From DEPARTMENT OF ONCOLOGY-PATHOLOGY
Karolinska Institutet, Stockholm, Sweden

CHARACTERIZING CANCER CELL SIGNALING AT THE PROTEIN LEVEL: FROM TARGETED TO PROTEOME AND PHOSPHOPROTEOME-WIDE ANALYSES

Elena Panizza

Karolinska Institutet

Stockholm 2017

Cover art by Elena Panizza

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by AJ E-Print AB

ISBN 978-91-7676-761-0

Characterizing cancer cell signaling at the protein level: from targeted to proteome and phosphoproteome-wide analyses

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Elena Panizza

Principal Supervisor:
Professor Janne Lehtiö, Ph.D.
Karolinska Institutet,
Department of Oncology-Pathology

Co-Supervisors:
Lukas M. Orre, Ph.D.
Karolinska Institutet,
Department of Oncology-Pathology

Rui M. Mamede Branca, Ph.D.
Karolinska Institutet,
Department of Oncology-Pathology

Professor Klas Wiman, Ph.D.
Karolinska Institutet,
Department of Oncology-Pathology

Opponent:
Professor Jesper Velgaard Olsen, Ph.D.
University of Copenhagen,
Faculty of Health and Medical Sciences

Examination Board:
Associate Professor Kaisa Lehti, Ph.D.
University of Helsinki,
Research Program in Genome-Scale Biology
Karolinska Institutet,
Department of Microbiology, Tumor and Cell Biology

Professor Ola Söderberg, Ph.D.
Uppsala Universitet
Department of Pharmaceutical Biosciences

Professor Jonas Bergquist, Ph.D.
Uppsala Universitet
Department of Chemistry

The public defence of this thesis will take place at the Rolf Luft auditorium (L1:00), Karolinska Institutet, Solna on Friday, 6th October, 2017 at 9:00

Venue Address:
Rolf Luft Auditorium,
Rolf Luft centrum, CMM
Karolinska vägen 6
17174 Solna, Sweden

"…to boldly go where no one has gone before."

Star Trek: The Next Generation

Opening narration

ABSTRACT

Proteins are the effectors of cellular functions and the constituting elements of cellular signaling cascades. The ability to analyze the abundances and the dynamics of proteins is central to dissect cellular signaling and its effects on cell physiology.

The aim of this thesis is to gain insight into protein level regulatory mechanisms that contribute to the development of cancer, by optimizing and employing targeted and large-scale methods. Specifically, to examine mechanisms regulating protein stability, localization, protein-protein interactions, and to characterize targets of a protein phosphatase enzymatic activity. Additionally, to optimize a workflow for quantitative phosphoproteomics analysis with the goal of improving the sensitivity and lower the requirement in terms of sample quantity of current methods.

Study I elucidated a mechanism by which S100A4 interacts with p53 in the nucleus thereby promoting its degradation, and the effects of this interaction on the growth and survival of lung and cervix adenocarcinoma cell lines, by employing targeted methods for the analyses of protein stability, protein localization and protein-protein interactions.

Study II elucidated a mechanism by which TRAP promotes metastasis-related cell properties in breast cancer cells via the TGFβ-pathway and CD44, by a combination of proteomics and phosphoproteomics analyses with targeted methods. Furthermore, a moderate-depth phosphoproteomic profiling of TRAP overexpressing cells was achieved by peptide fractionation by high-resolution isoelectric focusing (HiRIEF) on IPG strips pH range 2.5-3.7, and provided a list of putative targets of TRAP phosphatase activity.

Study III developed a workflow for in-depth quantitative phosphoproteomics analysis based on high-resolution isoelectric focusing (HiRIEF) fractionation on a wide pH range (2.5-10). The workflow employs phospho-enrichment by titanium dioxide coupled with isobaric labeling by TMT, and provides for good analytical depth and sensitivity, requiring a low amount of starting material. Application of this workflow for the analysis of cervix adenocarcinoma cells HeLa revealed 1,264 novel phosphorylation sites, of which 165 phospho-sites that are suggested to have a regulatory function during the mitotic phase, based on kinase-association analysis.

In summary, the work presented in this thesis contributes to the collective effort of improving and applying targeted and large-scale methods for the analysis of protein level regulatory mechanisms, particularly by focusing on the optimization of a workflow for phosphoproteomics analysis. Development of these methods and improvements in integrating discovery and validation efforts, will be central in the coming years and offer unprecedented opportunities for increasing our understanding of life and to discover new treatments and cures for diseases.

LIST OF SCIENTIFIC PAPERS

I. Lukas M. Orre, **Elena Panizza**, V.O. Kaminskyy, E. Vernet, T. Gräslund, Boris Zhivotovsky and Janne Lehtiö.
S100A4 interacts with p53 in the nucleus and promotes p53 degradation. *Oncogene (2013), 32, 5531–5540*

II. Anja Reithmeier#, **Elena Panizza**#, Michael Krumpel, Lukas M Orre, Rui MM Branca, Janne Lehtiö, Barbro Ek-Rylander, Göran Andersson (# Authors shall be considered as first authors).
Tartrate-resistant acid phosphatase (TRAP/ACP5) promotes metastasis-related properties via TGFβ2/TβR and CD44 in MDA-MB-231 breast cancer cells. *BMC Cancer, 2017. In press.*

III. **Elena Panizza**, Rui MM Branca, Peter Oliviusson, Lukas M Orre, and Janne Lehtiö.
Isoelectric point-based fractionation by HiRIEF coupled to LC-MS allows for in-depth quantitative analysis of the phosphoproteome. *Scientific Reports 7, Article number: 4513 (2017)*

Additional publications

Johansson HJ, Sanchez BC, Mundt F, Forshed J, Kovacs A, **Panizza E**, Hultin-Rosenberg L, Lundgren B, Martens U, Máthé G, Yakhini Z, Helou K, Krawiec K, Kanter L, Hjerpe A, Stål O, Linderholm BK, Lehtiö J. Retinoic acid receptor alpha is associated with tamoxifen resistance in breast cancer. *Nat Commun. (2013), 4:2175.*

Voena C, Di Giacomo F, **Panizza E**, D'Amico L, Boccalatte FE, Pellegrino E, Todaro M, Recupero D, Tabbò F, Ambrogio C, Martinengo C, Bonello L, Pulito R, Hamm J, Chiarle R, Cheng M, Ruggeri B, Medico E, Inghirami G. The EGFR family members sustain the neoplastic phenotype of ALK+ lung adenocarcinoma via EGR1. *Oncogenesis (2013), 2:e43.*

TABLE OF CONTENTS

INTRODUCTION 10

1 The hallmarks of cancer 10

1.1 Resistance to cell death – Inactivation of p53 11

1.1.1 Functions of the p53 protein 11

1.1.2 Regulation of MDM2-dependent p53 degradation and mechanisms of p53 inactivation in cancer 12

1.1.3 The S100 family of proteins and p53 13

1.2 Activation of invasion and metastasis – Mediators of cellular invasion 13

1.2.1 The invasion-metastasis cascade 13

1.2.2 Epithelial-to-mesenchymal transition and local invasion 14

1.2.3 Contrasting roles of TGFβ: a protein with tumor suppressor and tumor promoting capabilities 15

1.2.4 CD44 participates in cancer cells EMT, adhesion, migration and metastasis 16

1.2.5 Tartrate resistant acid phosphatase (TRAP) and bone metastasis 17

1.3 Genome instability – Deregulation of progression through mitosis and chromosomal instability 18

1.3.1 The cell cycle 18

1.3.2 Tight regulation of the progression through mitosis prevent errors in chromosomal segregation 19

1.3.3 Regulation of cell cycle progression and mitotic phosphorylation 20

2 Analyzing cellular signaling at the protein level 21

2.1 Complexity of the proteome and mechanisms of protein level regulation 21

2.2 Protein post-translational modifications – Protein phosphorylation 22

2.2.1 Protein kinases and protein phosphatases 23

2.2.2 Protein phosphorylation impacts protein function and has a central role in intracellular signaling 25

2.2.3 Deregulation of phosphorylation signaling in cancer 26

AIMS AND OBJECTIVES OF THE PRESENT THESIS 28

1 Overall aim of the thesis 28

2 Specific objectives 28

METHODS FOR THE ANALYSIS OF PROTEIN LEVEL REGULATION 29

1 Experimental methods 29

1.1 Study I 29

1.1.1 Protein turnover analyses 29

1.1.2 Protein-protein interaction analyses 30

1.2 Studies II and III 31

1.2.1 Mass spectrometry-based proteomics and phosphoproteomics analyses 31

1.2.2 Protein extraction and solubilization 32

1.2.3 Protein digestion to peptides 32

1.2.4 Phospho-peptide enrichment 33

1.2.5 Relative quantification by LC-MS: metabolic labeling by SILAC and isobaric labeling by TMT 34

1.2.6 Sample pre-fractionation by high-resolution isoelectric focusing (HiRIEF) and its application to phosphoproteomics analysis 34

1.2.7 Phosphoproteomics analysis by LC-MS/MS using Orbitrap mass spectrometers 36

2 Computational methods for the analysis and interpretation of large-scale proteomics and phosphoproteomics data 37

2.1 Studies II and III 37

2.1.1 Hierarchical clustering 37

2.1.2 Gene ontology enrichment analysis 39

2.1.3 Protein-protein interaction network analysis 40

2.1.4 Prediction of protein kinase-substrate association strength and the identification of putatively functional phosphorylation sites 41

RESULTS AND DISCUSSION 42
1 Study I - S100A4 interacts with p53 in the nucleus and promotes p53 degradation. 42
2 Study II - Tartrate-resistant acid phosphatase (TRAP/ACP5) promotes metastasis-related properties via TGFβ2/TβR and CD44 in MDA-MB-231 breast cancer cells..... 43
3 Study III - Isoelectric point-based fractionation by HiRIEF coupled to LC-MS allows for in-depth quantitative analysis of the phosphoproteome. 45
GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES 48
1 General conclusions 48
2 Future perspectives for the integration of large-scale and targeted methods..... 48
ACKNOWLEDGEMENTS..... 55
REFERENCES 57

LIST OF ABBREVIATIONS

ATC	Anatomical Therapeutic Chemical Classification
Cdc20	Cell division cycle 20
Cdc42	Cell division control protein 42
CDK	Cyclin-dependent kinase
CID	Collision-induced dissociation
CIN	Chromosomal instability
CTGF	Connective tissue growth factor
EGFR	Epidermal growth factor receptor
EMT	Epithelial-to-mesenchymal transition
ESI	Electrospray ionization
FAK	Focal adhesion kinase
FASP	Filter-aided sample preparation
FDA	Food and Drug Administration
FGF	Fibroblast growth factor
GO	Gene ontology
HAS	Hyaluronan synthase
HCD	Higher-energy collisional dissociation
HGF	Hepatocyte growth factor
HiRIEF	High-resolution isoelectric focusing
IL-1	Interleukin 1
IP	Immunoprecipitation
IPG	Immobilized pH gradient
JNK	C-Jun N-terminal kinase
LC-MS	Liquid chromatography-mass spectrometry
MAPK	Mitogen-activated protein kinase
MMP	Matrix metalloprotease
mTOR	Mammalian target of rapamycin
OPN	Osteopontin

PDGF	Platelet-derived growth factor
PDGFR	Platelet-derived growth factor receptor
pI	Isoelectric point
PI3K	Phosphoinositide-3-kinase
pKa	Constant of acid dissociation
PKC	Protein kinase C
pRb	Retinoblastoma-associated protein
PTB	Phospho-tyrosine binding
PTM	Post-translational modification
Rac1	Ras-related C3 botulinum toxin substrate 1
RhoA	Ras homolog family member A
RTK	Receptor tyrosine kinase
SAC	Spindle assembly checkpoint
SH2	Src homology 2
SILAC	Stable isotope labeling with amino acids in cell culture
SMAD	Mothers against decapentaplegic homolog
SP3	Single-Pot Solid-Phase-enhanced Sample Preparation
TCGA	The Cancer Genome Atlas
TβR	Transforming growth factor beta receptor
TGFα	Transforming growth factor alpha
TGFβ	Transforming growth factor beta
TNFα	Tumor necrosis factor alpha
TRAP	Tartrate resistant acid phosphatase
VEGF	Vascular endothelial growth factor
WB	Western Blot
wt	Wild type
w/w	Weight/weight

INTRODUCTION

1 The hallmarks of cancer

A seminal paper by Douglas Hanahan and Robert A. Weinberg published in 2000 defined six alterations in cell physiology that are essential for malignant cell growth, known as the hallmarks of cancer [1]. Two emerging hallmarks and two enabling characteristics were added in a second paper by the same authors published in 2011 [2]. The hallmarks include evading growth suppressors, inducing angiogenesis, deregulating cellular energetics, activating invasion and metastasis, sustaining proliferative signaling, enabling replicative immortality, avoiding immune destruction and resisting cell death, while the characteristics that enable tumor formation include genome instability and mutation and tumor-promoting inflammation (**Figure 1**). Despite the heterogeneity of human cancers, these hallmark features are shared by all human tumors and therefore provide a powerful theoretical framework to dissect the cellular processes underlying the development of malignancies.

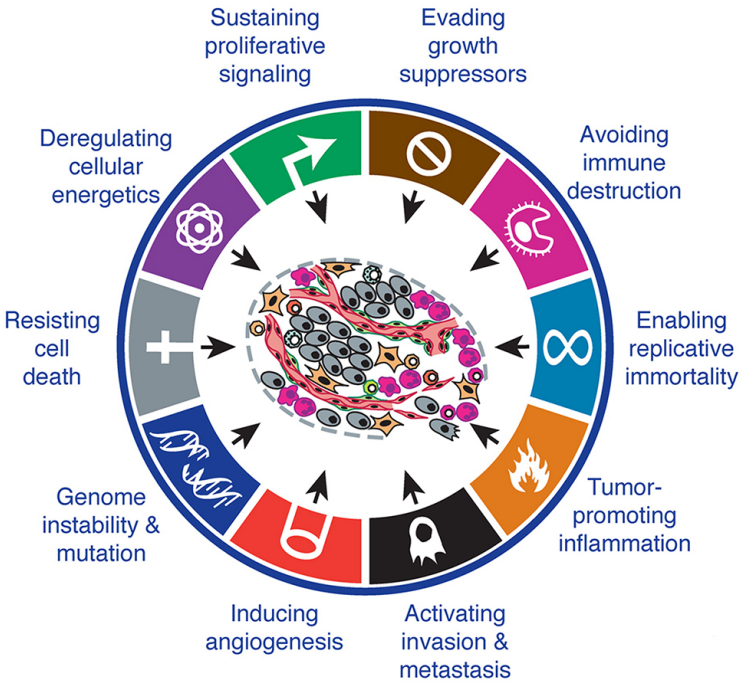

Figure 1. Hallmark alterations and characteristics enabling cancer.
Adapted from Cell. 2011 Mar 4;144(5):646-74, Hanahan D, Weinberg RA., "Hallmarks of cancer: the next generation", doi: 10.1016/j.cell.2011.02.013, with permission from Elsevier.

The doctoral work presented herein includes three studies, each of which is related to a different cancer hallmark or tumor-enabling characteristic: resistance to cell death, activation of invasion and metastasis and genome instability. These hallmarks will be discussed in paragraphs 1.1 to 1.3 of this section. The intent of these paragraphs is not to cover the subject

exhaustively, but rather to provide essential notions about the hallmarks, as well as to discuss specific aspects that are relevant to frame the work presented in this thesis.

Paragraph 1.1, related to Study I, will introduce fundamental concepts on tumors resistance to cell death and on the contribution of p53 inactivation to this cancer hallmark, as well as summarize the mechanisms behind p53 inactivation. Paragraph 1.2, related to Study II, will focus on the process of cell invasion and metastasis, particularly on the roles of TGFβ, CD44 and the protein phosphatase TRAP. Paragraph 1.3, related to Study III, will discuss the basics of the cell cycle and mention main mechanisms of surveillance of the cell genome integrity that are exerted at different points of the cell cycle.

1.1 Resistance to cell death – Inactivation of p53

Cell death is a common event in the human body. Cells can activate a program where an ordered series of events leads to cell death, called apoptosis. Apoptosis occurs physiologically in many tissues of the human organism to facilitate tissue renewal and maintain the appropriate number of cells. Apoptosis is also triggered by abnormalities in several cellular systems, including DNA damage, aberrant signaling mediated by oncogenes, lack of survival factors or hypoxia [3,4].

Cancer cells ability to avoid the activation of the apoptotic program in response to these stimuli grants them a survival advantage when conditions are harsh (e.g. lack of nutrients or of oxygen due to poor blood supply). Additionally, disruption of the normal apoptotic program increases DNA replication errors, and consequently the mutational rate, which in turn accelerate the rate at which mutations that favor cell proliferation occur, therefore enhancing tumor development [1,2].

1.1.1 Functions of the p53 protein

One of the main functions of the protein p53 is to activate the apoptotic program in case of errors in the cell genome or defects in the cell replicative machinery [5].

The p53 protein is normally present at very low levels in the cell. An increase in the level of p53 rapidly activates a series of events that leads to cell cycle arrest or to apoptosis. In order to maintain low levels of this protein, p53 is constantly degraded thus preventing its accumulation in the cell and activation of its function [6]. Main stimuli determining an increase in p53 levels include UV or ionizing radiations (causing double- or single-strand DNA breaks), hypoxia, lack of nucleotides and oncogene signaling (such as aberrant activation of the PI3K, Ras-Raf-MAPK pathway) [5,7].

p53 functions as a transcription factor. Wild type p53 exists in the cell nucleus as a tetramer [8,9]. Increase of p53 levels and post-translational modifications of its C-terminal domain lead to p53 activity, resulting in the transcription of several target genes. Two well studied targets of p53 transcriptional activity are $p21^{CIP1}$ [10,11], which activates a program that leads to cell cycle arrest, and MDM2, the E3 protein ubiquitin-ligase that targets p53 for

degradation [12–14]. Activation of MDM2 transcription by p53 establishes a negative feedback loop that keeps p53 levels low [15,16].

Besides $p21^{CIP1}$ and MDM2, p53 induces the transcription of a large number of other genes, overall exerting two main effects [5,7]:

1) cytostatic: growth arrest is induced to give time to the cell to repair errors in its DNA before allowing the cell cycle to proceed and cell division to occur. To help with this task, p53 also induces the expression of DNA repair genes.
2) apoptosis inducer: in case of major damages to the DNA and inability of the cell to repair them, or in case of anoxia or severe imbalances in cellular signaling, p53 up-regulates the expression of pro-apoptotic genes as well as hinders the expression of anti-apoptotic genes.

1.1.2 Regulation of MDM2-dependent p53 degradation and mechanisms of p53 inactivation in cancer

As mentioned in the previous paragraph, p53 is constantly degraded in the cells. The half-life of p53 is estimated to be around 20 minutes [17–19], therefore blockage of p53 degradation results in a rapid increase of p53 abundance in the cell.

p53 degradation is mediated by the E3 ubiquitin ligase MDM2. Binding of MDM2 to p53 immediately shuts off its transcriptional activity. Additionally, MDM2 catalyzes the addition of ubiquitin groups on the N-terminal region of p53, and mediates its export from the nucleus to the cytoplasm, where ubiquitinated p53 is degraded in the proteasome [12–14]. The protein ARF is another main regulator of p53 levels. ARF inhibits MDM2 activities by preventing its interaction with p53, either by sequestering MDM2 in the cytoplasm, or by targeting it to the nucleolus [20,21]. Cancer cells often show deregulation of either or both of these regulatory mechanisms. p53 wild type (wt) cancer cells often have both copies of the ARF genes inactivated, resulting in increased degradation of p53 by MDM2. Many cancers achieve a similar effect by overexpressing MDM2 [5].

p53 function is often lost in cancer due to mutations of the p53 gene, which are displayed by 30-50% of human cancers [22–24]. Contrary to the majority of tumor suppressor genes, which are mostly inactivated by frameshift or nonsense mutations, p53 is often found to be affected by mutations that lead to substitution of one of its amino acids (missense mutations) [25,26]. The resulting protein product is defective and cannot form a functional p53 tetramer. Thus, heterozygous frameshift or nonsense mutations are sufficient to mediate loss of p53 function as the protein product of the p53 mutated allele will form a non-functional complex when associating with the protein product of the p53 wt allele. On the contrary, in the case of heterozygous loss-of-function mutations (such as deletions), the protein product of the other (wt) allele suffice to maintain p53 functions, therefore loss-of-function mutations of p53 are not common in cancer [23]. As a result of the formation of non-functional p53 tetramers, MDM2 transcription is also not activated in p53 mutant cells, with the result that the inert p53

protein accumulates in the cell [27]. p53 is instead present at very low levels in p53 wt tumors as the negative feedback loop enacted by MDM2 is preserved.

1.1.3 The S100 family of proteins and p53

S100A4 belongs to the S100 family of Ca^{2+}-regulated proteins. The S100 family of proteins is composed of 21 members. The expression of multiple members of the S100 family is deregulated in different types of human cancers, including colorectal, pancreatic, lung, breast and prostate cancer [28], with each type of cancer being characterized by a different S100 proteins expression signature. Overexpression of S100A4 is associated with aggressive disease and development of metastasis in several human cancer types, and particularly in melanoma [29], breast [30,31] and lung [32,33] cancer. Research conducted in our group demonstrated that the expression of S100A4 and S100A6 in different lung cancer cell lines is up-regulated in response to ionizing radiation in a p53-dependent manner [34].

S100 proteins do not have any known enzymatic function, but are thought to regulate other proteins through direct interactions. Interaction of S100B with p53 is well studied in melanoma [35–37]. Several reports demonstrate the interaction of S100A4 with the cytoskeletal protein myosin IIA [38,39], possibly explaining the connection between S100A4 and the development of metastasis. At the time we initiated Study I, S100A4 had been demonstrated to interact *in vitro* with p53 nuclear export signal domain [40,41]. This domain is connected with p53 degradation, as export to the cytoplasm is necessary for the targeting of p53 to the proteasome. At about the same time, an interaction between p53 and MDM2 had also been demonstrated [42]. These studies were pivotal for our investigation of the interactions between the S100A4 protein and p53 and MDM2 in cells, as well as for studying the impact of these interactions on p53 degradation, and the consequences on cellular proliferation and survival. Results from this research furthered the understanding of the roles of S100A4 in cancer. The reader is referred to the "Results and Discussion" section of this thesis for a summary of the results of Study I and of their relevance in the field.

