer hydrophobic phosphotyrosine peptides. In contrast, the immuno-affinity strategy typically enriched for more hydrophobic phosphotyrosine peptides and peptides with centered phosphotyrosine residues in the sequences. Therefore, comprehensive phosphotyrosine peptide analysis by LC-MS/MS might require complementary enrichment methods to increase the number of identifications.

Future perspectives

A full understanding of the molecular mechanisms involved in pluripotent stem cells will allow precise manipulation of these cells for their application in therapies. In the past few years, several studies have been undertaken to unravel those signaling networks regulating the pluripotent stem cell state. However, many questions still remain to be answered. We know that growth factor signaling, including fibroblast growth factor-2 (FGF-2) and Activin A, play an important role in the regulation of stem cell self-renewal (*1*). Exogenous FGF-2 is typically used to effectively maintain human embryonic stem cells (hESCs) in an undifferentiated state for long term culture periods (*1-3*). The results described in chapter 2 together with a related study by Ding *et al.* (*4*) showed that many receptor tyrosine kinases and downstream effectors of these kinases are differentially phosphorylated upon FGF-2 stimulation. However, it is still unclear if FGF-2 signals exclusively through FGF receptors or activates additional cell surface proteins. More studies focused on receptor activation using for instance cell imaging based approaches, will be required which could shed some light into how activation of the FGF-2 dose-dependent signaling pathway occurs. It has been shown that low levels (< 10 ng/mL) of exogenous FGF-2 maintains self-renewal by mild activation of ERK, whereas high levels (> 50 ng/mL) activate the PI3K/AKT pathway in addition to ERK without increasing the activation of the MAPK/ERK pathway (*5*). However, the PI3K/AKT pathway seems to suppress the MAPK/ERK pathway. PI3K/AKT signaling retains ERK activity at basal lev-

els to maintain self-renewal since the elevated activity of ERK is required in the early differentiation (*6, 7*). How FGF-2 controls MAPK and PI3K signaling requires further investigation. In a similar manner, the level of activation of SMAD2/3 defines its role in either self-renewal or early differentiation. SMAD2/3 activates a different set of target genes in pluripotent cells and differentiating cells (*8-10*). Through Activin A and other TGF-β family members, SMAD2/3 is known to activate the transcription of NANOG (*11*). Understanding signaling pathways is extremely challenging due to, for instance, a complex cross-talk between them as in the case of ERK and SMAD2/3. The general vies emerging is that pathways are not at all linear, but more multi-dimensional networks. It is often difficult to define how changes in a single pathway influence other pathways because reported studies often include only the analysis of a specific linear signaling pathway. Ultimately, full understanding of the molecular mechanisms that regulate pluripotent stem cell fate would require the development of new technologies that can measure signaling activities in real time in homogeneous cell populations. Depending on the cell cycle stage, cell-to-cell differences might be observed at transcriptional and protein levels. Heterogeneity in pluripotent stem cells caused by genetic abnormalities, spontaneous mutations and reprogramming in the case iPSCs has been reported (*12-14*) . Therefore, analysis of more homogeneous cell populations can be more instructive on how cells respond to signals and other environmental triggers. Further, the involvement of signaling pathways and their impact in cellular reprogramming has not yet been adequately addressed. Important questions to answer are, for instance, which signaling pathways are important during the reprogramming process of somatic cells? Are these pathways also important in the pluripotent state? And are the same signaling pathways involved in the reprogramming processes of distinct somatic cells?

A system-level analysis is important to characterize, in standardized culture conditions, the processes that regulate stem cell self-renewal, pluripotency and differentiation towards a specific cell type. Such an approach would comprise the identification and quantification of proteins and their PTMs, deciphering protein interaction networks and analysis of the transcriptional circuitry. Bearing this in mind, an integrative approach using high throughput methods including proteomics, transcriptomics and epigenomics would provide an unprecedented molecular view of these processes. Ultimately, bioinformatics tools that can integrate inputs from these strategies