1.2 Activation of invasion and metastasis – Mediators of cellular invasion

1.2.1 The invasion-metastasis cascade

A metastasis is formed when cancer cells leave the primary site of a tumor, travel through the body's blood and lymphatic vessels and form a colony in a new site of the organism. Metastases are the cause of 90% of cancer-related deaths [43]. The tendency to form metastasis as well as the preferred site of metastasis differ based on tumor type. For example, melanomas often form metastases, while brain tumors rarely do [4]. Common sites of metastases are the bones, the brain, the liver and the lungs [43,44]. Outgrowth of metastatic colonies at these sites disrupts normal tissue function and is therefore life threatening.

Metastases are formed through a process named the invasion-metastasis cascade, which proceeds in several steps: local invasion, intravasation, transport through the circulation, arrest in a secondary site, extravasation and colonization of a new site in the organism. The

first steps of this process occur commonly, in fact metastasis-free patients often display a large number of cancer cells in circulation [45]. However, the metastatic process is inefficient; the rate-limiting step of this process is thought to be the colonization of a new site [46,47]. Execution of the steps of the invasion-metastasis cascade involve the acquisition by the migrating cell of a mesenchymal phenotype, the degradation of extracellular matrix to make space for invading cells and the reorganization of the migrating cell cytoskeleton and dynamic modulation of its adhesion to allow for cell movement [45]. A complex network of intracellular signals as well as signals between cells and the extracellular matrix contribute to the realization of these steps [48].

1.2.2 Epithelial-to-mesenchymal transition and local invasion

The process of local invasion is well defined for carcinomas, which represents 90% of human tumors. Carcinomas are cancers that develop from epithelial cells. Epithelial cells form the tissues that line the inner and outer cavities of the human bodies. Epithelial tissues do not contain blood vessels and get nutrients from the underlying connective tissue, by diffusion through the basement membrane, a specialized type of extracellular matrix. Carcinomas form at the epithelial side of basement membranes; after an initial phase in which they grow locally, they may acquire the ability to breach through the basement membrane and invade the underlying connective tissue. This ability, termed local invasion, defines a malignant tumor [45].

Local invasion is enabled by a change in the cancer cells' transcriptional programs, that determines a shift from an epithelial to a mesenchymal phenotype (epithelial-to-mesenchymal transition, EMT). During EMT cells stop expressing E-cadherin and start expressing N-cadherin. E-cadherin connects epithelial cells tightly to each other, forming ordered and stiff cells layers, while N-cadherin form weaker intracellular bonds that allow for cell motility, therefore permitting the cancer cells to break out of the ordered epithelial layer and move through the basement membrane into the underlying connective tissue [49]. Local invasion is also dependent on the secretion of matrix-metalloproteases (MMPs), which are synthesized by cancer cells after they have undergone EMT. MMPs digest the basement membrane providing cancer cells with extra space to expand, as well as with growth factors that are present in the proteinaceous matrix of the basement membrane itself. Ultimately, breaching through the basement membrane allows the cancer cells to reach the underlying connective tissue, where they get access to blood vessels [45].

EMT can be triggered by molecules derived from the connective tissue, such as TGFβ [50] or collagen-I [51,52]. Notably, cancer cells that have undergone EMT in response to TGFβ derived from the underlying connective tissue have been shown to be able to start produce TGFβ themselves and therefore become self-sufficient in maintaining the mesenchymal phenotype [53]. Another effector of EMT as well as a promoter of cell invasiveness is CD44, a cell receptor strongly involved in cellular adhesion [54].

1.2.3 Contrasting roles of TGFβ: a protein with tumor suppressor and tumor promoting capabilities

TGFβ is a well-studied molecule able to exert contrasting effects on cells. As we mentioned in the previous paragraph, TGFβ can trigger EMT and contribute to the maintenance of cells mesenchymal phenotype, therefore contributing to the formation of metastases. In normal cells though, TGFβ acts as a growth repressor. Therefore cancer cells need to evade this effect of TGFβ to benefit from its invasion-inducing capabilities [53].

TGFβ acts by binding to its receptors on the cell surface, and triggering the phosphorylation and translocation to the nucleus of the Smad transcription factors [53]. Smads induce the transcription of numerous target genes, including the two CDK inhibitors $p15^{INK4B}$ and $p21^{CIP1}$ [55–57], which halt cell cycle progression. This effect is counteracted by Myc, which represses the transcription of $p15^{INK4B}$ and $p21^{CIP1}$ [53,58–60]; TGFβ though, is able to shut down Myc transcription [61,62], thereby keeping full control on cell cycle progression. TGFβ also trigger another series of events in cells independently from the Smad transcription factors (non canonical TGFβ signaling) [53]. Effector molecules stimulated by TGFβ include several proteins in the MAPK pathway (MAPK1, MAPK3, MAPK14 and JNK) [63–65], PI3K, AKT and mTOR [66], and several GTP-binding proteins (Ras, RhoA, Rac1 and Cdc42) [67]. Amplified activation of these non-canonical TGFβ effectors in response to TGFβ contributes to cancer development and progression.

Several mechanisms are responsible for the resistance of cancer cells to the cytostatic effects of TGFβ, including:

1. Overexpression of Myc, which in turn inhibits $p15^{INK4B}$ and $p21^{CIP1}$ transcription through several mechanisms [56,68,69]. Myc overexpression can result from its transcription having become unresponsive to TGFβ [70], either because of deregulation of Myc transcription factors, or because of mutations in the Myc gene promoter leading to constitutive transcription [71]
2. Inactivation of the pRb signaling pathway: when pRb is eliminated, the sole presence of $p15^{INK4B}$ is not sufficient to halt the cell cycle [72]
3. Hyperactivation of the PI3K/AKT pathway, which results in inhibition of Smad3-dependent $p21^{CIP1}$ trascription [55,73–75]

Once cancer cells have evaded its cytostatic effects, TGFβ instead promotes tumor development by increasing cell proliferation and cell invasiveness. There are several mechanisms responsible for these effects, many of which have not yet been fully elucidated. For example, TGFβ mediates increased cellular proliferation by inducing the expression of cytokines and growth factors, including IL-1, CTGF, PDGF, TGF-α, EGFR and PDGFR [53]. A plethora of mechanisms responsible for TGFβ-mediated EMT and metastasis have been demonstrated [53,76,77], including:

1) Stimulation of EMT and metastasis through several molecules in the TGFβ canonical and non canonical pathways

2) Crosstalk between TGFβ and integrins: integrins have been shown to trigger EMT in a TGFβ-dependent manner in several types of cancer cells
3) TGFβ-dependent dissolution of cellular tight junctions and disassembly of cell actin cytoskeleton, leading to induction of EMT
4) TGFβ-induced production of factors mediating breakdown of mineralized bone, leading to enhanced osteolytic bone metastases

Contributions of TGFβ to cell invasiveness are still being defined and are an active area of research and will be discussed in Study II of this thesis.

1.2.4 CD44 participates in cancer cells EMT, adhesion, migration and metastasis

CD44 is a cell receptor whose major ligand is hyaluronan, the most abundant component of the basement membrane. Binding of CD44 to hyaluronan regulates EMT and inhibition of the synthesis of hyaluronan reduces EMT and the formation of metastases [78,79]. Hyaluronan, CD44, TNFα and TGFβ have been shown to collaborate in inducing EMT. In this model, TNFα induces CD44 activation by promoting the synthesis of hyaluronan; CD44 in turn activates TGFβ receptor II and its downstream signaling cascade mediated by the Smads, resulting in EMT [80] (**Figure 2**).

Additional CD44 ligands include osteopontin (OPN), FGF2, VEGF, HGF, collagen, laminin,

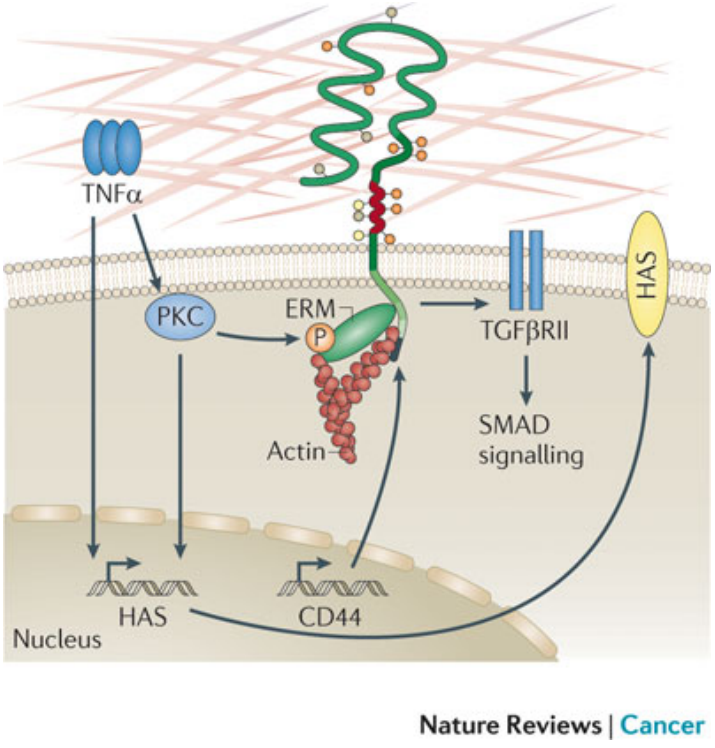

Figure 2. A signaling pathway leading to EMT dependent on the crosstalk between CD44, TNFα and TGFβ. This scheme illustrates an example of crosstalk between several molecules, including CD44, leading to EMT. TNFα activates the transcription of hyaluronan synthase (HAS) mediated by protein kinase C (PKC). Hyaluronan binds and activates CD44 on the plasma membrane; activated CD44 in turn initiates the activation of TGFβ receptor II and its downstream signaling through the SMAD proteins, which ultimately induce EMT [80]. *Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Cancer 11, 254-267 (April 2011), doi:10.1038/nrc3023, Margot Zöller, "CD44: can a cancer-initiating cell profit from an abundantly expressed molecule?", Copyright © 2011.*

and fibronectin. Binding of CD44 to its ligands is involved in the assembly of an extracellular matrix that supports tumor cell adhesion [54] and induces the upregulation of other adhesion molecules, such as integrins [81]. Migrating tumor cells use CD44 to adhere and roll on endothelial cells during the initial steps of extravasation. CD44 is pushed at the leading edge of migrating tumor cells by the actin cytoskeleton, which bind to the CD44 cytoplasmic tail [82,83]. CD44 is also able to concentrate MMPs at the cell surface, which locally degrade the extracellular matrix and create space for the invading cells [84,85]. Activation of TGFβ through CD44-associated MMP9 has been demonstrated, resulting in enhanced invasion and metastasis, as well as increased osteoclast activity and bone resorption [86,87].

1.2.5 Tartrate resistant acid phosphatase (TRAP) and bone metastasis

The bone is a plastic tissue whose turnover is regulated by two types of cells: osteoclasts, which have the function to breakdown mineralized bone, and osteoblasts, which have the function to reconstruct it. As much as 10% of skeletal bone mass is replaced every year. Osteoclasts break the bone down by first demineralizing it, and then by degrading the extracellular matrix which formed the scaffold for the calcium phosphate crystals, in a process termed bone resorption [88]. Metastasizing cancer cells manipulate osteoclasts and osteoblasts to remodel the bone in order to make space for their own growth. Breast cancer cells, for example, are able to activate the osteoclasts to form osteolytic metastases [76].

Tartrate resistant acid phosphatase (TRAP) is the protein encoded by the human gene ACP5, and is expressed by actively resorbing osteoclasts, therefore being a marker for bone resorption [89]. Besides its expression in osteoclasts, TRAP is also highly expressed in several human cancers, including melanoma, breast and ovarian cancers [90,91] and is a marker for bone metastases [92–96]. Even though the connections between TRAP and the development of metastases are numerous, the understanding of the molecular routes mediating this function is fragmented. A few molecular mechanisms regulating and/or mediating TRAP actions have been demonstrated. TRAP modulates the phosphorylation of the focal adhesion kinase (FAK), resulting in increased invasion and metastasis of melanoma and breast cancer [97], and it is a transcriptional target of Forkhead box M, promoting hepatocellular carcinoma metastasis [98]. The fundamental questions of whether and how TRAP phosphatase activity is causally involved in the development of metastases remain unanswered. As of today, TRAP is known to dephosphorylate osteopontin and bone sialoprotein [99], but no other substrates of its phosphatase activity are known. Study II of this thesis addresses this question by analyzing the phosphoproteome and the proteome of breast cells overexpressing TRAP and of the same cells after reverting the increased TRAP expression by shRNA mediated knockdown. The reader is referred to the “Results and Discussion” section for this discussion.

1.3 Genome instability – Deregulation of progression through mitosis and chromosomal instability

Cancer cells are characterized by a high degree of mutability of their genomes, which accelerates the rate at which these cells can acquire combinations of alleles advantageous for their proliferation and survival [2]. As a result of this increased mutability, cancer cells genomes are unstable. Normally, there are different mechanisms in place to halt the process leading to cell division when the DNA is damaged or errors occur during its replication, or when the chromosomes are in disarray prior to cell division. Inactivation of these mechanisms confers cancer cells genomic instability [100].

1.3.1 The cell cycle

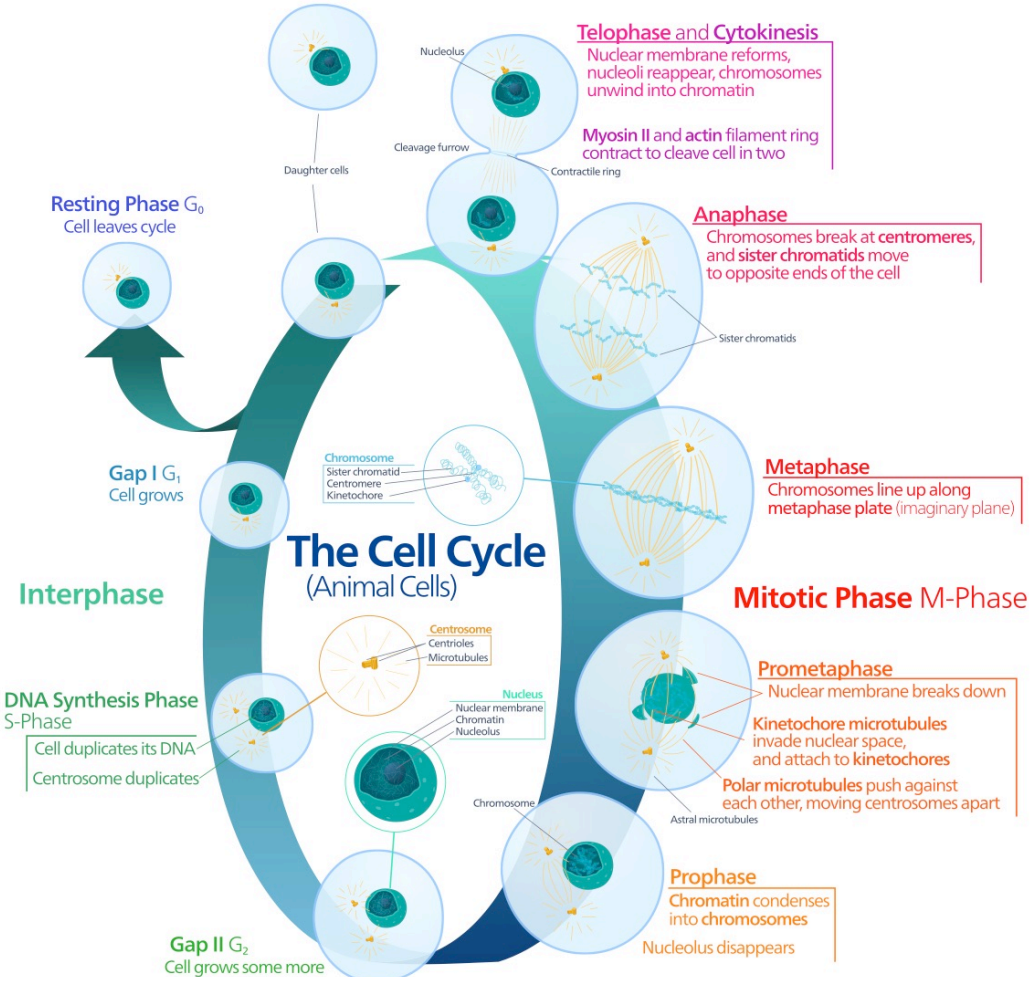

Figure 3. Stages of the cell cycle.

Reproduced from "Wikipedia: The free encyclopedia. (2004, July 22). FL: Wikimedia Foundation, Inc. Retrieved on August 24, 2017, from https://commons.wikimedia.org/wiki/File:Animal_cell_cycle-en.svg

The ordered sequence of events leading to cell division is termed the cell cycle. The cell cycle is conceptually divided into four major stages (**Figure 3**). After cell division, cells enter a phase called G1 (first gap), where they make a decision on whether to grow or to enter a

quiescent state, from which possibly they can differentiate. If the decision to grow is made, cells synthesize the nucleic acids required to duplicate their DNA and the other molecules necessary to form the components of two daughter cells. When all the necessary elements have been synthesized, different signals can trigger the crossing of a decision point in G1 called restriction point (R-point), after which the cell is committed to proceed through the rest of the cell cycle and divide. After traversing the R-point, cells enter the S (synthesis) phase, during which their genetic material is replicated and then progress through a G2 (second gap) phase [101]. Finally, cell division into two daughter cells is performed during the mitotic (M) phase. The mitotic phase is further divided into prophase, pro-metaphase, metaphase, anaphase, telophase and cytokinesis [4].

During the cell cycle, several surveillance mechanisms monitor each step and permit the cell to proceed to the next step only if the previous step has been completed successfully. In case of errors, these monitors arrest the progression through the cell cycle until the problems have been resolved. These monitoring mechanisms constitute the cell cycle checkpoints and can be summarized as [102]:

- DNA damage checkpoint, at the G1/S transition: the cell does not enter the S phase if the genome is damaged
- DNA damage checkpoint, during S phase: DNA replication is halted when errors in the replication occur or if the DNA is damaged, to give to the cell time to repair the damage
- entrance to M-phase: if DNA replication is not completed, entrance into the M-phase is blocked
- spindle assembly checkpoint (SAC): if the chromatids are not properly assembled on the mitotic spindle during pro-metaphase, the cell does not proceed to anaphase to avoid missegregation of the chromosomes to daughter cells

1.3.2 Tight regulation of the progression through mitosis prevent errors in chromosomal segregation

Alterations affecting different stages of the cell cycle contribute to cancer progression by providing growth, survival or transformation advantages to the cell. A main cell cycle deregulation is loss of the R-point in the G1 phase by inactivation of the pRb gene, which grants cancer cells unrestrained proliferative capability. The vast majority of human tumors display inactivation of the pRb pathway [103]. Another important surveillance mechanism, the SAC, in enacted during mitosis to halt mitotic progression until all chromosomes are correctly attached to the mitotic spindle [104,105]. Mitosis is a highly complex cell cycle stage, which proceeds in an orderly series of steps to ensure that errors in chromosome segregation are not made. Disruption of mechanisms governing the progression through the mitotic phase can result in chromosomal instability (CIN). CIN is defined as a persistently high rate of gain or loss of entire chromosomes or fractions of chromosomes during cell divisions [106,107]. About 85% of cancer cells characterized by CIN are aneuploid, which means that one or more of their chromosomes misses a copy or presents extra copies [4].

As mentioned earlier, mitosis is divided into several phases: prophase, pro-metaphase, metaphase, anaphase, telophase and cytokinesis. Briefly, during prophase the centrosomes migrate to the opposite sides of the cell nucleus and start to polymerize tubulin to form microtubules; chromosomes condense and structures called kinetochores that will be sites of microtubules attachment are assembled at the chromosome centromeric region. During pro-metaphase the nuclear envelope breaks down, so the microtubules produced by the centrosomes can reach and associate with the chromosomes kinetochores. When all chromosomes are attached to the spindle, with their sister chromatids attached to opposite spindle poles (bi-orientation), pro-metaphase is completed. Failure of proper attachment of all chromosomes to the mitotic spindle triggers the SAC. Mitosis continues with chromosomes alignment at an equal distance between the two spindle poles in metaphase (congression), after which sister chromatids are separated and pulled towards their respective pole (anaphase). Finally, the cell divides in two daughter cells (telophase and cytokinesis) [108].

In Study III, cells were treated with nocodazole, a small molecule inhibitor of tubulin polymerization that causes cell cycle arrest in mitosis [109]. Inhibition of tubulin polymerization by nocodazole results in the disruption of the mitotic spindle, which triggers the SAC. The SAC is enacted by the protein Mad2, whose soluble form is generated by binding to kinetochores that are not attached to microtubules. The soluble form of Mad2 is able to inhibit Cdc20, which is the initiator of the series of events that leads to the separation of the sister chromatids [104,105]. In summary, in the presence of nocodazole, unbound kinetochores prevent chromatid separation, thus the cell cycle is arrested in pro-metaphase.

1.3.3 Regulation of cell cycle progression and mitotic phosphorylation

Regulation of progression through the cell cycle is performed by cyclin/cyclin-dependent kinases (CDKs) complexes. Cyclins act as regulatory subunits and CDKs as catalytic subunits. Each type of cyclin pairs with a specific CDK: cyclins D bind CDK4/CDK6, cyclins E bind CDK2, cyclins A bind CDK2 or CDK1 and finally the cyclins B bind CDK1. Cyclin proteins level or availability is regulated by protein degradation or by intracellular localization, which restrict the expression and the activity of particular cyclins to the cell cycle stage that they promote. CDKs protein level instead is practically constant throughout the cell cycle, and their activity is regulated by the presence of the corresponding cyclin as well as by phosphorylation. Cyclin-CDK complexes activity can be abolished by CDK inhibitors, whose activity is enhanced by stressors such as DNA damage, or diminished for example by mitogenic signals [110].

The main regulator of the mitotic phase is the protein kinase CDK1. The control of mitosis involves more than 32,000 phosphorylation and dephosphorylation events [111], which drive cell rounding, condensation of the chromosomes, breakdown of the nuclear envelope, fragmentation of cell organelles, and finally the assembly of the mitotic spindle [112]. Not only do mitotic phosphorylations affect a very large number of substrates, but also display a high site occupancy (the proportion of proteins carrying the phosphorylation in a certain protein pool) compared to other cell cycle phases [113]. Most of these phosphorylation events

need to be reversed for the cell to exit mitosis and divide. The complex cyclin B1/2-CDK1 is responsible for phosphorylating a large number of mitotic substrates. In addition to CDK1, several other protein kinases are activated during the mitosis phase and govern various processes that are necessary for mitotic progression [114]. Phosphorylation events with regulatory functions are tightly controlled during mitosis both time- and space-wise (the latter by regulating the localization of the corresponding protein kinases). Though, the relevance and function of the majority of the phosphorylation events within the large network of mitotic phosphorylations is still unknown. It is hypothesized that a large number of the mitotic phosphorylation events serves to inhibit the function of proteins that are not involved in the processes occurring during mitosis [113]. Study III of this thesis addresses these fundamental questions on the functionality of specific protein phosphorylation events.

2 Analyzing cellular signaling at the protein level

2.1 Complexity of the proteome and mechanisms of protein level regulation

The proteins are the effectors of cellular functions and the constituting elements of cellular signaling cascades. The ability to analyze the abundances and the dynamics of proteins is thus central to dissect cellular signaling and its effects on cell physiology. The ensemble of proteins expressed by a cell (proteome) is extremely complex [115]: the human genome contains approximately 20,000 protein coding genes; from those, it is estimated that approximately 100,000 distinct protein isoforms are generated through alternative splicing [116]. Additionally, the set of expressed genes varies across different cell lineages [117,118]. Transcriptomics analysis can provide information on which genes are transcribed and at what quantity, but transcripts abundances do not always correlate with protein abundances due to the actions of several post-transcriptional regulatory mechanisms [119–122]. Furthermore, protein activity and dynamics (e.g. enzymatic activity or protein-protein interactions) can be analyzed uniquely at the protein level.

Several mechanisms of protein level regulation contribute to the complexity of the proteome (**Figure 4**). These mechanisms include:

- Translational regulation (impacting protein levels)
- Protein stability/turnover (impacting protein levels)
- Protein cleavage events (impacting protein activities)
- Protein-protein interactions (impacting protein levels and activities)
- Subcellular localization (impacting protein levels and activities)
- Post-translational modifications (PTMs, impacting protein activities and levels): more than 200 types of post-translational modifications have been listed [123].

The work presented in this thesis employs several methods for the analysis of mechanisms of protein level regulation in human cancer, including cellular treatments to assay protein turnover and immunoprecipitation and proximity ligation assay (PLA) to study protein-

protein interactions (Study I); mass-spectrometry-based proteomics and phosphoproteomics to analyze protein abundance and phosphorylation status (Studies II and III).

The reader is referred to the Methods section for a description of the basic principles of these methods.

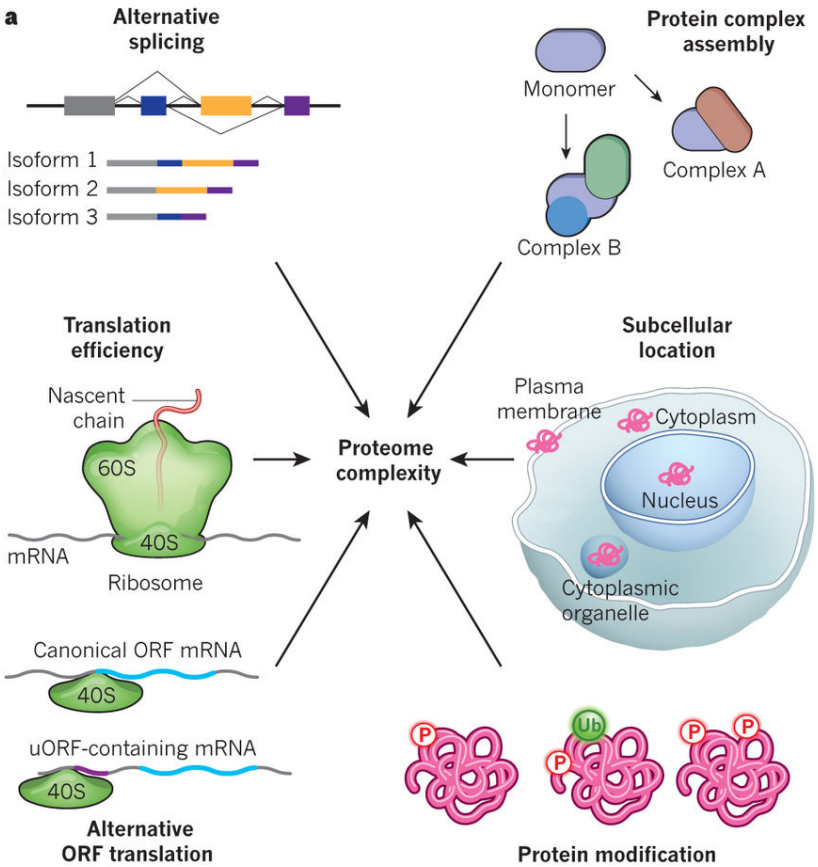

Figure 4. Sources of proteome complexity. Factors contributing to generate the complexity of proteomes include (clockwise from bottom left): translational regulation (use of alternative upstream open reading frames in mRNA translation or translation efficiency); alternative splicing; protein-protein interactions; subcellular localization of proteins; post-translational modifications. *Adapted by permission from Macmillan Publishers Ltd: Nature 537, 328–338 (15 September 2016) doi:10.1038/nature19947, "Proteome complexity and the forces that drive proteome imbalance", J. Wade Harper & Eric J. Bennett, copyright 2016.*

2.2 Protein post-translational modifications – Protein phosphorylation

Protein phosphorylation and protein dephosphorylation refer to the addition and removal of phosphate groups (PO_4^{3-}) to and from proteins. Phosphate can react with the alkyl hydroxyl groups of serine and threonine side chains (pSer and pThr), and with the aryl hydroxyl group of tyrosine side chains, forming phosphate monoesters (pTyr) (**Figure 5**, top) [124]. Phosphate esters are very stable, and have an half-life of 10^{12} years at 25°C [125]. In eukaryotes, phosphorylation occurs mainly on serine, threonine and tyrosine residues [124,126] and is a major mediator of intracellular signaling transduction. Phosphate can also react with the imidazole groups of histidine side chains, and with the amine groups of lysine

and arginine side chains, forming phosphoramidate bonds (**Figure 5**, bottom). Phosphoramidate bonds are rapidly hydrolyzed at high temperatures and at acidic pH [127].

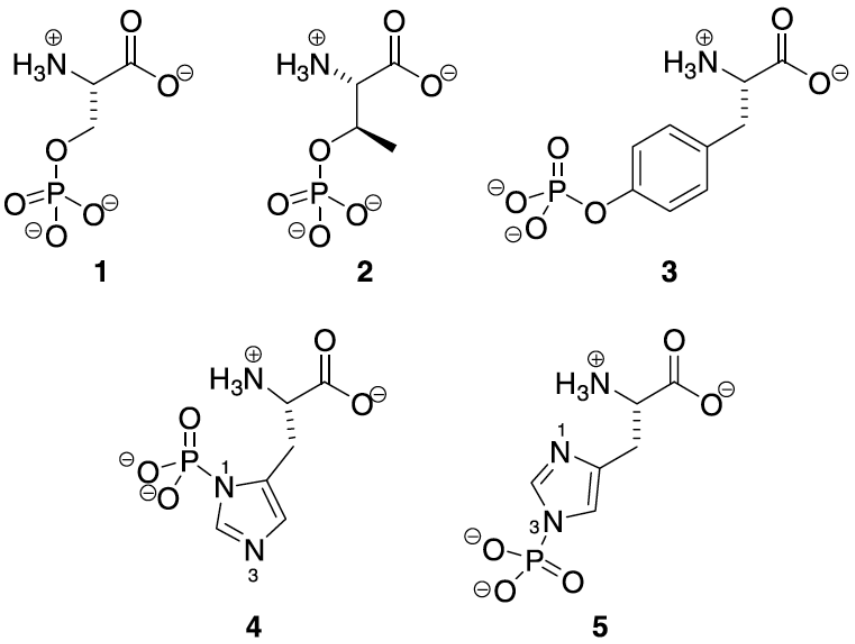

Figure 5. Common phosphorylated amino acids. 1-2-3) Phosphate esters, pSer, pThr and pTyr respectively. 4-5) Phosphoramidate bonds in two phosphohistidine isomers [247]**.** *Reprinted from Amino Acids, "Focus on phosphohistidine", January 2007, Volume 32, Issue 1, pp 145–156, P. V. Attwood et al., © Springer-Verlag 2006, with permission of Springer.*

Biochemical and proteomic procedures to study phosphoester amino acids (pSer, pThr and pTyr) commonly require these conditions, therefore the study of histidine phosphorylation has been very limited. Histidine phosphorylation has recently been shown to have regulatory roles in eukaryotic cells, but its abundance is still debated [128]. Finally, phosphate can also react with the thiol groups of cysteine side chains and with the carboxyl groups of aspartate and glutamate side chains; phosphorylation of these amino acids occur mainly in plants and in prokaryotes [124,126].

The fully protonated conjugate acid of phosphate is phosphoric acid (H_3PO_4); phosphoric acid has three constants of acid dissociation (pKa): 2.2, 7.2 (5.8 as an ester) and 12.4, each corresponding to the pH at which half of the molecules' population undergoes proton dissociation [124]. This means that at physiological pH phosphoric acid donates one or two protons, and therefore possesses one or two negative charges. These pKas values are particularly low, making phosphorylated amino acids stronger acids than their unphosphorylated counterparts, as well as than any other unmodified amino acid. This unique property can be exploited to separate, and consequently enrich for, phosphorylated amino acids, as it will be discussed in more details in paragraph 1.2.4 in the Methods section.

2.2.1 Protein kinases and protein phosphatases

Phosphate groups are added to proteins by protein kinases and removed by protein phosphatases. The transfer of phosphate groups from ATP molecules to specific amino acids

in proteins is catalyzed by protein kinases. Phosphorylation is reverted by hydrolysis catalyzed by protein phosphatases. Eukaryotic protein kinases and phosphatases responsible for the phosphorylation and dephosphorylation of serine, threonine and tyrosine residues are well studied, and eukaryotic histidine protein kinases and phosphatases have also been identified [128]. At least 518 protein kinases and 189 protein phosphatases exist in human [129,130].

Eukaryotic protein kinases derive from a common protein kinase ancestor gene and therefore possess a conserved catalytic domain [131]. Protein kinases can be classified in 10 groups based on sequence and function of their catalytic domain [126]. The group of the tyrosine kinases includes 90 proteins; protein tyrosine kinases have evolved more recently than the other protein kinases, around the same time when multicellular organisms emerged. It is thought that the central role of tyrosine phosphorylation in transducing extracellular signals was fundamental to provide the inter-cellular communication functions necessary for the establishment of multi-cellularity [132]. Protein kinases target specific linear motifs on their substrates; recognition motifs are known for 179 (approximately 35%) of the protein kinases [133]. This specificity can be exploited to predict protein kinase-substrate interactions, expanded upon in paragraph 2.1.4 in the Methods section.

The catalytic domains of protein phosphatases have several different evolutionary origins, unlike protein kinase domains [134,135]. Protein phosphatases often recognize their substrates based on docking interactions, rather than recognition of linear motifs at the site to

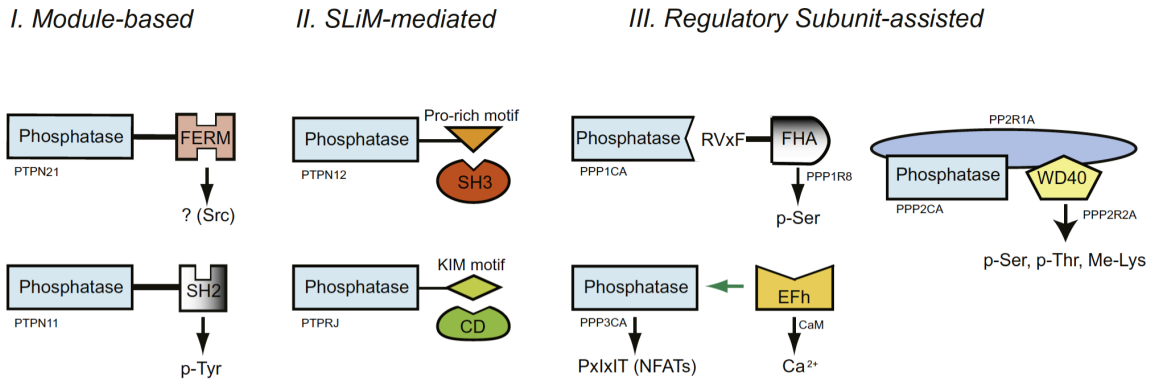

Figure 6. Mechanisms of substrate recognition by protein phosphatases. *Adapted from FEBS Lett. (2012) 586(17):2732-9. doi: 10.1016/j.febslet.2012.05.008. "The human phosphatase interactome: An intricate family portrait", Sacco F. et al.*

be dephosphorylated [136] (**Figure 6**).

Main modes of substrate recognition by protein phosphatases include [137]:

1) *Module-based*: the phosphatase catalytic domain is positioned next to a domain capable of recognizing the substrate (e.g. a SH2 domain, recognizing a specific pTyr site)
2) *Short linear motif (SLiM)-mediated*: the protein phosphatase contains one (or more) SLiM(s) that function as ligands for binding domains contained within the phosphatase substrate

3) *Regulatory subunit-assisted*: the protein phosphatase binds to an adaptor protein that targets it to its substrate; or, the protein phosphatase contains a SLiM that targets it to a subcellular localization where it can find specific substrates

While about 5,700 protein phosphorylated residues have been associated with their upstream protein kinase [138], only 626 have been associated to their protein phosphatase [139]. The variety of mechanisms mediating protein phosphatases' substrate recognition makes the task of enlarging the catalogue of protein phosphatase substrates particularly challenging. The work performed in Study II contributes to this effort by identifying a list of possible substrates of the protein phosphatase TRAP (see paragraph 2 in the Results and discussion section), by using a phosphoproteomics analysis.

2.2.2 Protein phosphorylation impacts protein function and has a central role in intracellular signaling

As illustrated in **Figure 7**, Protein phosphorylation affects protein function in different ways:

1) **Providing for a docking site or, in general, affecting protein-protein interactions.** Protein phosphorylation can create docking sites for proteins that contain domains recognizing phosphorylated amino acids (phospho-binding domains) [126]. More in general, the negative charge carried by the phosphate group impacts the contact surfaces between proteins, and consequently may determine or impede protein-protein contacts [140].
2) **Causing allosteric regulation.** Allosteric regulation occurs when the modification or the binding of a ligand in one region of a protein affects the physical and/or chemical properties of another, distant site of the same protein [141]. Examples of regions known to allosterically regulate domain activity are the activation loops and the glycine-rich regions of protein kinases. Post-translational modifications found in these regions are therefore more likely to affect protein activity [140]. This principle is applied in Study III of this thesis to provide additional clues on the possible functionality of phosphorylation sites.

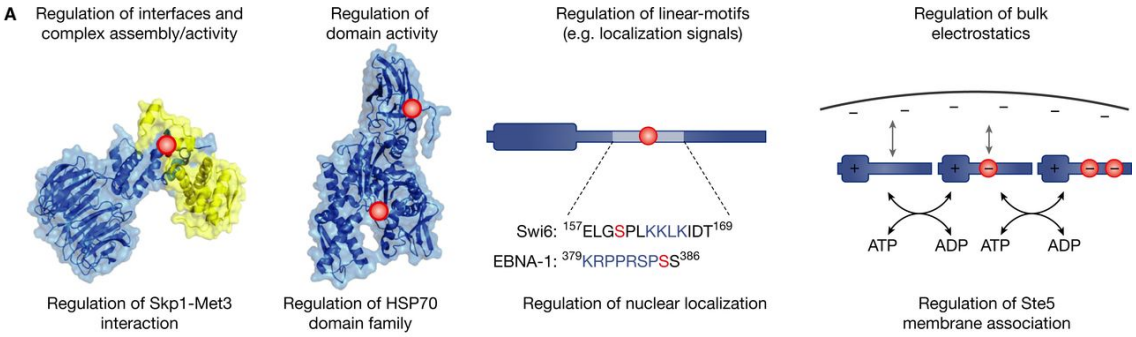

Figure 7. Modalities by which protein phosphorylation impacts protein functions. Main mechanisms by which protein phosphorylation impacts protein function include: regulation of protein-protein interaction, allosteric regulation of protein activity, participation in a linear motif and regulation of protein bulk electrostatic properties. *Adapted from Molecular Systems Biology (2013) 9, 714, DOI 10.1002/msb.201304521, "Evolution and functional cross-talk of protein post-translational modifications", Beltrao P. et al.*

3) **Creating linear motifs.** Phosphorylation can control the functional state of a motif, by switching it to an "on" or "off" status, or altering its binding strenght [142]. Specific phosphorylated linear motifs are responsible for protein re-localization, degradation or cleavage [142,143].
4) **Regulating the bulk electrostatics of the protein.** The addition of several negative charges when multiple sites on a protein are phosphorylated impacts the ability of the protein to interact with other charged surfaces in the cell, such as the cell membrane [144], and determines major rearrangements of the protein structure. This mechanism has been hypothesized to be responsible for inhibiting protein functionality during the mitotic phase [113].

Phospho-binding domains have a central role in the transduction of intracellular signals. The first of these domains to be discovered was the Src homology 2 (SH2) domain, which bind specific oligopeptide sequences containing a phosphorylated tyrosine [145]. As of today, 121 SH2 domains have been identified; SH2 domains are characterized by a common tridimensional structure, but each domain binds a distinct pTyr-containing oligopeptide sequence. Other phospho-binding modules include the pTyr-binding (PTB) domain, and at least 14 distinct pSer/pThr binding domains, including the 14-3-3 and WW domains [126] (**Figure 8**). Phosphorylated amino acids flanked by specific oligopeptide sequences function as ligands for these domains, thereby determining the formation of a physical association between the proteins that carry them. Other interaction domains are able to recognize unmodified peptide motifs, such as the SH3 domain, which recognizes proline-rich sequences. The evolution of these interaction domains provided for a flexible system that could be employed by any cellular protein to form specific protein-protein interactions and is the foundation for the transmission of information occurring in intracellular signaling pathways [126].

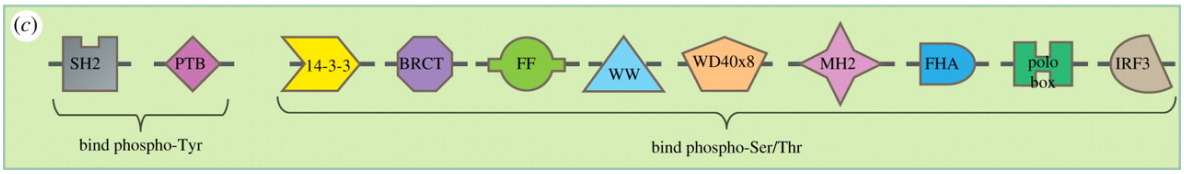

Figure 8 Phospho-tyrosine and phospho-serine/threonine binding domains. *Adapted from Jin et al., "Modular evolution of phosphorylation-based signalling systems", Philosophical Transactions B, 2012 Sep 19;367(1602):2540-55. doi: 10.1098/rstb.2012.0106, by permission of the Royal Society.*

2.2.3 Deregulation of phosphorylation signaling in cancer

Phosphorylation is an extremely common post-translational modification; at present, phosphorylation has been identified on more than 17,000 human proteins [138]. Protein phosphorylation is centrally involved in the transduction of intracellular signals and in the regulation of cellular functions. Sustained phosphorylation signaling resulting from the aberrant activation of protein kinases drive several of the hallmark phenotypes of human cancer, including sustained proliferation, survival, invasion, deregulation of cellular metabolism, angiogenesis, and evasion of immune destruction [146].

Ninety one genes encoding for protein kinases have been identified as cancer drivers, and 40% of those encode for receptor tyrosine kinases (RTKs) [147]. Deregulation of RTKs is particularly involved in sustaining proliferation signaling. Multicellular organisms depend upon the cooperation of large groups of cells mediated by intercellular communication; the functions of cells within tissues are coordinated in order to guarantee the wellbeing of the entire tissue and organism. Growth factors are major signaling molecules involved in intercellular communications. Growth factors act as mitogens, meaning that they provoke cell proliferation by favoring the progression through the R-point in G1. Several growth factors bind to receptor tyrosine kinases (RTKs) located on cell membranes to trigger cell proliferation. Protein tyrosine kinases are uniquely found in multicellular organisms (see paragraph 2.2.1 in this section), and have an exclusive role in receiving extracellular signals. Deregulation of protein kinases grants cancer cells the ability of sustaining proliferative signaling independently from extracellular signals [2]. Other protein kinases that act as drivers of cancer include TGFβ receptors; non-receptor tyrosine kinases (downstream of cell membrane receptors); intracellular protein kinases associated with the PI3K or MAPK pathways, which regulate multiple cellular functions; intracellular protein kinases regulating cytoskeleton organization, cell cycle progression, DNA repair or transcription [147].

Twelve protein phosphatases have been identified as being involved in the development of cancer. This group of protein phosphatases contains proteins acting either as oncogenes or as tumor suppressors, and some of these proteins phosphatases are cancer drivers [130].

AIMS AND OBJECTIVES OF THE PRESENT THESIS

1 Overall aim of the thesis

The overall aim of this thesis is to gain insight into protein level regulatory mechanisms that contribute to the development of cancer, by optimizing and employing targeted and large-scale methods. Specifically, to examine mechanisms regulating protein stability, localization, protein-protein interactions, and to characterize targets of a protein phosphatase enzymatic activity. Additionally, to optimize a workflow for quantitative phosphoproteomics analysis with the goal of improving the sensitivity and lower the requirement in terms of sample quantity of current methods.

2 Specific objectives

Study I:

- To investigate S100A4-p53 interaction and its consequences on p53 protein stability and on the proliferation and survival of lung and cervix adenocarcinoma cells, by employing targeted methods for the analyses of protein stability, protein localization and protein-protein interactions.

Study II:

- To elucidate the molecular mechanisms underlying the promotion of cancer metastasis by the protein TRAP, and to characterize how TRAP affects cellular signaling and identify possible targets of its protein phosphatase activity in breast cancer cells, by proteomics and phosphoproteomics analyses coupled with targeted methods.

Study III:

- To develop a workflow for in-depth quantitative phosphoproteomics analysis using tandem mass tags (TMT) for multiplex quantification, a method rarely employed for phosphoproteomics analysis at the time the study was initiated.
- To evaluate the analytical depth provided by high-resolution isoelectric focusing (HiRIEF) fractionation for phosphoproteomics analysis, by studying cervix adenocarcinoma cells treated with pervanadate or arrested in mitosis.

METHODS FOR THE ANALYSIS OF PROTEIN LEVEL REGULATION

This section presents methods and technologies that were employed in this thesis to analyze mechanisms of protein level regulation. The basis of the methods and the rationale for their application will be discussed. The reader is referred to the papers for technical details on the methods, and for information on other methods that were used in the studies.