would also be necessary. Regarding PTMs, it is evident that protein phosphorylation plays an important role in the regulation of many biological processes in cells including pluripotent stem cells. In general, large scale phosphoproteomic analyses using MS are performed to identify phosphorylation sites. However, the functions of the vast majority of identified phosphorylation sites are not known. Follow-up, much lower through-put, experiments such as site-specific mutations are required to determine the importance of individual phosphorylation sites as described for the transcription factor SOX2 in HUES-7 cells (*15*). Phosphorylation of SOX2 at three serine residues leads to its SUMOylation and thereby inhibits its DNA binding properties. Strong evidence supports the notion that other modifications are also required to modulate protein function. To facilitate the understanding of molecular mechanisms, more attention should be centered on the analysis of other PTMs. Besides phosphorylation, MS is starting to be used for the identification of other types of PTMs (e.g. glycosylation, acetylation, ubiquitination, sumoylation), and it has been shown that many proteins are multiply modified by different PTMs (e.g. phosphate group and glycosyl group) (*16-18*). The identification of individual proteins that form a proteome and their PTMs is important but is not sufficient to fully understand the molecular mechanisms regulating ESC properties. Many cellular processes are carried out through protein-protein interactions since proteins generally function as members of larger complexes. The characterization of protein interactions is therefore essential to unravel protein interacting networks that regulate the pluripotent cell state. Techniques such as yeast-two-hybrid (*19*) and affinity purification (*20*) have been used for protein-protein interaction studies. Affinity purification has been employed in combination with MS to analyze protein-protein interactions in pluripotent stem cells (*21*). The interacting partners of SOX2 (*22*), OCT4 (*23, 24*), and NANOG (*25*) are already characterized and have been reported for ESCs. Similar studies should be undertaken in iPSCs to identify the interactome of these transcription factors and determine the similarity of the protein networks in iPSCs and ESCs.

References

[1] L. Vallier, M. Alexander, R. A. Pedersen, Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic

stem cells, J Cell Sci **118** (2005) 4495-4509.

[2] M. E. Levenstein, T. E. Ludwig, R. H. Xu, R. A. Llanas, K. VanDenHeuvel-Kramer, D. Manning, J. A. Thomson, Basic fibroblast growth factor support of human embryonic stem cell self-renewal, Stem Cells **24** (2006) 568-574.

[3] R. H. Xu, R. M. Peck, D. S. Li, X. Feng, T. Ludwig, J. A. Thomson, Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells, Nat Methods **2** (2005) 185-190.

[4] V. M. Ding, P. J. Boersema, L. Y. Foong, C. Preisinger, G. Koh, S. Natarajan, D. Y. Lee, J. Boekhorst, B. Snel, S. Lemeer, A. J. Heck, A. Choo, Tyrosine phosphorylation profiling in FGF-2 stimulated human embryonic stem cells, PLoS One **6** (2011) e17538.

[5] A. M. Singh, D. Reynolds, T. Cliff, S. Ohtsuka, A. L. Mattheyses, Y. Sun, L. Menendez, M. Kulik, S. Dalton, Signaling network crosstalk in human pluripotent cells: a Smad2/3-regulated switch that controls the balance between self-renewal and differentiation, Cell Stem Cell **10** (2012) 312-326.

[6] F. Lanner, J. Rossant, The role of FGF/Erk signaling in pluripotent cells, Development **137** (2010) 3351-3360.

[7] J. Nichols, J. Silva, M. Roode, A. Smith, Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo, Development **136** (2009) 3215-3222.

[8] S. Brown, A. Teo, S. Pauklin, N. Hannan, C. H. Cho, B. Lim, L. Vardy, N. R. Dunn, M. Trotter, R. Pedersen, L. Vallier, Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors, Stem Cells **29** (2011) 1176-1185.

[9] A. B. McLean, K. A. D'Amour, K. L. Jones, M. Krishnamoorthy, M. J. Kulik, D. M. Reynolds, A. M. Sheppard, H. Liu, Y. Xu, E. E. Baetge, S. Dalton, Activin a efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed, Stem Cells **25** (2007) 29-38.

[10] A. K. Teo, S. J. Arnold, M. W. Trotter, S. Brown, L. T. Ang, Z. Chng, E. J. Robertson, N. R. Dunn, L. Vallier, Pluripotency factors regulate definitive endoderm specification through eomesodermin, Genes Dev **25** (2011) 238-250.

[11] R.-H. Xu, T. L. Sampsell-Barron, F. Gu, S. Root, R. M. Peck, G. Pan,

J. Yu, J. Antosiewicz-Bourget, S. Tian, R. Stewart, J. A. Thomson, NANOG Is a Direct Target of TGFβ/Activin-Mediated SMAD Signaling in Human ESCs, Cell Stem Cell **3** (2008) 196-206.

[12] T. Graf, M. Stadtfeld, Heterogeneity of embryonic and adult stem cells, Cell Stem Cell **3** (2008) 480-483.