1 Experimental methods

1.1 Study I

1.1.1 Protein turnover analyses

Proteins are turned over in cells as a result of their synthesis in the ribosomes and their degradation by the proteasome or by the lysosomes and their secretion. The proteasome degrade proteins that have been modified by the addition of one or more ubiquitin molecules (polyubiquitin chains); the addition of ubiquitin is catalyzed by ubiquitin ligases. Protein ubiquitination is a highly regulated process; human cells contain more than 1,000 ubiquitin ligases [148] with distinct specificities for particular amino acid sequences ("degron" motifs) on target proteins [149]. The affinity of ubiquitin ligases for degron motifs can be modulated by PTMs of the degron motifs; additionally, the enzymatic activity of ubiquitin ligases can also be regulated, either by PTMs or by binding to scaffold or adaptor proteins. Modulation of protein degradation by these mechanisms affects a large array of cellular processes, including: elimination of misfolded proteins, regulation of DNA repair, transcription, replication, and chromosome cohesion/segregation, regulation of the dynamics and functions of several intracellular proteins, control of subunit stoichiometries within protein complexes, regulation of apoptosis, neurodegeneration and many others [150]. Alterations of protein degradation circuits, such as mutations that alter the levels or the activity of ubiquitin ligases or change their substrate specificity, are causally involved in many diseases, including cancer.

In Study I, we employed cellular treatments that inhibit protein synthesis or protein degradation to analyze the rate of p53 degradation under different experimental conditions. Employed treatments include:

- *Cycloheximide:* a protein synthesis inhibitor of bacterial origin, that functions by impeding the movement of tRNA and mRNA molecules in relation to the ribosome (blockage of translational elongation) [151].
- *Nutlin-3A:* a compound that specifically blocks p53 degradation, by inhibiting the interaction between the ubiquitin ligase MDM2 and p53 [152,153].
- *MG132:* a cell permeable molecule that inhibits the protease activity of the proteasome [154,155].

Upon treatment with these compounds, p53 protein levels were measured by Western Blot (WB), to assess whether the employed experimental conditions affected p53 protein stability.

1.1.2 Protein-protein interaction analyses

As mentioned earlier, intracellular protein-protein interactions can impact protein localization, activity and stability (see paragraph 2.1 in the Introduction section). In Study I, we hypothesized that the degradation of p53 can be regulated via interaction with S100A4. In order to analyze the interaction between these two proteins, immunoprecipitation and proximity ligation assay were employed.

Immunoprecipitation (IP) is based on the use of an antibody to precipitate a protein of interest, together with its interaction partners. In Study I, Western Blot analyses was employed to characterize the immunoprecipitated proteins. IPs are performed on protein extracts, therefore they do not provide information on the subcellular location of the identified interactions. Additionally, detergents employed for protein extraction can disrupt some protein-protein interactions. To avoid this undesired effect, proteins complex were cross-linked by incubating the cells with a 0.125% paraformaldehyde solution prior to protein extraction. Crosslinking works by introducing covalent bonds between proteins that are in close vicinity. The advantages of using formaldehyde are that it is a small molecule, and therefore it links only closely associated proteins, as well as its ease of use, as it freely crosses cell membranes (**Figure 9**) [156].

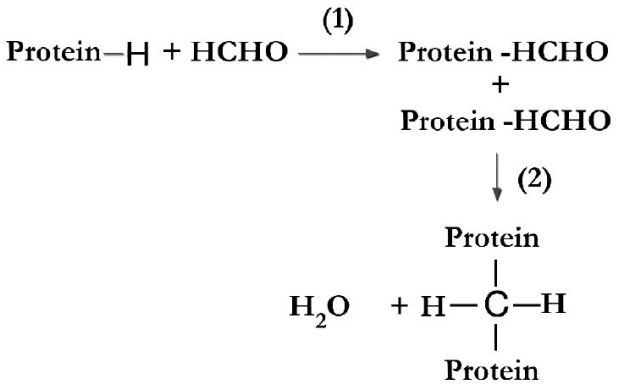

Figure 9. Reaction of formaldehyde with proteins. 1) Formaldehyde reacts with amino acids containing primary amines (lysine), purines and thiols (cysteine). **2)** Two modified amino acids react with each other forming a methylene (-CH2-) bridge. The cross-linking reaction also involves other functional groups such as amides, asparagines and guanidine and tyrosine carbon rings, but at slower rate [248]. *Reproduced from Journal of oral and maxillofacial pathology (2012), 16(3): 400–405, DOI: 10.4103/0973-029X.102496, "Chemical and physical basics of routine formaldehyde fixation", Thavarajahet R. et al., copyright : © Journal of Oral and Maxillofacial Pathology.*

The *proximity ligation assay (PLA)* is a method recently developed, which allows detection and quantification of protein complexes in intact cells, thus providing information on their subcellular localization. The method makes use of two antibodies attached to oligonucleotide molecules. When the two antibodies bind to two cellular proteins located in proximity to each other, the oligonucleotides can be ligated forming a small circle, which is then extended by rolling-circle amplification by a DNA polymerase. The resulting rolling-circle product is then

hybridized with complementary fluorophore-labeled probes, resulting in the emission of strong fluorescent signals that can be captured as dots with a fluorescent microscope (**Figure 10**) [157].

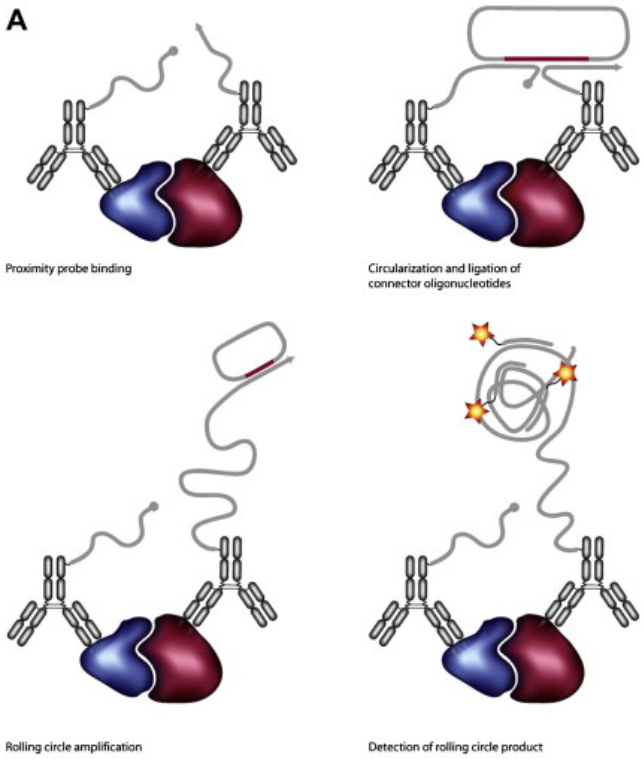

Figure 10. In situ PLA. Two antibodies attached to oligonucleotide molecules bind to proteins that are localized in proximity. The oligonucleotides can therefore be ligated into a circular DNA molecule, which is then amplified by rolling-circle amplification (RCA). Finally, the rolling circle product is hybridized with complementary fluorophore-labeled probes, allowing for its detection. *Reprinted from Methods, Volume 45, Issue 3, July 2008, Söderberg O. et al., "Characterizing proteins and their interactions in cells and tissues using the in situ proximity ligation assay", Pages 227-232, Copyright 2008, with permission from Elsevier.*

Microscopic images of the fluorescent dots can be overlaid with images of stained cellular nuclei and/or membranes from the same microscopic fields, to assign a nuclear or cytoplasmic localization to the protein complexes, as well as to count their number per cell.

In Study I, PLA was employed to examine the interactions between S100A4 and p53, as well as between S100A4 and MDM2 and between p53 and MDM2. Since antibodies might bind other proteins than the one they are intended to (off-target binding), three different antibody pairs were employed to validate the interaction between S100A4 and p53.

1.2 Studies II and III

1.2.1 Mass spectrometry-based proteomics and phosphoproteomics analyses

Mass spectrometry (MS) is a powerful and flexible technology for proteomics analysis. Bottom-up proteomics, which uses peptide identifications to infer protein identities, is the most widely employed proteomic approach to characterize proteins within complex

biological samples. Current MS-based proteomics technologies allow the identification and quantification of the relative abundances of more than 10,000 proteins in a cell [158].

The following paragraphs describe the steps of the workflows for bottom-up MS-based proteomics and phosphoproteomics analyses employed in Studies II and III. Large-scale phosphoproteomics analysis is challenging due to the low abundance and poor ionization efficiency of phosphorylated peptides [159]. To address these challenges, the workflows for phosphoproteomics analyses include phospho-enrichment by HiRIEF (Study II) or by titanium dioxide (Study III).

1.2.2 Protein extraction and solubilization

Proteins were extracted using a solution of 0.5% sodium deoxycholate (SDC) and 0.35% sodium lauroyl sarcosinate (SLS). These detergents are able to efficiently extract a wide variety of cellular proteins, including hydrophobic membrane-embedded proteins, which solubilize poorly [160]. Protein extraction and denaturation, was efficiently achieved by combining the use of these detergents with heating (95°C for 10 minutes) and chemical reduction and alkylation of cysteine residues, to disrupt disulfide bonds. Denaturation is necessary to render the protein accessible for enzymatic digestion. Detergents were removed by filter-aided sample preparation (FASP) [161]. This step is necessary for subsequent mass spectrometric analysis as detergents act as surfactants, impeding to the peptide ions to escape the electrospray droplets and enter the gas phase in the electrospray ionization (ESI) source (see paragraph 1.2.7 in this section), resulting in ion suppression. Recently, a novel method for detergent removal requiring a much shorter sample processing time and providing for higher experimental reproducibility, has been introduced. This method, employing paramagnetic beads and termed Single-Pot Solid-Phase-enhanced Sample Preparation (SP3), will likely be largely used in the coming years [162].

1.2.3 Protein digestion to peptides

Proteins were digested to peptides using the endoprotease trypsin alone (Study II) or in combination with Lys-C (Study III). Trypsin cleaves polypeptide molecules at the C-terminal side of lysine and arginine residues, while Lys-C only cleaves at the C-terminal side of lysine residues. The combination of Lys-C and trypsin increases the efficiency of digestion; Lys-C cleavages at lysine sites mediate partial protein unfolding, which facilitates the access of trypsin to tightly folded protein regions [163]. Additionally, a high enzyme to protein ratio (1:25 w/w) was employed in Study III to improve the digestion efficiency of phosphorylated proteins, as negative charges carried by pSer and pThr residues reduce the accessibility to cleavage sites positioned in their vicinity [164]. A high digestion efficiency is important to limit the prevalence of peptides containing missed cleaved sites, which increase the complexity of the peptide mixtures to be analyzed, impacting negatively on the depth of identifications. Additionally, missed cleaved peptides introduces errors in relative quantification, as the relative proportion of cleaved to missed cleaved peptides tend to vary across samples.

1.2.4 Phospho-peptide enrichment

Phosphorylation typically occurs on a small proportion of a protein pool, and consequently phosphorylated peptides are of low abundance. Most of serine and threonine phosphorylation events have been estimated to occur on less than 25% of the proteins in a protein pool (this proportion is referred to as fractional occupancy). The fractional occupancy is even lower for tyrosine phosphorylation events (less than 10%) [165]. Given their low abundance, phosphorylated peptides are often not identified when analyzing complex peptide samples. In order to increase the depth of phosphoproteomics analysis, several methods to enrich for phosphorylated peptides have been introduced. As tyrosine phosphorylation is particularly low abundant, specific enrichment methods have been introduced for the enrichment of tyrosine phosphorylated peptides, such as immunoprecipitation with antibodies targeting pTyr and purification using SH2 domain binders [166,167]. Very recently, it has been demonstrated that the employment of extensive pre-fractionation combined with high peptide loads can overcome the low abundance of phosphorylated species and allows to achieve good analytical depth without performing phospho-enrichment [168].

Several approaches can be used for phospho-peptide enrichment; most of these approaches make use of positively charged metal ions which are able to bind to the negatively charged phosphorylated peptides. Employed metals include Fr^{3+}, Ga^{3+}, Zr^{4+} and Ti^{4+} for immobilized metal affinity chromatography (IMAC), and again titanium, in the form of TiO_2 beads [166,169]; TiO_2 beads were employed in Study III. The affinity of peptides for TiO_2 beads correlates positively with the presence of acidic residues within the peptide sequence: the higher the number of acidic residues within a peptide and their acidic strength (measured by their pKa), the higher the affinity for TiO_2 beads. The phosphate group of phosphorylated amino acids has a higher acidic strength than the side chains of the acidic amino acids aspartic acid and glutamic acid (phosphate group pKa is 2.2, while the pKas of the side chains of aspartic acid and glutamic acid are 3.67 and 4.25, respectively [170]). This means that in most pH conditions, phosphate groups have a high capacity to donate protons and therefore carry a negative charge (see paragraph 2.2 in the Introduction section). The more negatively charged phosphate groups on a peptide, the higher the affinity for TiO_2. Consequently, the degree of binding affinity to TiO_2 beads proceeds in the order: multiply phosphorylated peptides > singly phosphorylated peptides > non-phosphorylated peptides containing acidic residues [171]. In Study III, two parameters were adjusted to increase the specificity of the binding of TiO_2 beads to phosphorylated peptides:

- **pH conditions employed for the binding.** In Study III a solution containing 6% trifluoroacetic acid (TFA) (pH ≅ 2) was employed for binding phospho-peptides to TiO_2 beads. At this pH the majority of Asp and Glu side chains are protonated and therefore neutral (no negative charge), while the phosphate group, whose pKa is 2.2, is negatively charged and can be specifically bound by the TiO_2 beads.
- **Proportion of TiO_2 beads to peptides.** The availability of binding sites on the TiO_2 beads impacts the equilibrium of bound versus unbound molecules. A high proportion of

TiO_2 beads to peptides results in the presence of residual free binding sites after high affinity peptides-beads interactions have taken place (i.e. phospho-peptides-beads interactions). In that case, low affinity interactions between residual free binding sites and non phosphorylated peptides may take place. On the other hand, if the proportion of TiO_2 beads to peptides is too low, there aren't enough sites on the beads to bind the phosphorylated peptides. In Study III a ratio of TiO_2 beads to peptides of 1:4 was employed, which was previously shown to provide for a good binding specificity [171].

1.2.5 Relative quantification by LC-MS: metabolic labeling by SILAC and isobaric labeling by TMT

Two methods for relative quantification were employed: stable isotope labeling with amino acids in cell culture (SILAC, Study II), and tandem mass tags (TMT, Study III). Both methods allow for multiplex analysis, which means that multiple samples can be analyzed in a single LC-MS run. For multiplexing, samples from different experimental conditions are labeled and pooled prior to MS analysis. SILAC multiplexing capability is limited to two or three samples (two were used in Study II), whereas TMT allow to simultaneously quantify ten samples.

In Study III, TMT labeling was performed after phospho-enrichment. Phospho-enrichment highly reduces the quantity of peptides per sample, therefore reducing the amount of TMT reagent required for labeling (a single set of TMT reagents was employed in Study III). For a detailed discussion of the basis of SILAC and TMT labeling, refer to published literature ([172], [173]).

1.2.6 Sample pre-fractionation by high-resolution isoelectric focusing (HiRIEF) and its application to phosphoproteomics analysis

Human cell proteomes are constituted by a very large number of chemically diverse proteins generated as a result of differential gene expression, alternative splicing, post-translational modifications and cleavage events (see paragraph 2.1 in the Introduction section), whose concentrations span seven orders of magnitude [174]. Additionally, protein digestion into peptides further increases the number of distinct molecular species present in the sample. Such complexity, both in terms of number of different molecules and in terms of their abundance, makes the task of identifying all of the molecules contained in a sample very challenging, disfavoring in particular the detection of low abundant proteins. Sample pre-fractionation is a widely employed approach to increase the analytical depth of proteomics analyses based on reducing the complexity of the samples to be analyzed by LC-MS [166]. s discussed above, phosphorylated proteins are of particularly low abundance, and therefore chromatographic pre-fractionation is often employed to improve their detection. Commonly employed pre-fractionation methods for phosphoproteomics analysis include Strong Cation Exchange (SCX) and high-pH reversed phase chromatography [175,176].

In Studies II and III of this thesis, HiRIEF was employed as a pre-fractionation method. HiRIEF separates peptides based on their isoelectric point (pI), using immobilized pH gradient (IPG) strips with varying pH ranges. After focusing on IPG strips, the peptides are extracted into 72 fractions, which are then analysed individually by LC-MS. Generated fractions are highly orthogonal, as the majority of peptides (typically 85-90%) are identified in one or at most two consecutive fractions. Analysis of the generated fractions by LC-MS results in an in-depth profiling of the cellular proteome [177].

Study II and Study III of this thesis describe for the first time the application of the HiRIEF fractionation method for phosphoproteomics analysis. At the time this thesis work was

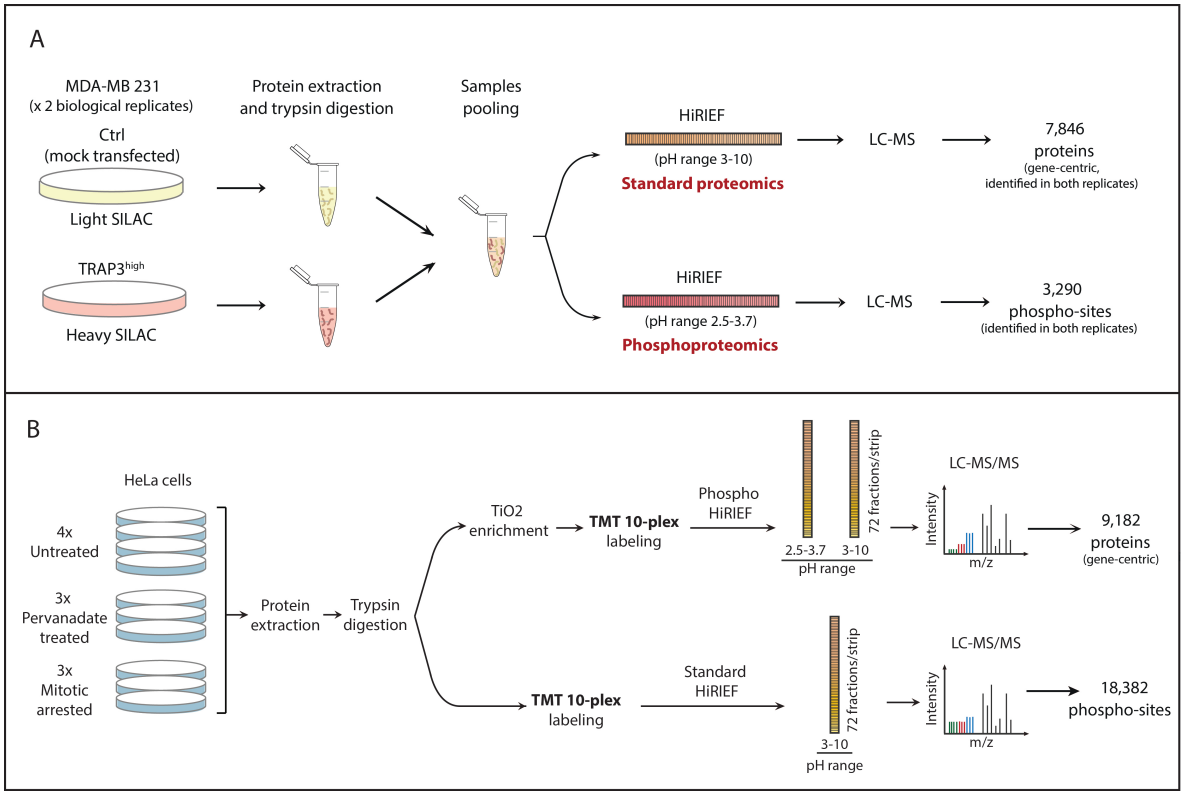

Figure 11. Workflows for phosphoproteomic analysis based on HiRIEF. A) Phosphoproteomics analysis based on the enrichment using a "ultra-acidic" IPG strip. **B)** TiO_2 enrichment is performed prior to fractionation of the samples on both the "ultra-acidic" and wide pH range (3-10) IPG strip, resulting in a high analytical depth. *Adapted from "Tartrate-resistant acid phosphatase (TRAP/ACP5) promotes metastasis-related properties via TGFβ2/TβR and CD44", Reithmeier A, Panizza E. et al., BMC Cancer (2017). In press; and from Scientific Reports 7, Article number: 4513 (2017), doi:10.1038/s41598-017-04798-z, "Isoelectric point-based fractionation by HiRIEF coupled to LC-MS allows for in-depth quantitative analysis of the phosphoproteome", Panizza E. et al.*

initiated we intended to test whether separation of peptides using a novel, prototypical, IPG gel strip with an "ultra-acidic" pH range (2.5-3.7) would be an efficient strategy to enrich for phosphorylated peptides. The rationale behind this idea is that phosphorylation decreases the pI of peptides due to the low pKa of the phosphate group (see paragraph 2.2 in the Introduction section). This idea had been tested before [178,179], but never with a separation technique with a high resolving power like HiRIEF. In Study II, we employed ultra-acidic IPG strips to separate samples containing 1 mg of peptides derived from breast cancer cells (**Figure 11A**). As it will be expanded upon in the Results section (paragraph 2*)*, this approach

resulted in a limited analytical depth, which then motivated us to start the optimization of a more complex workflow. Such effort resulted in the method presented in Study III, where peptides are enriched with TiO_2 beads prior to fractionation on IPG strips with pH range 2.5-3.7 and 3-10 (**Figure 11B**). Employment of the wide pH range 2.5-10 allowed for the identification of a large number of phospho-peptides with varied amino acid compositions and number of phosphorylations.

1.2.7 Phosphoproteomics analysis by LC-MS/MS using Orbitrap mass spectrometers

Phosphoproteomics analysis was performed using a LTQ Orbitrap Velos mass spectrometer in Study II, and a Q Exactive mass spectrometer in Study III. We will review here the main components of the instrumental setups employed in these studies as well as highlight the differences between the two instruments.

1) Nanoflow reversed-phase liquid chromatography (nano LC) system, coupled online to the mass spectrometer. Even after pre-fractionation, the degree of complexity of peptide mixtures derived from human cell samples is still very high. Separation by reversed-phase LC prior to MS analysis further simplifies the sample, thereby contributing to increase the analytical depth. Introduction of nano LC systems brought about major improvements for modern MS-based proteomics analysis [180]. The term nano LC derives from the rate of injection flow that these systems provide, which is on the order of nanoliters/minute. Such a low rate of injection flow permits the coupling to the nano-electrospray ionization (ESI) source, which in turns provide for high sensitivity due to the small size of the droplets it produces. Nano LC systems make use of chromatographic columns with a very small internal diameter. For this thesis work, chromatographic separation was performed on a 15 cm long C18 PicoFrit column with 100 μm of internal diameter in Study II (rate of injection flow 4 μl/min), and on a 50 cm long nano EASY-spray C18 column with 75 μm of internal diameter in Study III (rate of injection flow 250 nl/min).
2) Nano-ESI source: ESI is a so-called "soft" ionization system, as it can generate molecular ions without significant fragmentation, allowing the anaysis of intact polypeptides by MS [181].
3) Mass analyzer: both the LTQ Orbitrap Velos and the Q Exactive instruments are equipped with an high-resolution Orbitrap mass analyzer. The design of the Orbitrap mass analyzer allows for high resolution (the capacity to distiguish two peaks whose m/z differ by a very small value) and accurate mass detection [182]. The LTQ Velos also performs mass analysis in a ion trap for the the fragment ions produced by collision-induced dissociation (CID). Ions are selected for fragmentation using a quadrupole analyzer in Q Exactive instruments and using a ion trap in the LTQ Velos. Ion selection is faster in quadrupole analyzers than in ion traps [183].
4) Collision chamber: for MS2 analysis, ion fragmentation is performed by collision with gas molecules (typically N_2) in a collision chamber. The LTQ Velos and the Q Exactive are equipped with a different set of collision chambers: the LTQ Velos contains both a ion trap (used for CID) and an higher-energy collisional dissociation (HCD) collision

chamber, while the Q Exactive contains only a HCD collision chamber. There are several differences between those two fragmentation methods. Compared to CID, HCD produces more ion fragments, resulting in higher quality of the MS2 spectra. Additionally, CID fragmentation does not produce fragments ions with a m/z <= $0.3*(m/z)_{precursor\ ion}$ [184] resulting in a loss of information on low-mass m/z fragment ions. As reporter ions from isobaric labels such as TMT have a very low mass (the largest m/z is 131), HCD is the fragmentation method of election for isobaric labeling-based relative quantification. When operating in CID (collision-induced dissociation) mode, the LTQ Velos performs ion selection, fragmentation and mass analysis sequentially in the ion trap. Instead, the Q Exactive only operates in HCD (higher-energy collisional dissociation) mode, using the Orbitrap for mass analysis. This means that is is able to fill ions for fragmentation in the HCD collision chamber while scanning the fragments generated by the previous fragmentation event in the Orbitrap. Finally, fragment ions produced by HCD are analyzed at high resolution in the Orbitrap, while fragment ions produced by CID are detected at low resolution in the ion trap [185]. Detection in the ion trap is faster than in the Orbitrap.

2 Computational methods for the analysis and interpretation of large-scale proteomics and phosphoproteomics data

Quantitative large-scale proteomics and phosphoproteomics analyses generate datasets containing thousands of variables measured across several experimental conditions. Appropriate statistics and bioinformatic approaches are fundamental to analyze and interpret these large-scale datasets in order to formulate biological hypotheses. Some of the key computational approaches employed in Studies II and III of this thesis are summarized in the following sections.

2.1 Studies II and III

2.1.1 Hierarchical clustering

Hierarchical clustering is a method that allows for the grouping of sets of observations based on their similarity. In Studies II and III, sets of observations include the expression levels of proteins or phospho-sites across different experimental conditions. By grouping similar sets of observations, it is possible to find clusters of proteins (or phospho-sites), whose expression (or phosphorylation level) is regulated similarly across experimental conditions. Proteins (or phospho-sites) participating in the same cellular processes or that are part of a protein complex are often contained in such co-regulated clusters, therefore hierarchical clustering is a powerful approach for the interpretation of large-scale data.

In hierarchical clustering, clusters of observations are organized in several levels based on the degree of similarity between clusters, forming a tree-shaped diagram called a dendrogram. Dendrograms are built in a step-wise manner. At the lower level of a dendrogram, each single observation constitutes a cluster. A distance representing the degree of similarity between

each pair of observations is then calculated using a function (*distance metric*), and the two most similar observations are joined into a new cluster. The similarity between this new cluster and each of the other (old) clusters is then calculated (based on a certain *linkage method*) and the two most similar clusters are joined again; this step is repeated until all observations are joined into a single cluster. This bottom-up clustering approach is termed *agglomerative clustering*. In the resulting dendrogram, the height of each node is proportional to the degree of similarity between the two clusters it joins: the higher the node, the less similar the clusters (**Figure 12**).

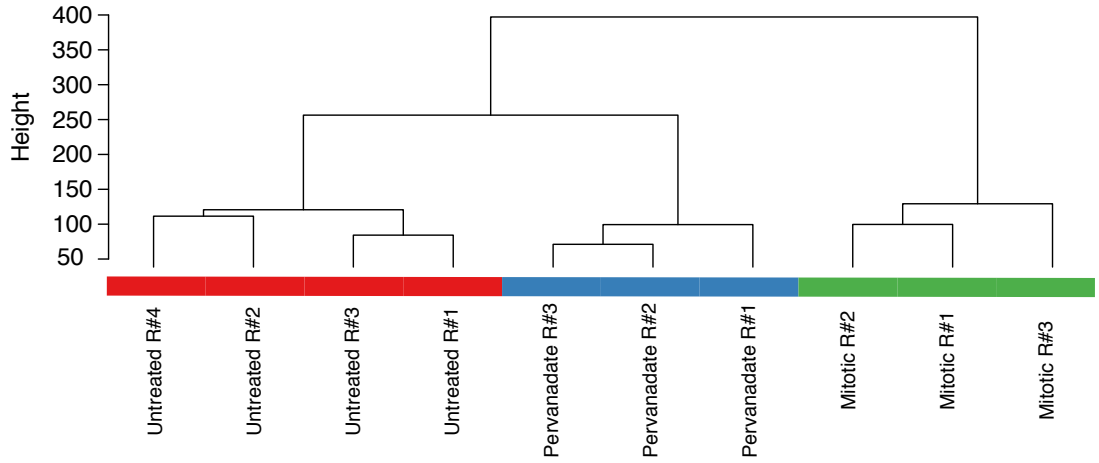

Figure 12. Dendrogram illustrating the arrangement of the clusters produced by hierarchical clustering based on Euclidian distance and complete linkage of ten observations. Ten observations derived from three experimental conditions measured in replicates (four, three and three replicates each) are clustered using Euclidian distance as distance metric and complete linkage as linkage method. Based on the dendrogram arrangement and node height, it is possible to observe that replicates of the same experimental condition cluster together, as well as that the cluster containing the three mitotic arrested samples differs the most compared to all of the other clusters. The color coding represents the cellular treatment: red - untreated samples; blue - pervanadate treated samples; green - mitotic arrested samples. *Adapted from Scientific Reports 7, Article number: 4513 (2017), doi:10.1038/s41598-017-04798-z, "Isoelectric point-based fractionation by HiRIEF coupled to LC-MS allows for in-depth quantitative analysis of the phosphoproteome", Panizza E. et al.*

Different approaches can be employed to calculate the distances between observations (by a *distance metric)* and between clusters (by a *linkage method)*. In Study II and III of this thesis Euclidian distance was employed as *distance metric* and complete linkage was employed as *linkage method* [186].

Distance metric – Euclidian distance

Euclidian distance is defined as:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

where *(1)*:

- *x* and y are vectors, each containing a set of observations (e.g. two vectors each containing the expression levels of a protein, or of a phospho-site, across different experimental conditions)
- *d(x,y)* is the distance between the two vectors (sets of observations)
- *i* is each observation in a set (e.g., each sample or replicate)
- *n* is the number of observations per set (e.g. the number of samples or replicates)

Euclidian distances always have a positive value, as the differences between each pair of measurements are squared. Another distance metric commonly used for hierarchical clustering is Pearson correlation; Pearson correlation coefficients have values ranging between -1 and 1, where -1 indicates a negative linear correlation, 0 indicates no correlation, and 1 indicates a positive linear correlation. Considering two sets of observations (e.g. two proteins) with a similar pattern of expression, Euclidian distance increases with the difference between the average of the two proteins expression level (different magnitude of regulation). The Pearson correlation coefficient instead would be close to 1 even in case the magnitude of regulation is different between the two proteins, as it expresses the existence of a linear correlation between the two sets of observations. Therefore, Euclidian distance will have a larger value in case of proteins regulated in the same direction (both up-regulated or both down-regulated) but to a different extent (different magnitude of regulation), than in the case of proteins regulated in the same direction and at the same extent, while the Pearson correlation coefficient will not differ. Therefore, Euclidian distance separates proteins (or phospho-sites), based not only on the pattern but also on the magnitude of expression. For this reason it was chosen for the analyses of the data from Study II and Study III of this thesis.

Linkage method – Complete linkage

Complete linkage is a method to calculate cluster distance, that defines the distance between two clusters as the longest distance between any two members of each of the two clusters. In other words, the distance between the two most dissimilar observations in each cluster defines the distance between the clusters. Other commonly used linkage methods include single linkage and average linkage, where the distance between two clusters are defined respectively as the shortest or as the average distance between the members of two clusters [187,188].

2.1.2 Gene ontology enrichment analysis

The gene ontology (GO) project provides a database of gene annotations describing the functionality of genes. This resource was employed in Studies II and III to examine the functionality of the identified proteins. In GO, gene functionality is described using terms belonging to three separate categories: biological process, molecular function and cellular component. Biological process terms describe the biological objective to which a gene or protein contributes. Molecular function terms describe the activities carried out at the

molecular level by a protein. Cellular component terms describe the parts of a cell or of its extracellular environment where a protein is present and/or active. Biological processes are accomplished through the execution of one or more molecular functions [189,190]. GO terms are organized hierarchically: child terms describe specialized processes that contribute to master term processes. For example, child terms of "cell cycle process" include: "cell cycle phase transition", "cytokinesis", "attachment of spindle microtubules to kinetochore", "centrosome duplication", "centrosome separation", and so on. As of today, the GO database contains 44,898 terms.

In Studies II and III, the GO database was employed to investigate whether the sets of proteins or phospho-sites that were differentially regulated in the examined experimental conditions were enriched in particular functions, as compared to the set of all proteins expressed in the studied cells. In order to perform this analysis, the terms describing the subset of regulated proteins (target set) are compared to the terms describing all the proteins identified in the analysis (background set). Over- or under-representation of GO terms in the target set compared to the background set were evaluated in terms of significance and fold-enrichment using the web tool GOrilla [191,192]. GO terms that are enriched in the target set represent the main cellular processes and functions that are regulated in the examined experimental conditions.

2.1.3 Protein-protein interaction network analysis

Protein-protein interaction analysis allows to infer the function of proteins based on their association with other proteins. For example, proteins displaying a high degree of connectivity tend to have important regulatory functions, and groups of highly connected proteins are usually involved in the same processes [192]. In Studies II and III, protein interaction networks were visualized using the software Cytoscape [193]. In protein-protein interaction networks, proteins are represented as nodes connected by edges representing some type of association between the proteins. The definition of association between proteins depends on the database used for the analysis.

In Study II, the database STRING was used to derive information on protein-protein interactions. The confidence of each protein-protein interaction contained in the database is expressed by a probability score, termed "combined score", which ranges between 0 and 1 (no confidence to very high confidence). The combined score is a corrected score computed combining the probabilities derived from the different evidences available for an interaction [194]. STRING takes into account 7 types of evidences for protein-protein interactions: experimental (e.g. derived from biochemical, biophysical or genetic experiments); database-derived (mainly from pathway databases, evaluated by a human expert curator); text mining (using Pubmed); co-expression (based on large-scale microarray and RNAseq data); neighborhood (assigned based on the conservation across species); fusion (assigned based on the existence of at least one organism where the pair of proteins considered has fused into a single protein-coding gene); co-occurrence (assigned based on the co-occurrence of the pair of genes examined in several organisms) [195,196]. In Study II, protein-protein interactions

with a combined score of at least 0.6 (representing a medium-high confidence of interaction) were employed in the analysis.

In Study III, the Cytoscape plugin PhosphoPath was employed for analysis [197]. PhosphoPath renders a phosphorylation site-specific visualization of the data; proteins and phospho-sites are represented as separate nodes. Phospho-site nodes can be colored based on the phospho-site quantitative information, therefore providing for an efficacious representation of phosphoproteomics data. Each phospho-site is connected to the protein on which it occurs by an edge. Edges are also present between proteins that interact with each other (based on BioGrid [198]) and between protein kinases and phospho-sites that are known to be their substrate (based on the PhosphoSitePlus database [199]).

2.1.4 Prediction of protein kinase-substrate association strength and the identification of putatively functional phosphorylation sites

NetworKIN was employed in Study III to obtain scores expressing kinases-phospho-sites association strength [200], as a means to predict phospho-site functionality [201]. The rationale underlying this approach is that a strong and highly specific association with a protein kinase is necessary to mediate quick and precise phosphorylation events (as regulatory phosphorylation events typically are). Therefore, the strength of association with protein kinases can be used as a predictor of phospho-site functionality.

NetworKIN scores combine two parameters to represent the strength of association between a phospho-site and a protein kinase: STRING scores (based on known protein-protein interactions [202]), and NetPhorest scores (based on the degree of matching between the sequence around the phosphorylated site and the motif recognized by the protein kinase [133]). NetworKIN reports the probability of association between 48,003 phosphorylation sites and 207 protein kinases. NetworKIN scores do not exist for the remaining 311 protein kinases due to lack of information on their interaction partners and/or on the motif they recognize on their substrates.

RESULTS AND DISCUSSION

1 Study I - S100A4 interacts with p53 in the nucleus and promotes p53 degradation.

S100A4 is overexpressed in several types of cancer and its expression is associated with poor prognosis and the development of metastasis [203]. A study conducted by Orre et al. had demonstrated p53-dependent expression of S100A4, as well as regulation of S100A4 in response to ionizing radiation [34]. Several studies had demonstrated an interaction between S100A4 and p53 by *in vitro* assays [40,204–208], but the interaction had never been shown in cells. In this study, we investigated the interaction between S100A4 and p53, and its consequences on p53 stability and on the growth and survival of the lung and cervix adenocarcinoma cell lines A549 and HeLa.

The main findings from the study include:

- S100A4 interacts with p53 in the nucleus of A549 and HeLa cells.
- S100A4-p53 interaction increases upon inhibition of p53 proteasomal degradation (by the MDM2 inhibitor Nutlin-3a or by the proteasome inhibitor MG132).
- S100A4 interacts with MDM2 in A549 cells, and the interaction between p53 and MDM2 increases upon inhibition of p53 proteasomal degradation by the proteasome inhibitor MG132.
- p53 protein but not transcript levels are increased upon knockdown of S100A4 in A549 and HeLa cells.
- p53 degradation rate is decreased upon knockdown of S100A4.
- The cell cycle arrests in G1 in response to S100A4 knockdown.
- Upon S100A4 knockdown, apoptosis increases and anchorage-independent growth decreases in response to cisplatin.

S100A4 was shown to interact with both p53 and MDM2 by PLA. A possible dynamic of these interactions might be that S100A4 binding to p53 favors its subsequent interaction with MDM2, thereby promoting the degradation of p53. Additional analyses would be necessary to clarify whether the three proteins exist as a complex or if these interactions occur sequentially. Furthermore, the interaction between S100A4 and MDM2 was demonstrated by PLA but was not confirmed by IP. Unbiased investigation of S100A4 interaction partners by MS analysis might clarify the nature and dynamics of these interactions, as well as possibly reveal additional S100A4 interaction partners.

2 Study II - Tartrate-resistant acid phosphatase (TRAP/ACP5) promotes metastasis-related properties via TGFβ2/TβR and CD44 in MDA-MB-231 breast cancer cells.

The protein phosphatase TRAP is overexpressed in several human cancers and is a marker for bone metastases [90–96]. In order to investigate the mechanisms mediating TRAP effects in cancer, we created three models: MDA-MB-231 breast cancer cells overexpressing TRAP ($TRAP3^{high}$ cells), and the same cells where TRAP overexpression was reverted using two different shRNA sequences (sh2 and sh3+4 cells). The proteome and phosphoproteome of these cell lines, as well as their phenotypes, were profiled and the main findings include:

- MDA-MB-231 cells display increased anchorage-independent growth, increased proliferation, migration and transwell invasion upon TRAP overexpression.
- Phosphoproteomics and proteomics analyses demonstrate a regulation of cellular adhesion and extra-cellular matrix organization processes upon perturbation of TRAP levels.
- Regulated processes involve several proteins in the TGFβ pathway, as well as three highly up-regulated phospho-sites on CD44.
- The proliferation and migration of TRAP overexpressing cells is reduced upon TGFβ2 and TGFβ receptor 1/2 (TβR1/2) inhibition.
- The proliferation of TRAP overexpressing cells is reduced upon blockage of CD44
- 116 phospho-sites significantly down-regulated upon TRAP overexpression were identified. The twenty top down-regulated phospho-sites are reported in **Table 1**

The 116 phospho-sites down-regulated in response to TRAP overexpression are candidates to be targets of TRAP phosphatase activity, but independent analyses would be necessary for their validation. Osteopontin and bone sialoprotein, which were previously described to be targets of TRAP [99], were not identified as dephosphorylated in response to TRAP overexpression in this study. The study might have failed to identify these proteins due to the limited analytical depth of the phospho-analysis (3,290 unique phosphorylation sites identified), which might have also hindered the identification of additional TRAP targets. As discussed in paragraph 1.2.6 in the Methods section, in this study the enrichment of phospho-peptides was solely based on HiRIEF fractionation using an ultra-acidic IPG strip (pH range 2.5-3.7). Phospho-peptides with such an acidic pI are mostly multiply phosphorylated peptides, or singly phosphorylated peptides containing a high number of acidic residues. Singly phosphorylated peptides with mildly acidic or basic amino acid composition are not identified by this approach, posing a major limitation to achieving a good analytical depth (**Figure 13A**). Besides those technical limitations, an additional challenge of identifying the targets of a protein phosphatase is that the abundance of the corresponding phospho-peptides is lowered (as they are dephosphorylated). In the case of a SILAC experiment, as the one utilized here, the control sample is the primary contributor to their quantity in the experimental sample.

Gene	Protein description	Phospho-site sequence window	Phospho site	Uniprot ID	Phospho-site log2(ratio) (duplicates average)
TEX2	Testis-expressed sequence 2 protein	APSSPLTSPSDTRSF	S270	Q8IWB9	-2.792
EIF5	Eukaryotic translation initiation factor 5	KEAEEESSGGEEEDE	S390	P55010	-2.362
EIF5	Eukaryotic translation initiation factor 5	LKEAEEESSGGEEED	S389	P55010	-2.362
FAM91A1	Protein FAM91A1	EDPADTASVSSLSLS	S364	Q658Y4	-2.353
STAG2	Cohesin subunit SA-2	DDTMSVISGISSRGS	S1061	Q8N3U4	-2.205
ITGB4	Isoform Beta-4D of Integrin beta-4	TSVSSHDSRLTAGVP	S1451	P16144-4	-2.118
RHBDF2	Isoform 2 of Inactive rhomboid protein 2	LPSQEAPSFQGTESP	S142	Q6PJF5-2	-2.022
RHBDF2	Isoform 2 of Inactive rhomboid protein 2	PSFQGTESPKPCKMP	S148	Q6PJF5-2	-2.022
RHBDF2	Isoform 2 of Inactive rhomboid protein 2	QRDLELPSQEAPSFQ	S137	Q6PJF5-2	-2.022
PNKP	Isoform 2 of Bifunctional polynucleotide phosphatase/kinase	EETRTPESQPDTPPG	S75	Q96T60-2	-1.972
SENP3	Sentrin-specific protease 3	AHPKNHLSPQQGGAT	S169	Q9H4L4	-1.902
SENP3	Sentrin-specific protease 3	GATPQVPSPCCRFDS	S181	Q9H4L4	-1.902
SENP3	Sentrin-specific protease 3	SPQQGGATPQVPSPC	T176	Q9H4L4	-1.902
PTGES3	Isoform 2 of Prostaglandin E synthase 3	ADDDSQDSDDEKMPD	S118	Q15185-2	-1.871
PTGES3	Isoform 2 of Prostaglandin E synthase 3	VDGADDDSQDSDDEK	S115	Q15185-2	-1.871
PNKP	Isoform 2 of Bifunctional polynucleotide phosphatase/kinase	QPDTPPGTPLVSQDE	T83	Q96T60-2	-1.863
PNKP	Isoform 2 of Bifunctional polynucleotide phosphatase/kinase	TPESQPDTPPGTPLV	T79	Q96T60-2	-1.863
MDC1	Isoform 3 of Mediator of DNA damage checkpoint protein 1	QTTTERDSDTDVEEE	S453	Q14676-3	-1.853
ARID4B	Isoform 2 of AT-rich interactive domain-containing protein 4B	IEVLSEDTDYEEDEV	T707	Q4LE39-2	-1.836
ARID4B	Isoform 2 of AT-rich interactive domain-containing protein 4B	RKDIEVLSEDTDYEE	S704	Q4LE39-2	-1.836

Table 1. Top 20 phospho-sites down-regulated in response to TRAP overexpression.

3 Study III - Isoelectric point-based fractionation by HiRIEF coupled to LC-MS allows for in-depth quantitative analysis of the phosphoproteome.

The low abundance of phosphorylated proteins constitutes a major challenge in achieving a good analytical depth in phosphoproteomics studies. This study presents an optimized workflow for phosphoproteomics analysis based on fractionation by HiRIEF on a wide pH range (IPG 2.5-3.7, "ultra-acidic" strip; and IPG 3-10 strip). A low amount of peptides per sample (300 µg/sample) was employed for phospho-enrichment with TiO_2, followed by labeling with TMT. Performance of TMT labeling after TiO_2 enrichment limits the amount of TMT reagent required, therefore considerably reducing the cost of the experiment.

The main findings from the study include:

- 18,382 phosphorylation sites were identified with high confidence localization and quantified in cervix adenocarcinoma cells HeLa.
- Phosphoproteomics analysis by HiRIEF requires low amount of material (300 µg/sample) and allows for multiplex quantification using TMT.
- 1,264 of the identified phospho-sites were novel.
- A high proportion of tyrosine phosphorylation sites was identified: 1,203 pY sites (6.3% of all the identified phospho-sites).
- HiRIEF fractionation using an IPG strip pH range 2.5-3.7 provides for enrichment of multiply phosphorylated peptides.
- 165 novel phospho-sites with a putative regulatory function during mitotic progression were identified by kinase association analysis.

Employment of a wide pH range (2.5-3.7 and 3-10) for fractionation resulted in the identification of phospho-peptides with both acid and basic amino acid composition (**Figure 13B**), highly increasing the depth of the analysis as compared to the workflow presented in Study II.