[13] A. Maitra, D. E. Arking, N. Shivapurkar, M. Ikeda, V. Stastny, K. Kassauei, G. Sui, D. J. Cutler, Y. Liu, S. N. Brimble, K. Noaksson, J. Hyllner, T. C. Schulz, X. Zeng, W. J. Freed, J. Crook, S. Abraham, A. Colman, P. Sartipy, S. Matsui, M. Carpenter, A. F. Gazdar, M. Rao, A. Chakravarti, Genomic alterations in cultured human embryonic stem cells, Nat Genet **37** (2005) 1099-1103.

[14] K. H. Narsinh, N. Sun, V. Sanchez-Freire, A. S. Lee, P. Almeida, S. Hu, T. Jan, K. D. Wilson, D. Leong, J. Rosenberg, M. Yao, R. C. Robbins, J. C. Wu, Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells, J Clin Invest **121** (2011) 1217-1221.

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

[16] H. Zhang, T. Guo, X. Li, A. Datta, J. E. Park, J. Yang, S. K. Lim, J. P. Tam, S. K. Sze, Simultaneous Characterization of Glyco- and Phosphoproteomes of Mouse Brain Membrane Proteome with Electrostatic Repulsion Hydrophilic Interaction Chromatography, Molecular & Cellular Proteomics **9** (2010) 635-647.

[17] B. Guo, A. D. Sharrocks, Extracellular signal-regulated kinase mitogen-activated protein kinase signaling initiates a dynamic interplay between sumoylation and ubiquitination to regulate the activity of the transcriptional activator PEA3, Mol Cell Biol **29** (2009) 3204-3218.

[18] Z. Tian, N. Tolic, R. Zhao, R. J. Moore, S. M. Hengel, E. W. Robinson, D. L. Stenoien, S. Wu, R. D. Smith, L. Pasa-Tolic, Enhanced top-down characterization of histone post-translational modifications, Genome Biology **13** (2012) R86.

[19] S. Fields, O. Song, A novel genetic system to detect protein-protein interactions, Nature **340** (1989) 245-246.

[20] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Seraphin,

A generic protein purification method for protein complex characterization and proteome exploration, Nat Biotechnol **17** (1999) 1030-1032.

[21] M. Yousefi, V. Hajihoseini, W. Jung, B. Hosseinpour, H. Rassouli, B. Lee, H. Baharvand, K. Lee, G. H. Salekdeh, Embryonic Stem Cell Interactomics: The Beginning of a Long Road to Biological Function, Stem Cell Rev (2012).

[22] S. K. Mallanna, B. D. Ormsbee, M. Iacovino, J. M. Gilmore, J. L. Cox, M. Kyba, M. P. Washburn, A. Rizzino, Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate, Stem Cells **28** (2010) 1715-1727.

[23] M. Pardo, B. Lang, L. Yu, H. Prosser, A. Bradley, M. M. Babu, J. Choudhary, An expanded Oct4 interaction network: implications for stem cell biology, development, and disease, Cell Stem Cell **6** (2010) 382-395.

[24] D. L. van den Berg, T. Snoek, N. P. Mullin, A. Yates, K. Bezstarosti, J. Demmers, I. Chambers, R. A. Poot, An Oct4-centered protein interaction network in embryonic stem cells, Cell Stem Cell **6** (2010) 369-381.

[25] J. Wang, S. Rao, J. Chu, X. Shen, D. N. Levasseur, T. W. Theunissen, S. H. Orkin, A protein interaction network for pluripotency of embryonic stem cells, Nature **444** (2006) 364-368.

Appendices

Samenvatting

Eiwitfosforylering speelt een belangrijke rol in cellulaire communicatie. Fosforylering kan de functie van een eiwit veranderen: het kan een eiwit "aan" zetten, ofwel activeren, of juist "uit", ofwel remmen. In de laatste jaren heeft massaspectrometrie (MS) zich ontwikkeld tot de ultieme methode om niet alleen grootschalig eiwitten te identificeren, maar ook om effectief eiwitfosforylering kwalitatief en kwantitatief in kaart te brengen. De belangrijkste doelstelling van het werk beschreven in dit proefschrift is het vergroten van onze kennis over de dynamiek in eiwitfosforylering en de rol van deze posttranslationele modificatie in pluripotente stamcellen. Om dit doel te bereiken hebben we nieuwe methoden ontwikkeld voor de analyse van fosfoproteomen en deze gebruikt in ons stamcel onderzoek. In **hoofdstuk 1** worden enkele algemene aspecten van pluripotente stamcellen beschreven, met name de moleculaire mechanismen die de pluripotente staat van de stam cellen reguleren. Daarnaast is er een overzicht van massaspectrometrie instrumenten en methoden in opgenomen die voor het analyseren van het proteoom worden gebruikt. Een aantal van deze massaspectrometrische technieken hebben we ook gebruikt in de fosfoproteoom studies beschreven in dit proefschrift.