Specifically, the amino acid composition of the phospho-peptides impacts their distribution across HiRIEF fractions; very acidic phospho-peptides (DE>KHR) are identified in the pH range 2.5-3.7 strip and in the first 15 fractions of the pH range 3-10 strip, while basic phospho-peptides (DE<KR) are identified after fraction 15 of the IPG 3-10 strip (**Figure 14**). A possible improvement for future studies would be to reduce the extent of fractionation in order to decrease the analysis time, especially in the case of the IPG strip pH range 2.5-3.7, which has a relatively low complexity (6,537 phospho-peptides identified). In such a reduced fractionation protocol, keeping a separation between multiply phosphorylated and singly phosphorylated peptides (fractions 10-25 and fractions 40-72, respectively) would be beneficial in order to avoid the suppression of multiply phosphorylated peptide ionization by the singly phosphorylated peptides.

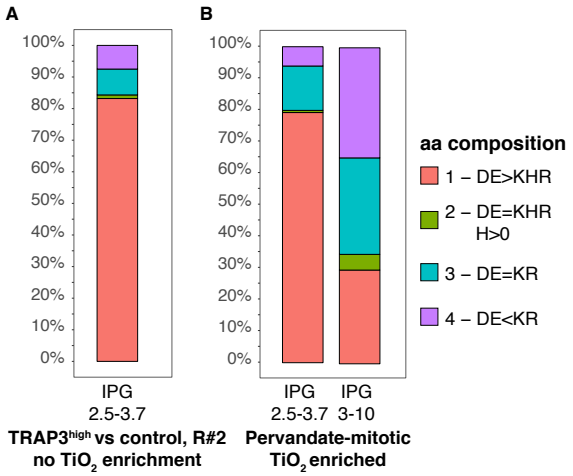

Figure 13. Distribution across the pI range of phospho-peptides based on their amino acid composition. A) 1 mg of peptides were separated on an IPG 2.5-3.7 strip in Study II to achieve phospho-enrichment; no TiO_2 phospho-enrichment was performed. The plot represents only phosphorylated peptides (3,892 phospho-peptides, from replicate #2). **B)** TiO_2 enriched peptides were further separated on IPG 2.5-3.7 and 3-10 strip in Study III. The plot represents only phosphorylated peptides (6,537 phospho-peptides in the IPG 2.5-3.7 and 15,294 phospho-peptides in the IPG 3-10).

An important issue that was addressed in this study concerns phospho-site functionality: as of today, only 3% of the reported human phosphorylation sites have a known regulatory function (total reported human phospho-sites: 224,661, corresponding to 17,495 gene products; reported regulatory phospho-sites: 6,695, corresponding to 2,187 gene products; PhosphoSitePlus, release 2017-07-19 [199]). In this study, phospho-site functionalities were predicted based on their strengths of interaction with protein kinases using NetworKIN [200,201], identifying 165 novel phospho-sites with a putatively regulatory function during mitosis. A limitation of this analysis is that the prediction is restricted to known protein kinase–substrate interactions. As discussed in paragraph 2.1.4 in the Methods section, the knowledge on protein kinases substrates is still partial, and association to substrates is not predicted for 311 of the 518 known human protein kinases. Consequently, the analysis is blind to the identification of functional phospho-sites that are phosphorylated by under-studied protein kinases, whose substrate consensus motif and/or interaction partners are not yet known.

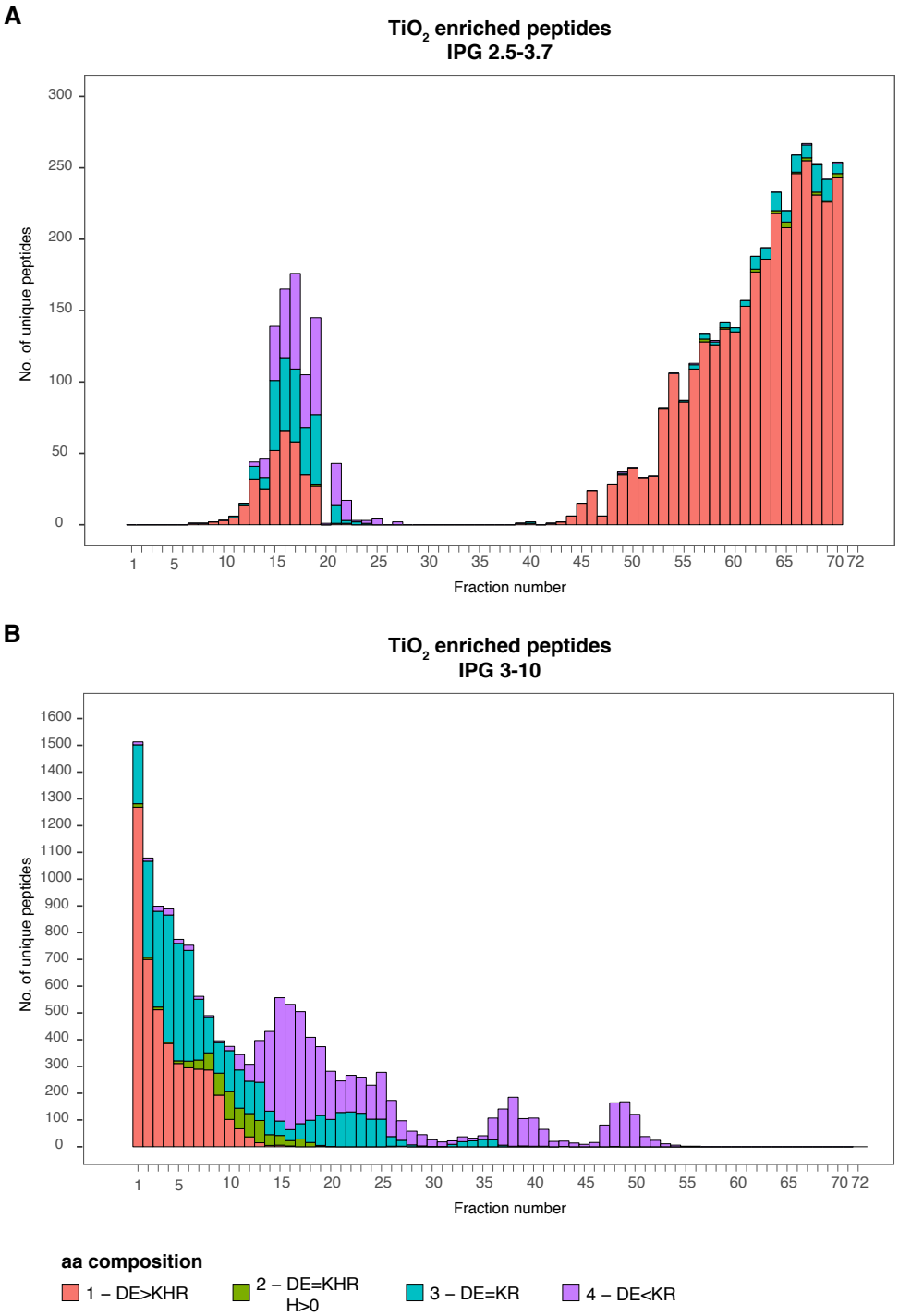

Figure 14. Distribution of unique phospho-peptides across HiRIEF fractions, broken down by amino acid composition. Phospho-peptides were enriched by TiO_2 prior to separation by HiRIEF (Study III). **A)** IPG 2.5-3.7; most identified phospho-peptides are very acidic (DE > KHR). The small number of phospho-peptides with a more basic amino acid composition identified in this pH range are multiply phosphorylated (fractions 10 to 25, see Fig. 2B in Paper III). **B)** IPG 3-10; the distribution clearly shows a separation between acidic phospho-peptides (DE > KHR), identified predominantly before fraction 15, and basic phospho-peptides (DE < KR), identified predominantly after fraction 15.

GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

1 General conclusions

This thesis focuses on investigating protein level regulatory mechanisms that contribute to the development of cancer, by applying and optimizing targeted and large-scale methods.

In Study I, targeted methods for the analyses of protein stability, protein localization and protein-protein interactions allowed to describe a mechanism by which S100A4 interacts with p53 in the nucleus thereby promoting its degradation, in lung and cervix adenocarcinoma cells.

In Studies II and III, the main objective was to optimize a workflow for phosphoproteomic analysis that would provide improvements over current methods, and to apply this workflow. In Study II, proteomics and phosphoproteomics analyses combined with targeted methods were employed to demonstrate that TRAP promotes metastasis-related cell properties in breast cancer cells via the TGFβ-pathway and CD44. In this study, a moderate-depth phosphoproteomic profiling of TRAP overexpressing cells was achieved by peptide fractionation by HiRIEF on IPG strips pH range 2.5-3.7, and provided a list of putative targets of TRAP phosphatase activity. In Study III, a workflow for in-depth quantitative phosphoproteomics analysis employing high-resolution isoelectric focusing (HiRIEF) fractionation on a wide pH range (2.5-10) was optimized. The workflow employs phospho-enrichment by titanium dioxide coupled with isobaric labeling by TMT, and provides for good analytical depth and sensitivity, requiring a low amount of starting material. Application of this workflow for the analysis of cervix adenocarcinoma cells HeLa revealed 1,264 novel phosphorylation sites, of which 165 phospho-sites that are suggested to have a regulatory function during the mitotic phase based on kinase-association analysis.

2 Future perspectives for the integration of large-scale and targeted methods

The investigation of different mechanisms of protein level regulation relies on the existence of dedicated analytical approaches. The work presented in this thesis contributes to the collective effort of improving these methods by focusing on the optimization of a workflow for phosphoproteomics analysis. Considering the proteomics field at large, an array of methods for the large-scale analysis of protein levels [119,209–213], post-translational modifications [214–219], protein-protein interactions [220–229], protein stability [230–233] and subcellular localization [230,234,235] exists. These methods can further be improved in terms of proteome coverage, lowering the amount of material and time required for the analysis, as well as providing the possibility to analyze a larger number of samples.

In the last decade, large-scale methods increasingly became important for biological research. Large-scale methods offer the possibility to obtain an unbiased global profiling of the system

being investigated. Large-scale methods for genomics and transcriptomics analysis have reached a significant level of maturity. The human genome project was initiated in 1990 and the first drafts of the human genome were published in 2001 [236,237]. Since then, extensive research efforts in the fields of genomics and transcriptomics led to the creation of resources that allow browsing of the human genome, both in a healthy state (UCSC Genome Browser [238]; https://genome.ucsc.edu) and in cancer (The Cancer Genome Atlas, TCGA [239,240]; https://cancergenome.nih.gov). The TCGA includes genomic maps of 33 types of cancer, generated from tumor tissues and matched normal tissues derived of more than 11,000 patients, corresponding to the astronomic total of 2.5 petabytes of data.

The proteomic field is much younger than the genomic field. The first publications containing the terms "human proteomic" and "human genomic" date back to 1992 and 1946, respectively, and the total number of publications in the field of genomic surpasses the ones in the field of proteomic by one order of magnitude (**Table 2**).

Search term	Number of publications	Year of first entry
Human proteomic	53,695	1992
Human phosphoproteomic	1,457	2001
Human genomic	697,769	1946
Cancer proteomic	20,831	1996
Cancer phosphoproteomic	775	2001
Cancer genomic	266,708	1951

Table 2. Number of publications in the proteomic and genomic field. The indicated terms were searched in Pubmed on August 25, 2017 (https://www.ncbi.nlm.nih.gov/pubmed).

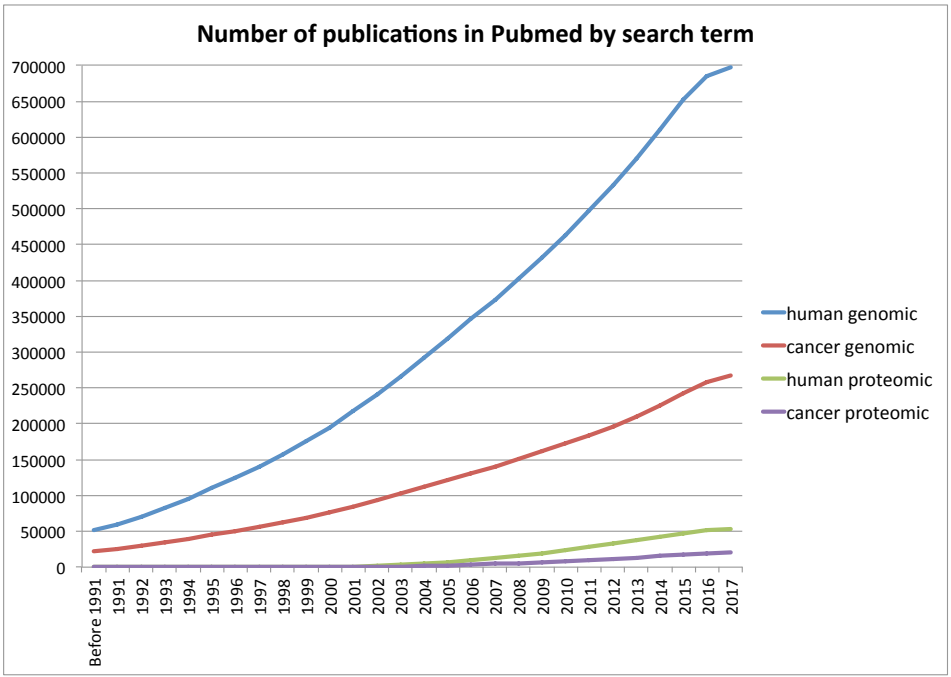

Figure 15. Rate of increase in the number of publications in the proteomic and genomic field since 1991. The indicated terms were searched in Pubmed on August 25, 2017 (https://www.ncbi.nlm.nih.gov/pubmed).

Promisingly, the number of publications in the proteomic field has been increasing rapidly in the last 10 years (**Figure 15**), and the first drafts of the human proteome were published in 2014 and in 2015, based on mass spectrometry analysis [117,241] and on microarray-based immunohistochemistry [118], respectively. Further developments of proteomics technology will be important to improve the description of the molecular phenotypes of biological systems, which cannot be obtained from genomic and transcriptomics analyses.

The emergence of large-scale methods is resulting in the generation of biological data faster than ever before, and the scientific community is currently faced with the challenge of transforming this enormous amount of data in biological understanding. An estimation of the ultimate impact of large-scale data can be provided by the examination of the number of drugs approved by the Food and Drug Administration (FDA) in recent years.

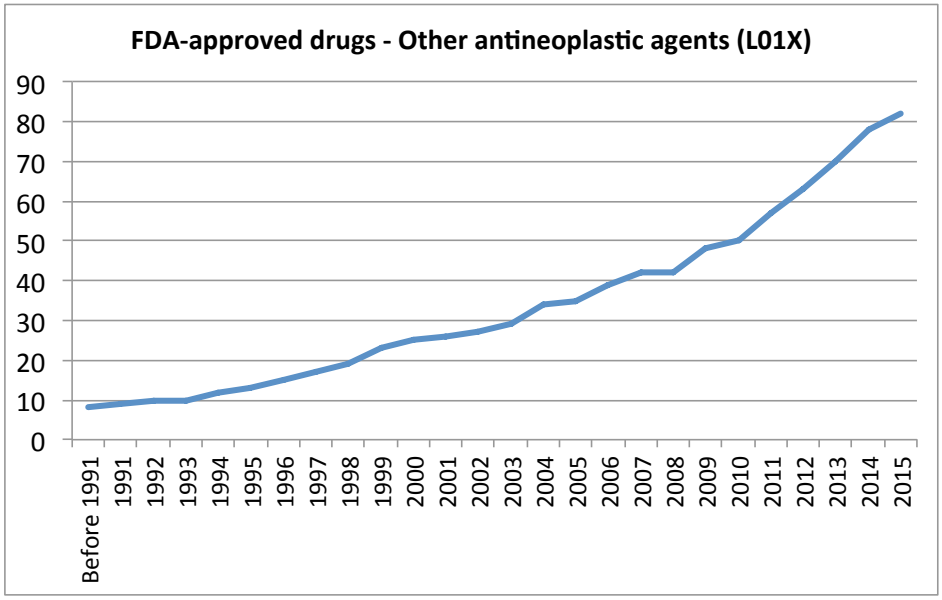

Figure 16. Number of FDA approved drugs in the category "Other antineoplastic agents (L01X)". The number of approved drugs per year was previously published [242]. Represented drugs are classified as "Other antineoplastic agents" (code L01X) according to the Anatomical Therapeutic Chemical Classification (ATC)**.**

Figure 16 represents the number of approved drugs in the category "Other antineoplastic agents". This category has been the fastest growing during the last 5 years [242], and includes monoclonal antibodies and kinase inhibitors, drugs that can be assigned to defined protein targets (other categories of antineoplastic drugs contain compounds that act as cytotoxic agents). Specifically, in this category twenty protein kinase inhibitors have been approved between 2011 and 2015. Even though these figures are positive, and despite the exponential growth of approved cancer drugs with well-defined targets, the absolute numbers of drugs being approved are still relatively small compared to the amount of biological data that is being produced. Additionally, the targets of FDA approved cancer drugs overlap only marginally with the cancer drivers discovered by genomic analyses, due to the fact that many of these cancer drivers are newly discovered cancer-associated genes on which little biological investigation has been performed [242,243] (**Figure 17**). Altogether, this data

highlights the need of improving the efficiency of generating biological understanding from large-scale data.

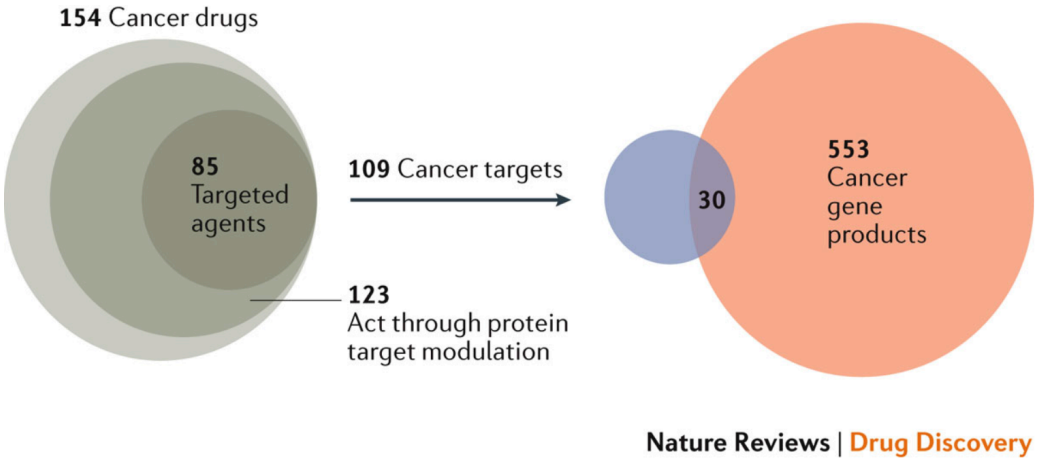

Figure 17. Overlap of FDA approved cancer drug targets with cancer drivers identified by genomic analyses. The 154 cancer drugs represented on the left hand side of the figure include broadly cytotoxic agents, such as DNA intercalating agents, cytotoxic agents that act by modulating a target protein that is not unique of cancer cells such as tubulin inhibitors, and targeted agents that have a clear, tumor-specific target such as kinase inhibitors and nuclear hormone receptor antagonists. The targets of the drugs contained in the latter group (109 cancer targets) were compared with a consensus reference list of cancer driver genes [243]. *Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Drug Discovery 16(1):19-34. A comprehensive map of molecular drug targets, Santos R et al.; doi: 10.1038/nrd.2016.230. Copyright 2016.*

In order to dissect this problem, consider the scheme presented in **Figure 18**. This schematic representation allows to decompose and analyze the different steps involved in the process leading from the production of large-scale data to the generation of novel biological understanding, and the expertise involved in each step. *Data* are generated by large-scale analyses (discovery phase); from the data, *information* is extracted by computational analyses (hypothesis building phase); the information is in turn used to produce *mechanistic knowledge* on the biological system being analyzed by targeted analyses (validation phase). Each step of this process requires different expertise. Successful completion of the process leads to the generation of novel *biological understanding* (i.e. explanation of a biological or molecular mechanism or process).

As of today, the scientific community is quite successful in converting data into information, which is made available in databases, or presented in scientific publications as hypothetical biological models. The conversion of such information into knowledge and subsequently into biological understanding though is lagging behind due to the contributions of different factors, including:

1) Lack of high throughput validation methods. Most of the methods currently available for targeted validation are low throughput. Very few methods exist that allow targeted analyses of a large number of samples in a limited time exist, e.g. luminex assays [244], proximity extension assays [245] and targeted proteomics [246]. Additional methods are

necessary to probe the roles of candidate molecules in determining the observed phenotypes, and prioritize their selection for further analyses.

2) Low level of integration between computationally-oriented and targeted biology-oriented research efforts. As of today, large-scale analyses and targeted analyses are often performed by distinct research groups. Groups performing targeted analyses often lack the expertise required to formulate biological questions based on large-scale data; similarly, groups performing large-scale analyses tend to have a strong computational component and do not follow-up their studies past the stage of formulating hypotheses. The result is a divide in the flow leading to the generation of biological understanding, which strongly hinders the successful completion of the process. Different measures can be taken to address this issue, including:
   - creation of mixed research groups, including both scientists with computational and wet lab skills, to promote communication and exchange of ideas
   - promoting the learning and the application of computational skills in wet lab research training positions, to reduce the distance between scientists with different expertise

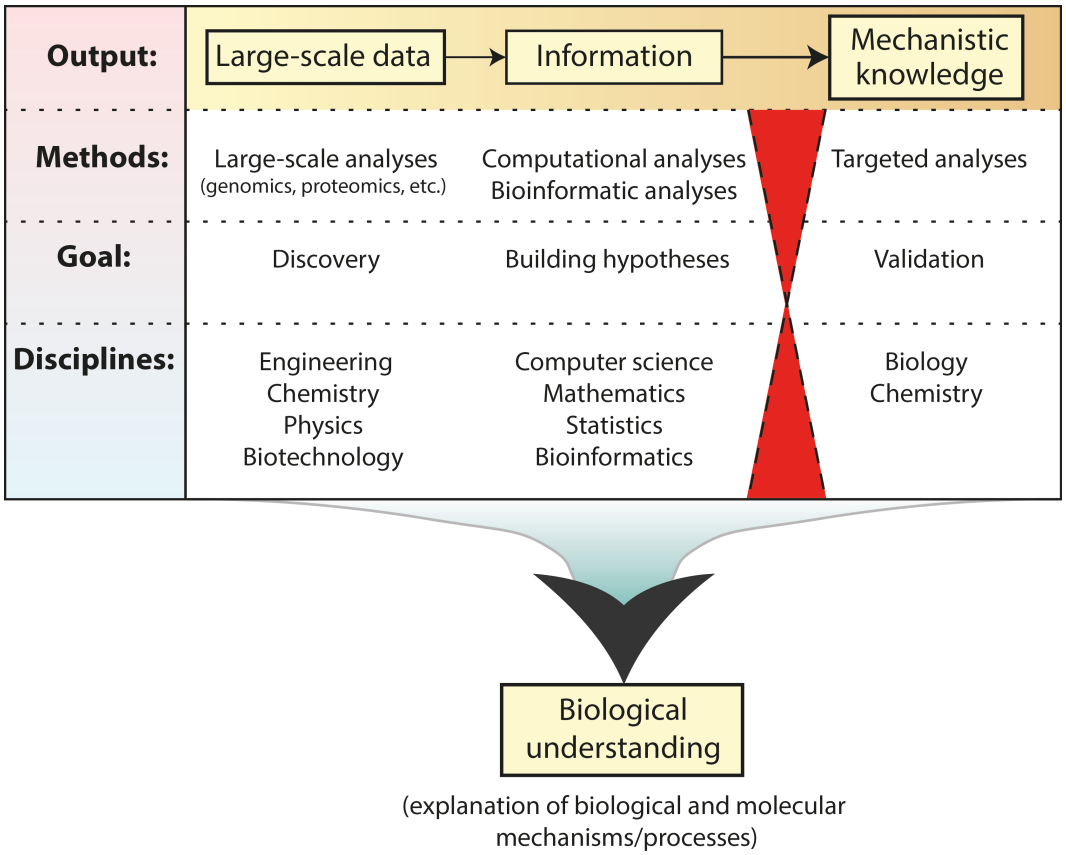

Figure 18. The interrupted flow of biological sciences. Schematic representation of the steps involved in the generation of novel biological understanding starting from large-scale biological data. The red double-funnel shape symbolically represents the current bottleneck in biological research, which, once overcome, will facilitate the use of the full potential of large-scale data. The methods, goals and disciplines involved in each step of the process are listed below the schematic of the flow. Logical dissection of the flow into its components provides a framework to address current shortcomings and to develop ways to address the bottleneck and increase the generation of novel biological understanding.

In conclusion, this is an incredibly exciting time for biological research. Large-scale methods provide unprecedented opportunities for increasing our understanding of life and to discover new treatments and cures for diseases. Major challenges lying ahead involve the sphere of methodological improvements (both for large-scale analyses, e.g. proteomics, and for targeted analyses) and progresses of the current models of research towards a higher level of integration of different areas of expertise.

ACKNOWLEDGEMENTS

My time as a doctoral student at Karolinska Institutet and in Sweden has been an incredible path of scientific and personal learning and development.

I would like first of all to thank my main supervisor Janne Lehtiö, for getting me on board in his lab and introducing me to the world of omics, always feeding me with new challenges and opportunities, and giving me freedom and space in my research. I also thank my co-supervisors Lukas Orre and Rui Branca: I was fortunate to have you with me during my Ph.D. Thanks for accompanying me through the everyday issues in the lab and for all the support. Lukas, thanks for the fun biology conversations! And Rui, thanks for the chemistry, and for teaching me some Portuguese, obrigada! I would also like to thank my co-supervisor Klas Wiman for our discussions on the projects progress.

Many thanks to all my colleagues, which made my everyday life in the lab a lot of fun! In particular, thanks to Jürgen Eirich for being my beers-mate; to Jorrit Boekel for the kind and fun conversations; to Rozbeh Jafari for being the party man; to Yafeng Zhu for being my doctoral studies mate and for all the adventures together, inside and outside of the lab; to Annsofi Sandberg, Maria Pernemalm and Henrik Johansson for the office chats; to Helena Bäckvall and Hillevi Andersson-Sand for being dear friends and of unique help in managing the lab life and for brightening it up with social events.

I would also like to thank all my collaborators and in particular Anja Reithmeier, Barbro Ek-Riklander and Göran Andersson for the study on TRAP. Thanks Anja for being such a sincere and collaborative co-author. It has been a lot of fun! And thanks also to Anita Aperia and Jacopo Fontana for involving me in the exciting adventure on the Na^+-K^+-ATPase and Ouabain.

And then of course, it has been great to share this adventure with the friends that I met during these years in Sweden (almost seven years!). First of all, thanks to Catia Cerquéira, Ulrika Morris and Ganna Olynik: thanks for all the nights together, the dances, the beans, the crumbles, the wine...I love you girls! And thanks to Daniel and Pedro for the fun times together, your archipelago kräftskiva and basement parties (respectively) have no rivals! Grazie a Walter Basile….Basile!! Grazie per i piani (!) che abbiamo fatto insieme, le birre, i discorsi, le risate. Grazie a Claudia Fredolini, Davide Tamburro e Elisa Pin per le chiacchierate e per portarmi ogni giorno un po' di vivacità italiana nella vita svedese. Thanks also to all the other friends I met during these years and that shared some of this adventure with me: Kate Brunner; Luisa Hugert; Alex and Christine; Lasse, Johan and Rikard; Mark and Hermes.

Grazie anche a tutti i miei amici in Italia, che sono sempre stati capaci di portarmi il sorriso e mi sono rimasti vicini anche nella lontananza: Francesca G. e Omar, Francesca P. e Andrea, Isabel, Veronica e Michela, Maristella, Alessandro, Feno e Sara. Grazie ragazzi! Vi voglio

bene!!!! Grazie anche a Martina per tutte le avventure insieme, dai tempi dell'università fino ad ora! Infine, grazie ad Elisabetta Marone, la mia amica-sorella, per esserci sempre. Che dire di più? …..Bettaaaaaaa!! ☺

Vorrei infine fare un ringraziamento particolare alla mia famiglia, che mi ha da sempre sostenuto e incoraggiato in innumerevoli modi: per aver incentivato il mio interesse per i libri, per aver sempre appoggiato le mie scelte di studio e per avermi insegnato a pensare liberamente. Grazie mamma e papà! Grazie Andrea e Roberta! Grazie alle piccole Matilde e Adelaide, sono felice di essere la vostra zia! E grazie alla nonna Lina!

Last but not least, thanks to my dear Nate for the time together in this adventure, for all the help and the support, for the fun conversations on science and academia and for inspiring me to take on new challenges. E per tutto il resto! ☺

This PhD has been funded by grants from Vetenskapsrådet (VR), Stiftelsen för strategisk forskning (SSF), Cancerfonden, Radiumhemmetsforskningsfonder, EU FP7 GlycoHit, GE Healthcare.

REFERENCES

1. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000; 100: 57–70. doi: 10.1007/s00262-010-0968-0.

2. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. Cell. 2011. p. 646–74. doi: 10.1016/j.cell.2011.02.013.

3. Elmore S. Apoptosis: a review of programmed cell death. Toxicol Pathol. NIH Public Access; 2007; 35: 495–516. doi: 10.1080/01926230701320337.

4. Weinberg RA (Robert A. The biology of cancer. 2. ed. New York: Garland Science; 2014. 28 p.

5. Bieging KT, Mello SS, Attardi LD. Unravelling mechanisms of p53-mediated tumour suppression. Nat Rev Cancer. NIH Public Access; 2014; 14: 359–70. doi: 10.1038/nrc3711.

6. Kubbutat MH., Vousden KH. Keeping an old friend under control: regulation of p53 stability. Mol Med Today. 1998; 4: 250–6. doi: 10.1016/S1357-4310(98)01260-X.

7. Bates S, Vousden KH. p53 in signaling checkpoint arrest or apoptosis. Curr Opin Genet Dev. 1996; 6: 12–8.

8. Friedman PN, Chen X, Bargonetti J, Prives C. The p53 protein is an unusually shaped tetramer that binds directly to DNA. Proc Natl Acad Sci U S A. 1993; 90: 3319–23. doi: 10.1073/pnas.90.8.3319.

9. Rotter V, Abutbul H, Ben-Ze'ev A. P53 transformation-related protein accumulates in the nucleus of transformed fibroblasts in association with the chromatin and is found in the cytoplasm of non-transformed fibroblasts. EMBO J. 1983; 2: 1041–7.

10. El-Deiry WS, Tokino T, Velculescu VE, Levy DB, Parsons R, Trent JM, Lin D, Mercer WE, Kinzler KW, Vogelstein B. WAF1, a potential mediator of p53 tumor suppression. Cell. 1993; 75: 817–25. doi: 10.1016/0092-8674(93)90500-P.

11. El-Deiry WS, Tokino T, Waldman T, Oliner JD, Velculescu VE, Burrell M, Hill DE, Healy E, Rees JL, Hamilton SR, Kinzler KW, Vogelstein B. Topological Control of p21WAF1/CIP1 Expression in Normal and Neoplastic Tissues. Cancer Res. 1995; 55: 2910–9.

12. Kubbutat MHG, Jones SN, Vousden KH. Regulation of p53 stability by Mdm2. Nature. 1997; 387: 299–303. doi: 10.1038/387299a0.

13. Haupt Y, Maya R, Kazaz A, Oren M. Mdm2 promotes the rapid degradation of p53. Nature. 1997; 387: 296–9. doi: 10.1038/387296a0.

14. Honda R, Tanaka H, Yasuda H. Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. FEBS Lett. 1997; 420: 25–7.

15. Barak Y, Juven T, Haffner R, Oren M. mdm2 expression is induced by wild type p53 activity. EMBO J. 1993; 12: 461–8.

16. Wu X, Bayle JH, Olson D, Levine AJ. The p53-mdm-2 autoregulatory feedback loop. Genes Dev. 1993; 7: 1126–32.

17. Talis AL, Huibregtse JM, Howley PM. The role of E6AP in the regulation of p53 protein levels in human papillomavirus (HPV)-positive and HPV-negative cells. J Biol Chem. American Society for Biochemistry and Molecular Biology; 1998; 273: 6439–45. doi: 10.1074/JBC.273.11.6439.

18. Baresova P, Musilova J, Pitha PM, Lubyova B. p53 tumor suppressor protein stability and transcriptional activity are targeted by Kaposi's sarcoma-associated herpesvirus-encoded viral interferon regulatory factor 3. Mol Cell Biol. American Society for Microbiology (ASM); 2014; 34: 386–99. doi: 10.1128/MCB.01011-13.

19. Ju J, Schmitz JC, Song B, Kudo K, Chu E. Regulation of p53 Expression in Response to 5-Fluorouracil

in Human Cancer RKO Cells. Clin Cancer Res. 2007; 13.

20. Honda R, Yasuda H. Association of p19(ARF) with Mdm2 inhibits ubiquitin ligase activity of Mdm2 for tumor suppressor p53. EMBO J. 1999; 18: 22–7. doi: 10.1093/emboj/18.1.22.

21. Sherr CJ, Weber JD, Taylor LJ, Roussel MF, Bar-Sagi D. Nucleolar Arf sequesters Mdm2 and activates p53. Nat Cell Biol. Nature Publishing Group; 1999; 1: 20–6. doi: 10.1038/8991.

22. Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. Nature. Nature Publishing Group; 2000; 408: 307–10. doi: 10.1038/35042675.

23. Muller PAJ, Vousden KH. p53 mutations in cancer. Nat Cell Biol. 2013; 15: 2–8. doi: 10.1038/ncb2641.

24. Rivlin N, Brosh R, Oren M, Rotter V. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. Genes Cancer. Impact Journals, LLC; 2011; 2: 466–74. doi: 10.1177/1947601911408889.

25. Hussain SP, Harris CC. Molecular Epidemiology of Human Cancer: Contribution of Mutation Spectra Studies of Tumor Suppressor Genes. Cancer Res. 1998; 58.

26. Hainaut P, Hollstein M. P53 and Human Cancer: the First Ten Thousand Mutations. Adv Cancer Res. 2000; 77: 81–137. doi: 10.1016/S0065-230X(08)60785-X.

27. Bártek J, Bártková J, Vojtěsek B, Stasková Z, Lukás J, Rejthar A, Kovarík J, Midgley CA, Gannon J V, Lane DP. Aberrant expression of the p53 oncoprotein is a common feature of a wide spectrum of human malignancies. Oncogene. 1991; 6: 1699–703.

28. Bresnick, Weber, Zimmer. S100 proteins in cancer. Nat Rev Cancer. 2015; 15: 96–109. doi: 10.1038/nrc3893.

29. Nonaka D, Chiriboga L, Rubin BP. Differential expression of S100 protein subtypes in malignant melanoma, and benign and malignant peripheral nerve sheath tumors. J Cutan Pathol. 2008; 35: 1014–9. doi: 10.1111/j.1600-0560.2007.00953.x; 10.1111/j.1600-0560.2007.00953.x.

30. Lee WY, Su WC, Lin PW, Guo HR, Chang TW, Chen HHW. Expression of S100A4 and Met: Potential Predictors for Metastasis and Survival in Early-Stage Breast Cancer. Oncology. 2004; 66: 429–38. doi: 10.1159/000079496.

31. de Silva Rudland S, Martin L, Roshanlall C, Winstanley J, Leinster S, Platt-Higgins A, Carroll J, West C, Barraclough R, Rudland P. Association of S100A4 and osteopontin with specific prognostic factors and survival of patients with minimally invasive breast cancer. Clin Cancer Res. 2006; 12: 1192–200. doi: 10.1158/1078-0432.CCR-05-1580.

32. Stewart RL, Carpenter BL, West DS, Knifley T, Liu L, Wang C, Weiss HL, Gal TS, Durbin EB, Arnold SM, O'Connor KL, Chen M. S100A4 drives non-small cell lung cancer invasion, associates with poor prognosis, and is effectively targeted by the FDA-approved anti-helminthic agent niclosamide. Oncotarget. Impact Journals, LLC; 2016; 7: 34630–42. doi: 10.18632/oncotarget.8969.

33. Rud AK, Lund-Iversen M, Berge G, Brustugun OT, Solberg SK, Mælandsmo GM, Boye K. Expression of S100A4, ephrin-A1 and osteopontin in non-small cell lung cancer. BMC Cancer. 2012; 12: 333. doi: 10.1186/1471-2407-12-333.

34. Orre LM, Pernemalm M, Lengqvist J, Lewensohn R, Lehti? J. Up-regulation, Modification, and Translocation of S100A6 Induced by Exposure to Ionizing Radiation Revealed by Proteomics Profiling. Mol Cell Proteomics. 2007; 6: 2122–31. doi: 10.1074/mcp.M700202-MCP200.

35. Lin J, Blake M, Tang C, Zimmer D, Rustandi RR, Weber DJ, Carrier F. Inhibition of p53 transcriptional activity by the S100B calcium-binding protein. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2001; 276: 35037–41. doi: 10.1074/jbc.M104379200.

36. Baudier J, Delphin C, Grunwald D, Khochbin S, Lawrence JJ. Characterization of the tumor suppressor protein p53 as a protein kinase C substrate and a SlOOb-binding protein. Biochemistry. 1992; 89: 11627–31.

37. Lin J, Yang Q, Yan Z, Markowitz J, Wilder PT, Carrier F, Weber DJ. Inhibiting S100B restores p53

levels in primary malignant melanoma cancer cells. J Biol Chem. 2004; 279: 34071–7. doi: 10.1074/jbc.M405419200.

38. Li ZH, Bresnick AR. The S100A4 metastasis factor regulates cellular motility via a direct interaction with myosin-IIA. Cancer Res. 2006; 66: 5173–80. doi: 10.1158/0008-5472.CAN-05-3087.

39. Ramagopal UA, Dulyaninova NG, Varney KM, Wilder PT, Nallamsetty S, Brenowitz M, Weber DJ, Almo SC, Bresnick AR. Structure of the S100A4/myosin-IIA complex. BMC Struct Biol. 2013; 13: 31. doi: 10.1186/1472-6807-13-31.

40. Fernandez-Fernandez MR, Veprintsev DB, Fersht AR. Proteins of the S100 family regulate the oligomerization of p53 tumor suppressor. Proc Natl Acad Sci. 2005; 102: 4735–40. doi: 10.1073/pnas.0501459102.

41. van Dieck J, Brandt T, Teufel DP, Veprintsev DB, Joerger AC, Fersht AR. Molecular basis of S100 proteins interacting with the p53 homologs p63 and p73. Oncogene. 2010; 29: 2024–35. doi: 10.1038/onc.2009.490.

42. van Dieck J, Lum JK, Teufel DP, Fersht AR. S100 proteins interact with the N-terminal domain of MDM2. FEBS Lett. 2010; 584: 3269–74. doi: 10.1016/j.febslet.2010.06.024.

43. Chaffer CL, Weinberg R a. A perspective on cancer cell metastasis. Science. 2011; 331: 1559–64. doi: 10.1126/science.1203543.

44. Hess KR, Varadhachary GR, Taylor SH, Wei W, Raber MN, Lenzi R, Abbruzzese JL. Metastatic patterns in adenocarcinoma. Cancer. 2006; 106: 1624–33. doi: 10.1002/cncr.21778.

45. Valastyan S, Weinberg RA, Gerald W, Hudis CA, Norton L, Smid M, Foekens JA, Massagué J, Lee YC, Chen HS, al. et. Tumor metastasis: molecular insights and evolving paradigms. Cell. Elsevier; 2011; 147: 275–92. doi: 10.1016/j.cell.2011.09.024.

46. Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, Groom AC. Multistep Nature of Metastatic Inefficiency. Am J Pathol. 1998; 153: 865–73. doi: 10.1016/S0002-9440(10)65628-3.

47. Kim J, Wong CW, Goldsmith JD, Song C, Fu W, Allion M-B, Herlyn M, Al-Mehdi AB, Muschel RJ. Rapid apoptosis in the pulmonary vasculature distinguishes non-metastatic from metastatic melanoma cells. Cancer Lett. 2004; 213: 203–12. doi: 10.1016/j.canlet.2004.03.042.

48. Gonzalez DM, Medici D. Signaling mechanisms of the epithelial-mesenchymal transition. Sci Signal. NIH Public Access; 2014; 7: re8. doi: 10.1126/scisignal.2005189.

49. Wheelock MJ, Shintani Y, Maeda M, Fukumoto Y, Johnson KR. Cadherin switching. J Cell Sci. 2008; 121.

50. Miettinen PJ, Ebner R, Lopez AR, Derynck R. TGF-?? induced transdifferentiation of mammary epithelial cells to mesenchymal cells: Involvement of type I receptors. J Cell Biol. Rockefeller Univ Press; 1994; 127: 2021–36. doi: 10.1083/jcb.127.6.2021.

51. Shintani Y, Wheelock MJ, Johnson KR. Phosphoinositide-3 Kinase-Rac1-c-Jun NH2-terminal Kinase Signaling Mediates Collagen I-induced Cell Scattering and Up-Regulation of N-Cadherin Expression in Mouse Mammary Epithelial Cells. Mol Biol Cell. 2006; 17: 2963–75. doi: 10.1091/mbc.E05-12-1123.

52. Shintani Y, Hollingsworth MA, Wheelock MJ, Johnson KR. Collagen I Promotes Metastasis in Pancreatic Cancer by Activating c-Jun NH2-Terminal Kinase 1 and Up-regulating N-Cadherin Expression. Cancer Res. 2006; 66: 11745–53. doi: 10.1158/0008-5472.CAN-06-2322.

53. Tian M, Neil JR, Schiemann WP. Transforming growth factor-β and the hallmarks of cancer. Cell Signal. 2011; 23: 951–62. doi: 10.1016/j.cellsig.2010.10.015.

54. Zöller M. CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? Nat Rev Cancer. Nature Publishing Group; 2011; 11: 254–67. doi: 10.1038/nrc3023.

55. Seoane J, Le H-V, Shen L, Anderson SA, Massagué J. Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. Cell. 2004; 117: 211–23.

56. Seoane J, Pouponnot C, Staller P, Schader M, Eilers M, Massagué J. TGFβ influences Myc, Miz-1 and Smad to control the CDK inhibitor p15INK4b. Nat Cell Biol. 2001; 3: 400–8. doi: 10.1038/35070086.

57. Moustakas A, Kardassis D. Regulation of the human p21ŴAF1Ćip1 promoter in hepatic cells by functional interactions between Sp1 and Smad family members. Biochemistry. 1998; 95: 6733–8.

58. Staller P, Peukert K, Kiermaier A, Seoane J, Lukas J, Karsunky H, Möröy T, Bartek J, Massagué J, Hänel F, Eilers M. Repression of p15INK4b expression by Myc through association with Miz-1. Nat Cell Biol. 2001; 3: 392–9. doi: 10.1038/35070076.

59. Claassen GF, Hann SR. A role for transcriptional repression of p21CIP1 by c-Myc in overcoming transforming growth factor beta -induced cell-cycle arrest. Proc Natl Acad Sci U S A. National Academy of Sciences; 2000; 97: 9498–503. doi: 10.1073/pnas.150006697.

60. Gartel AL, Shchors K. Mechanisms of c-myc-mediated transcriptional repression of growth arrest genes. Exp Cell Res. 2003; 283: 17–21.

61. Frederick JP, Liberati NT, Waddell DS, Shi Y, Wang X-F. Transforming growth factor beta-mediated transcriptional repression of c-myc is dependent on direct binding of Smad3 to a novel repressive Smad binding element. Mol Cell Biol. 2004; 24: 2546–59.

62. Chen C-R, Kang Y, Siegel PM, Massagué J. E2F4/5 and p107 as Smad cofactors linking the TGFbeta receptor to c-myc repression. Cell. 2002; 110: 19–32.

63. Perlman R, Schiemann WP, Brooks MW, Lodish HF, Weinberg RA. TGF-beta-induced apoptosis is mediated by the adapter protein Daxx that facilitates JNK activation. Nat Cell Biol. 2001; 3: 708–14. doi: 10.1038/35087019.

64. Zavadil J, Bitzer M, Liang D, Yang YC, Massimi A, Kneitz S, Piek E, Bottinger EP. Genetic programs of epithelial cell plasticity directed by transforming growth factor-beta. Proc Natl Acad Sci U S A. National Academy of Sciences; 2001; 98: 6686–91. doi: 10.1073/pnas.111614398.

65. Bakin A V, Rinehart C, Tomlinson AK, Arteaga CL. p38 mitogen-activated protein kinase is required for TGFbeta-mediated fibroblastic transdifferentiation and cell migration. J Cell Sci. 2002; 115: 3193–206.

66. Bakin A V., Tomlinson AK, Bhowmick NA, Moses HL, Arteaga CL. Phosphatidylinositol 3-Kinase Function Is Required for Transforming Growth Factor β-mediated Epithelial to Mesenchymal Transition and Cell Migration. J Biol Chem. 2000; 275: 36803–10. doi: 10.1074/jbc.M005912200.

67. Lamouille S, Derynck R. Cell size and invasion in TGF-beta-induced epithelial to mesenchymal transition is regulated by activation of the mTOR pathway. J Cell Biol. The Rockefeller University Press; 2007; 178: 437–51. doi: 10.1083/jcb.200611146.