In **hoofdstuk 2** hebben we de veranderingen in eiwitfosforylering geanalyseerd in humane embryonale stamcellen (hESC) die veroorzaakt worden door de fibroblast groeifactor-2 (FGF-2), een factor essentieel voor de stabiliteit van stamcellen. In dit hoofdstuk hebben we een combinatie van "strong cation exchange" (SCX) chromatografie en titaniumdioxide (TiO_2) gebruikt voor de verrijking van de fosfopeptiden. Daarnaast is dimethyl stabiele isotoop labeling gebruikt voor de relatieve kwantificering van deze fosfopeptiden. FGF-2 is een groeifactor, essentieel voor het langdurige kweken van hESC. Dit maakt een grondig onderzoek naar de invloed van deze groeifactor op de embryonale stamcellen interessant. In deze studie hebben we 3.261 unieke eiwitten geïdentificeerd en gekwantificeerd waarvan 1.064 eiwitten bleken te zijn gefosforyleerd op een of meer aminozuren. Ongeveer 40% van de laatste genoemde groep eiwitten hebben een differentiële fosforylering ondergaan ten gevolge van FGF-2 stimulatie. Verschillende van deze eiwitten behoren tot de algemene signaaltransductie paden die betrokken zijn bij het reguleren van de pluripotentie en proliferatie van hESC (zo-

als Wnt, en PI3K/AKT). Verder hebben we ook hESC-geassocieerde eiwitten gevonden, zoals SOX2, RIF1, SALL4, DPPA4 en DNMT3B alsmede een aantal eiwitten waarvan de transcriptie gereguleerd wordt door de pluripotentie transcriptiefactoren SOX2, OCT4 en NANOG. Onze data geven nieuwe inzichten in hoe FGF-2 helpt bij het handhaven van de ongedifferentieerde toestand van hESC.

Met de bovengenoemde methoden, die gebruikt worden voor de specifieke analyse van het fosfoproteoom, worden voornamelijk peptiden met fosforylering op serine en threonine gedetecteerd. Er worden bijna geen peptiden met tyrosine fosforylering met die methoden gevonden. Het zeer lage aantal peptiden met tyrosine fosforylering geïdentificeerd in deze meer globale fosfoproteoom studies weerspiegelt waarschijnlijk de relatieve zeldzaamheid van deze modificatie. Toch weten we dat juist tyrosine fosforylering een belangrijke rol speelt in signaaltransductie en cellulaire communicatie. Voor mijn studie naar tyrosine fosforylering ging dus de voorkeur uit naar een specifiekere methode, zoals die op basis van immunoprecipitatie. In **hoofdstuk 3** beschrijven we een aanpak met fosfotyrosine specifieke antilichamen om tyrosine gefosforyleerde peptiden te verrijken uit een complex monster van gedigesteerde eiwitten, waarna de verrijkte peptiden geanalyseerd worden met LC-MS/MS. Door gebruikmaking van deze methode, kon een dataset van ongeveer 1000 unieke fosfotyrosine peptiden gegenereerd worden uit HeLa-cellen die gestimuleerd waren met pervanadaat. Na vergelijking van deze dataset met drie veelgebruikte databases, namelijk de Human Protein Reference Database (HRPD), PhosphoSitePlus en phospho. ELM bleek dat er toch nog vele nieuwe gefosforyleerde tyrosines konden worden geïdentificeerd in onze dataset.

Zoals eerder beschreven is FGF-2 een belangrijk groeifactor voor pluripotente stamcellen. Het activeert receptor tyrosine kinasen (in het bijzonder FGFR) om zo differentiatie te voorkomen. Dit geeft nogmaals de belangrijke rol aan van tyrosine fosforylering in het reguleren van de pluripotentie van die stamcellen. Om die reden hebben we de methode, die beschreven is in hoofdstuk 3 gebruikt in combinatie met de reeds genoemde stabiele isotoop dimethyl labeling om het niveau van tyrosine fosforylering te ver-

gelijken tussen twee typen pluripotente stamcellen namelijk: gedifferentieerde cellen die hergeprogrammeerd zijn tot stamcellen (genaamd geïnduceerde pluripotente stamcellen (hiPSC)) en humane embryonale stamcellen (hESC). Vergelijking van hiPSC en hESC is nodig om de overeenkomsten en verschillen tussen deze cellen op moleculair niveau in kaart te brengen. De resultaten uit deze vergelijking zijn beschreven in **hoofdstuk 4**. We vonden dat 31 eiwitten een substantieel verschil in tyrosine fosforylering vertoonden tussen hiPSC en hESC. In deze groep zaten tyrosine kinasen zoals FYN, LYN, LCK, EphA2, FGFR1 en eiwitten geassocieerd aan het cytoskelet die betrokken zijn bij de handhaving van de pluripotente toestand van de stamcellen (bijvoorbeeld TJP2, F11R, cdk5 en PTK2). Dit resultaat suggereert dat er wat verschillen zijn in tyrosine fosforylering geïnduceerd signaaltransductie tussen hiPSC en hESC. Vervolgexperimenten zullen echter nodig zijn om de specifieke rol van de tyrosine fosforylering van deze eiwitten in hiPSCs en hESC beter te begrijpen.

Een vergelijking van deze resultaten met een recente grootschalige analyse van het globale fosfoproteoom in hESC en hiPSC door Phanstiel *et al*. (2011) liet weinig overlap zien in de geïdentificeerde peptiden met gefosforyleerde tyrosines. Om dit verder te onderzoeken hebben we een systematische vergelijking uitgevoerd tussen twee verschillende methoden: analyse van globale fosforylering m.b.v. metaal-gebaseerde affiniteit verrijking (vaak gebruikt in globale fosfoproteoom analysen zoals ook beschreven door Phanstiel *et al.*) en fosfotyrosine specifieke immunoprecipitatie. De voorlopige resultaten van dit werk worden beschreven in **hoofdstuk 5**. In zijn totaliteit was de overlap tussen de twee methoden uiterst klein, in overeenstemming met onze eerdere conclusie. Fosfotyrosine peptide datasets die door middel van metaal-gebaseerde affiniteitschromatografie aanpak (Ti^{4+}-IMAC gevolgd door HILIC) gegenereerd werden, bestonden voor een groter deel uit peptiden die meervoudig gefosforyleerd waren. We concluderen dat voor uitvoerige fosfotyrosine studies m.b.v. LC-MS/MS beiden complementaire verrijkingsmethoden gebruikt moeten worden om het aantal identificaties te verhogen.

In zijn algemeenheid kunnen we stellen dat fosfoproteomics, en in het bij-

zonder fosfotyrosine specifieke proteomics, een enorme vooruitgang hebben doorgemaakt in de laatste tien jaar en ons nu in staat stellen effectief eiwitfosforylering kwalitatief en kwantitatief in kaart te brengen. Echter, ons onderzoek geeft ook aan dat nog geen enkele methode de complete cellulaire communicatie die plaatsvindt via fosforylering kan oppikken. Dientengevolge valt te concluderen dat de methodologie zullen nog steeds verder verbeterd moet worden.

Curriculum Vitae

Name: Adja Damba Zoumaro-Djayoon

Date and place of birth: 12 June 1982, Sokodé (Togo)

Experience

Research assistant, LUMC **2008 - 2008**

Project: coupling immuno-affinity chromatography columns on a HPLC system and develop methods for human plasma protein fractionation prior mass spectrometry analysis.

Education

MSc, Biomoleculaire wetenschappen, VU Amsterdam **2006 – 2008**

Second internship at the Leiden University Medical center (LUMC), Project: study of protein glycosylation in the cerebella of fast ageing mice. Supervised by: Dr. Oleg Klychnikov and Dr. Oleg Mayboroda

First internship at the department of medical oncology in Erasmus Medical Center, Project: the role of membrane transporters of the solute carrier family in the transport of anticancer drugs, platinum compounds. Supervised by: Dr. Herman Burger and Dr. Erik A.C Wiemer

List of publications

Serena Di Palma, **Adja Zoumaro-Djayoon**, Mao Peng, Harm Post, Christian Preisinger, Shabaz Mohammed, Javier Munoz, Albert J.R. Heck. A comparative analysis of phosphotyrosine peptides enriched by means of immuno-affinity precipitation and metal-based affinity chromatography. Manuscipt in preparation.

Zoumaro-Djayoon AD, Heck AJ, Muñoz J. Targeted analysis of tyrosine phosphorylation by immuno-affinity enrichment of tyrosine phosphorylated peptides prior to mass spectrometric analysis. Elsevier Methods 2012 Feb;56(2):268-274.

Zoumaro-Djayoon AD, Ding V, Foong LY, Choo A, Heck AJ, Muñoz J. Investigating the role of FGF-2 in stem cell maintenance by global phosphoproteomics profiling. Wiley Online Library Proteomics 2011 Oct; 11(20):3962-3971.

Acknowledgements