68. Wu S, Cetinkaya C, Munoz-Alonso MJ, von der Lehr N, Bahram F, Beuger V, Eilers M, Leon J, Larsson L-G. Myc represses differentiation-induced p21CIP1 expression via Miz-1-dependent interaction with the p21 core promoter. Oncogene. Nature Publishing Group; 2003; 22: 351–60. doi: 10.1038/sj.onc.1206145.

69. Brenner C, Deplus R, Didelot C, Loriot A, Vir E, De Smet C, Gutierrez A, Danovi D, Bernard D, Boon T, Giuseppe Pelicci P, Amati B, Kouzarides T, et al. Myc represses transcription through recruitment of DNA methyltransferase corepressor. EMBO J. 2005; 24: 336–46. doi: 10.1038/sj.emboj.7600509.

70. Chen CR, Kang Y, Massagué J. Defective repression of c-myc in breast cancer cells: A loss at the core of the transforming growth factor beta growth arrest program. Proc Natl Acad Sci U S A. National Academy of Sciences; 2001; 98: 992–9. doi: 10.1073/pnas.98.3.992.

71. Boxer LM, Dang C V. Translocations involving c-myc and c-myc function. Oncogene. Nature Publishing Group; 2001; 20: 5595–610. doi: 10.1038/sj.onc.1204595.

72. Laiho M, Rönnstrand L, Heino J, Decaprio JA, Ludlow JW, Livingston DM, Massagué J. Control of junB and extracellular matrix protein expression by transforming growth factor-beta 1 is independent of simian virus 40 T antigen-sensitive growth-sensitive growth-inhibitory events. Mol Cell Biol. American Society for Microbiology; 1991; 11: 972–8. doi: 10.1128/MCB.11.2.972.

73. Conery AR, Cao Y, Thompson EA, Townsend CM, Ko TC, Luo K. Akt interacts directly with Smad3 to regulate the sensitivity to TGF-beta induced apoptosis. Nat Cell Biol. 2004; 6: 366–72.

74. Atfi A, Abécassis L, Bourgeade M-F. Bcr-Abl activates the AKT/Fox O3 signalling pathway to restrict transforming growth factor-beta-mediated cytostatic signals. EMBO Rep. European Molecular Biology Organization; 2005; 6: 985–91. doi: 10.1038/sj.embor.7400501.

75. Remy I, Montmarquette A, Michnick SW. PKB/Akt modulates TGF-beta signalling through a direct interaction with Smad3. Nat Cell Biol. 2004; 6: 358–65. doi: 10.1038/ncb1113.

76. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordón-Cardo C, Guise TA, Massagué J. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell. 2003; 3: 537–49.

77. Yin JJ, Selander K, Chirgwin JM, Dallas M, Grubbs BG, Wieser R, Massagué J, Mundy GR, Guise TA. TGF-beta signaling blockade inhibits PTHrP secretion by breast cancer cells and bone metastases development. J Clin Invest. American Society for Clinical Investigation; 1999; 103: 197–206. doi: 10.1172/JCI3523.

78. Toole BP, Slomiany MG. Hyaluronan: A constitutive regulator of chemoresistance and malignancy in cancer cells. Seminars in Cancer Biology. 2008. p. 244–50. doi: 10.1016/j.semcancer.2008.03.009.

79. Yoshihara S, Kon A, Kudo D, Nakazawa H, Kakizaki I, Sasaki M, Endo M, Takagaki K. A hyaluronan synthase suppressor, 4-methylumbelliferone, inhibits liver metastasis of melanoma cells. FEBS Lett. 2005; 579: 2722–6. doi: 10.1016/j.febslet.2005.03.079.

80. Takahashi E, Nagano O, Ishimoto T, Yae T, Suzuki Y, Shinoda T, Nakamura S, Niwa S, Ikeda S, Koga H, Tanihara H, Saya H. Tumor necrosis factor-alpha regulates transforming growth factor-beta-dependent epithelial-mesenchymal transition by promoting hyaluronan-CD44-moesin interaction. J Biol Chem. 2010; 285: 4060–73. doi: 10.1074/jbc.M109.056523.

81. Lundell BI, McCarthy JB, Kovach NL, Verfaillie CM. Activation of beta1 integrins on CML progenitors reveals cooperation between beta1 integrins and CD44 in the regulation of adhesion and proliferation. Leukemia. 1997; 11: 822–9.

82. Lokeshwar VB, Fregien N, Bourguignon LYW. Ankyrin-binding domain of CD44(GP85) is required for the expression of hyaluronic acid-mediated adhesion function. J Cell Biol. 1994; 126: 1099–109. doi: 10.1083/jcb.126.4.1099.

83. Fehon RG, McClatchey AI, Bretscher A. Organizing the cell cortex: the role of ERM proteins. Nat Rev Mol Cell Biol. 2010; 11: 674–674. doi: 10.1038/nrm2955.

84. Yu Q, Stamenkovic I. Localization of matrix metalloproteinase 9 to the cell surface provides a mechanism for CD44-mediated tumor invasion. Genes Dev. 1999; 13: 35–48. doi: 10.1101/gad.13.1.35.

85. Desai B, Ma T, Zhu J, Chellaiah MA. Characterization of the expression of variant and standard CD44 in prostate cancer cells: identification of the possible molecular mechanism of CD44/MMP9 complex formation on the cell surface. J Cell Biochem. 2009; 108: 272–84. doi: 10.1002/jcb.22248.

86. Yu Q, Stamenkovic I. Cell surface-localized matrix metalloproteinase-9 proteolytically activates TGF-β and promotes tumor invasion and angiogenesis. Genes Dev. 2000; 14: 163–76. doi: 10.1101/gad.14.2.163.

87. Wilson TJ, Nannuru KC, Futakuchi M, Singh RK. Cathepsin G-mediated enhanced TGF-β signaling promotes angiogenesis via upregulation of VEGF and MCP-1. Cancer Lett. 2010; 288: 162–9. doi: 10.1016/j.canlet.2009.06.035.

88. US Department of Health and Human Services. Bone health and osteoporosis: a report of the Surgeon General. US Heal Hum Serv. 2004; : 437. doi: 10.2165/00002018-200932030-00004.

89. Halleen JM, Tiitinen SL, Ylipahkala H, Fagerlund KM, Väänänen HK. Tartrate-resistant acid phosphates 5b (TRACP 5b) as a marker of bone resorption. Clin Lab. 2006; 52: 499–509.

90. Adams LM, Warburton MJ, Hayman AR. Human breast cancer cell lines and tissues express tartrate-resistant acid phosphatase (TRAP). Cell Biol Int. 2007; 31: 191–5. doi: 10.1016/j.cellbi.2006.09.022.

91. Honig A, Rieger L, Kapp M, Krockenberger M, Eck M, Dietl J, Kämmerer U. Increased tartrate-resistant acid phosphatase (TRAP) expression in malignant breast, ovarian and melanoma tissue: an investigational study. BMC Cancer. 2006; 6: 199. doi: 10.1186/1471-2407-6-199.

92. Capeller B, Caffier H, Sütterlin MW, Dietl J. Evaluation of tartrate-resistant acid phosphatase (TRAP) 5b as serum marker of bone metastases in human breast cancer. Anticancer Res. 2003; 23: 1011–5.

93. Chao TY, Wu YY, Janckila AJ. Tartrate-resistant acid phosphatase isoform 5b (TRACP 5b) as a serum maker for cancer with bone metastasis. Clin Chim Acta. Elsevier B.V.; 2010; 411: 1553–64. doi: 10.1016/j.cca.2010.06.027.

94. Lyubimova N V., Pashkov M V., Tyulyandin SA, Gol'dberg VE, Kushlinskii NE. Tartrate-resistant acid phosphatase as a marker of bone metastases in patients with breast cancer and prostate cancer. Bull Exp Biol Med. 2004; 138: 77–9. doi: 10.1023/B:BEBM.0000046945.95479.d6.

95. Wu YY, Janckila AJ, Ku CH, Yu CP, Yu JC, Lee SH, Liu HY, Yam LT, Chao TY. Serum tartrate-resistant acid phosphatase 5b activity as a prognostic marker of survival in breast cancer with bone metastasis. BMC Cancer. 2010; 10: 158. doi: 10.1186/1471-2407-10-158.

96. Zenger S, He W, Ek-Rylander B, Vassiliou D, Wedin R, Bauer H, Andersson G. Differential expression of tartrate-resistant acid phosphatase isoforms 5a and 5b by tumor and stromal cells in human metastatic bone disease. Clin Exp Metastasis. 2011; 28: 65–73. doi: 10.1007/s10585-010-9358-4.

97. Scott KL, Nogueira C, Heffernan TP, van Doorn R, Dhakal S, Hanna JA, Min C, Jaskelioff M, Xiao Y, Wu CJ, Cameron LA, Perry SR, Zeid R, et al. Proinvasion Metastasis Drivers in Early-Stage Melanoma Are Oncogenes. Cancer Cell. 2011; 20: 92–103. doi: 10.1016/j.ccr.2011.05.025.

98. Xia L, Huang W, Tian D, Chen Z, Zhang L, Li Y, Hu H, Liu J, Tang G, Dou J, Sha S, Xu B, Liu C, et al. ACP5, a direct transcriptional target of FoxM1, promotes tumor metastasis and indicates poor prognosis in hepatocellular carcinoma. Oncogene. 2014; 33: 1395–406. doi: 10.1038/onc.2013.90.

99. Ek-Rylander B, Flores M, Wendel M, Heinegard D, Andersson G. Dephosphorylation of osteopontin and bone sialoprotein by osteoclastic tartrate-resistant acid phosphatase. Modulation of osteoclast adhesion in vitro. J Biol Chem. 1994; 269: 14853–6.

100. Aguilera A, Gómez-González B. Genome instability: a mechanistic view of its causes and consequences. Nat Rev Genet. Nature Publishing Group; 2008; 9: 204–17. doi: 10.1038/nrg2268.

101. Weinberg RA. The retinoblastoma protein and cell cycle control. Cell. 1995; 81: 323–30. doi: 10.1016/0092-8674(95)90385-2.

102. Kastan MB, Bartek J. Cell-cycle checkpoints and cancer. Nature. Nature Publishing Group; 2004; 432: 316–23. doi: 10.1038/nature03097.

103. Ortega S, Malumbres M, Barbacid M. Cyclin D-dependent kinases, INK4 inhibitors and cancer. Biochim Biophys Acta - Rev Cancer. 2002; 1602: 73–87. doi: 10.1016/S0304-419X(02)00037-9.

104. Musacchio A. The Molecular Biology of Spindle Assembly Checkpoint Signaling Dynamics. Curr Biol. 2015; 25: R1002–18. doi: 10.1016/j.cub.2015.08.051.

105. Lara-Gonzalez P, Westhorpe FG, Taylor SS. The Spindle Assembly Checkpoint. Curr Biol. 2012; 22: R966–80. doi: 10.1016/j.cub.2012.10.006.

106. Geigl JB, Obenauf AC, Schwarzbraun T, Speicher MR. Defining ?chromosomal instability? Trends Genet. 2008; 24: 64–9. doi: 10.1016/j.tig.2007.11.006.

107. Thompson SL, Bakhoum SF, Compton DA. Mechanisms of chromosomal instability. Curr Biol. NIH Public Access; 2010; 20: R285-95. doi: 10.1016/j.cub.2010.01.034.

108. Lodish HF. Molecular cell biology. 6. ed. New York: W.H. Freeman; 2008.

109. Zieve GW, Turnbull D, Mullins JM, McIntosh JR. Production of large numbers of mitotic mammalian cells by use of the reversible microtubule inhibitor Nocodazole. Nocodazole accumulated mitotic cells. Exp Cell Res. 1980; 126: 397–405. doi: 10.1016/0014-4827(80)90279-7.

110. Morgan DO. Principles of CDK regulation. Nature. 1995; 374: 131–4. doi: 10.1038/374131a0.

111. Burgess A, Vuong J, Rogers S, Malumbres M, O'Donoghue SI. SnapShot: Phosphoregulation of Mitosis. Cell. 2017; 169: 1358–1358.e1. doi: 10.1016/j.cell.2017.06.003.

112. Medema RH, Lindqvist A. Boosting and suppressing mitotic phosphorylation. Trends in Biochemical Sciences. 2011. p. 578–84. doi: 10.1016/j.tibs.2011.08.006.

113. Olsen J V, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ, Gnad F, Cox J, Jensen TS, Nigg EA, Brunak S, Mann M, Martin L, et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal. 2010; 3: ra3. doi: 10.1126/scisignal.2000475.

114. Nigg EA. Mitotic kinases as regulators of cell division and its checkpoints. Nat Rev Mol Cell Biol. 2001; 2: 21–32. doi: 10.1038/35048096.

115. Harper JW, Bennett EJ. Proteome complexity and the forces that drive proteome imbalance. Nature. 2016; 537: 328–38. doi: 10.1038/nature19947.

116. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008; 40: 1413–5. doi: 10.1038/ng.259.

117. Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, Mathieson T, Lemeer S, Schnatbaum K, et al. Mass-spectrometry-based draft of the human proteome. Nature. 2014; 509: 582–7. doi: 10.1038/nature13319.

118. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, et al. Tissue-based map of the human proteome. Science (80-). 2015; 347: 1260419–1260419. doi: 10.1126/science.1260419.

119. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. Nature. 2011; 473: 337–42. doi: 10.1038/nature10098.

120. Maier T, Schmidt A, Guell M, Kuhner S, Gavin A-C, Aebersold R, Serrano L. Quantification of mRNA and protein and integration with protein turnover in a bacterium. Mol Syst Biol. 2014; 7: 511–511. doi: 10.1038/msb.2011.38.

121. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012; 13: 227–32. doi: 10.1038/nrg3185.

122. Lawless C, Holman SW, Brownridge P, Lanthaler K, Harman VM, Watkins R, Hammond DE, Miller RL, Sims PFG, Grant CM, Eyers CE, Beynon RJ, Hubbard SJ. Direct and Absolute Quantification of over 1800 Yeast Proteins via Selected Reaction Monitoring. Mol Cell Proteomics. 2016; 15: 1309–22. doi: 10.1074/mcp.M115.054288.

123. Minguez P, Parca L, Diella F, Mende DR, Kumar R, Helmer-Citterich M, Gavin A-C, van Noort V, Bork P. Deciphering a global network of functionally associated post-translational modifications. Mol Syst Biol. 2012; 8: 1–14. doi: 10.1038/msb.2012.31.

124. Hunter T. Why nature chose phosphate to modify proteins. Philos Trans R Soc B Biol Sci. 2012; 367: 2513–6. doi: 10.1098/rstb.2012.0013.

125. Lad C, Williams NH, Wolfenden R. The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. Proc Natl Acad Sci. 2003; 100: 5607–10. doi: 10.1073/pnas.0631607100.

126. Jin J, Pawson T. Modular evolution of phosphorylation-based signalling systems. Philos Trans R Soc B Biol Sci. 2012; 367: 2540–55. doi: 10.1098/rstb.2012.0106.

127. Hultquist DE. The preparation and characterization of phosphorylated derivatives of histidine. Biochim Biophys Acta - Bioenerg. 1968; 153: 329–40. doi: http://dx.doi.org/10.1016/0005-2728(68)90078-9.

128. Fuhs SR, Hunter T. pHisphorylation: the emergence of histidine phosphorylation as a reversible regulatory modification. Curr Opin Cell Biol. 2017; 45: 8–16. doi: 10.1016/j.ceb.2016.12.010.

129. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The Protein Kinase Complement of the Human Genome. Science (80-). 2002; 298: 1912–34. doi: 10.1126/science.1075762.

130. Chen MJ, Dixon JE, Manning G. Genomics and evolution of protein phosphatases. Sci Signal. 2017; 10: eaag1796. doi: 10.1126/scisignal.aag1796.

131. Manning G, Plowman GD, Hunter T, Sudarsanam S. Evolution of protein kinase signaling from yeast to man. Trends in Biochemical Sciences. 2002. p. 514–20. doi: 10.1016/S0968-0004(02)02179-5.

132. Hunter T. Tyrosine phosphorylation: thirty years and counting. Current Opinion in Cell Biology. 2009. p. 140–6. doi: 10.1016/j.ceb.2009.01.028.

133. Miller ML, Jensen LJ, Diella F, Jørgensen C, Tinti M, Li L, Hsiung M, Parker SA, Bordeaux J, Sicheritz-Ponten T, Olhovsky M, Pasculescu A, Alexander J, et al. Linear motif atlas for phosphorylation-dependent signaling. Sci Signal. 2008; 1: ra2. doi: 10.1126/scisignal.1159433.

134. Andersen JN, Mortensen OH, Peters GH, Drake PG, Iversen LF, Olsen OH, Peter G, Andersen HS, Tonks NK, Møller PH, Peters NH, Møller NPH. Structural and Evolutionary Relationships among Protein Tyrosine Phosphatase Domains. Mol Cell Biol. 2001; 21: 7117–36. doi: 10.1128/MCB.21.21.7117.

135. Shi Y. Serine/Threonine Phosphatases: Mechanism through Structure. Cell. 2009. p. 468–84. doi: 10.1016/j.cell.2009.10.006.

136. Roy J, Cyert MS. Cracking the phosphatase code: docking interactions determine substrate specificity. Sci Signal. 2009; 2: re9. doi: 10.1126/scisignal.2100re9.

137. Sacco F, Perfetto L, Castagnoli L, Cesareni G. The human phosphatase interactome: An intricate family portrait. FEBS Lett. Elsevier; 2012; 586: 2732–9. doi: 10.1016/j.febslet.2012.05.008.

138. Hornbeck P V., Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan M. PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. 2012; 40. doi: 10.1093/nar/gkr1122.

139. Duan G, Li X, Köhn M. The human DEPhOsphorylation database DEPOD: A 2015 update. Nucleic Acids Res. 2015; 43: D531–5. doi: 10.1093/nar/gku1009.

140. Beltrao P, Bork P, Krogan NJ, Van Noort V. Evolution and functional cross-talk of protein post-translational modifications. Molecular Systems Biology. 2013. doi: 10.1002/msb.201304521.

141. Ribeiro AAST, Ortiz V. A Chemical Perspective on Allostery. Chem Rev. American Chemical Society; 2016; 116: 6488–502. doi: 10.1021/acs.chemrev.5b00543.

142. Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, Davey NE, Babu MM, Lee R van der, Groot NS de, Gsponer J, Schwanhäusser B, Busse D, Li N, Dittmar G, et al. The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. Sci Signal. 2013; 6: rs7. doi: 10.1126/scisignal.2003345.

143. Nardozzi JD, Lott K, Cingolani G. Phosphorylation meets nuclear import: a review. Cell Commun Signal. 2010; 8: 32. doi: 10.1186/1478-811X-8-32.

144. Serber Z, Ferrell JE. Tuning Bulk Electrostatics to Regulate Protein Function. Cell. 2007; 128: 441–4. doi: 10.1016/j.cell.2007.01.018.

145. Mayer BJ, Hanafusa H. Association of the v-crk oncogene product with phosphotyrosine-containing proteins and protein kinase activity. Proc Natl Acad Sci U S A. 1990; 87: 2638–42.

146. Gross S, Rahal R, Stransky N, Lengauer C, Hoeflich KP. Targeting cancer with kinase inhibitors. Journal of Clinical Investigation. 2015. p. 1780–9. doi: 10.1172/JCI76094.

147. Fleuren EDG, Zhang L, Wu J, Daly RJ. The kinome "at large" in cancer. Nat Rev Cancer. 2016; 16: 83–98. doi: 10.1038/nrc.2015.18.

148. Gao T, Liu Z, Wang Y, Cheng H, Yang Q, Guo A, Ren J, Xue Y. UUCD: A family-based database of

ubiquitin and ubiquitin-like conjugation. Nucleic Acids Res. 2013; 41. doi: 10.1093/nar/gks1103.

149. Varshavsky A. Naming a targeting signal. Cell. 1991; 64: 13–5.

150. Varshavsky A. The Ubiquitin System, Autophagy, and Regulated Protein Degradation. Annu Rev Biochem. 2017; 86: 123–8. doi: 10.1146/annurev-biochem-061516-044859.

151. SIEGEL MR, SISLER HD. INHIBITION OF PROTEIN SYNTHESIS IN VITRO BY CYCLOHEXIMIDE. Nature. 1963; 200: 675–6.

152. Vassilev LT, Vu BT, Craves B, Carvajal D, Podlaski F, Filipovic Z, Kong N, Kammlott U, Lukacs C, Klein C, Fotouhi N, Liu EA. In Vivo Activation of the p53 Pathway by Small-Molecule Antagonists of MDM2. Science (80-). 2004; 303: 844–8. doi: 10.1126/science.1092472.

153. Shangary S, Wang S. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. Annu Rev Pharmacol Toxicol. 2009; 49: 223–41. doi: 10.1146/annurev.pharmtox.48.113006.094723.

154. Rock KL, Gramm C, Rothstein L, Clark K, Stein R, Dick L, Hwang D, Goldberg AL. Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. Cell. 1994; 78: 761–71. doi: 10.1016/S0092-8674(94)90462-6.

155. Jensen TJ, Loo MA, Pind S, Williams DB, Goldberg AL, Riordan JR. Multiple proteolytic systems, including the proteasome, contribute to CFTR processing. Cell. 1995; 83: 129–35. doi: 10.1016/0092-8674(95)90241-4.

156. Klockenbusch C, Kast J. Optimization of formaldehyde cross-linking for protein interaction analysis of non-tagged integrin beta1. J Biomed Biotechnol. Hindawi; 2010; 2010: 927585. doi: 10.1155/2010/927585.

157. Söderberg O, Leuchowius KJ, Gullberg M, Jarvius M, Weibrecht I, Larsson LG, Landegren U. Characterizing proteins and their interactions in cells and tissues using the in situ proximity ligation assay. Methods. 2008; 45: 227–32. doi: 10.1016/j.ymeth.2008.06.014.

158. Richards AL, Merrill AE, Coon JJ. Proteome sequencing goes deep. Curr Opin Chem Biol. 2015; 24: 11–7. doi: 10.1016/j.cbpa.2014.10.017.

159. Solari FA, Dell'Aica M, Sickmann A, Zahedi RP. Why phosphoproteomics is still a challenge. Mol Biosyst. 2015; 11: 1487–93. doi: 10.1039/c5mb00024f.

160. Masuda T, Tomita M, Ishihama Y. Phase Transfer Surfactant-Aided Trypsin Digestion for Membrane Proteome Analysis. J Proteome Res. American Chemical Society; 2008; 7: 731–40. doi: 10.1021/pr700658q.

161. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009; 6: 359–62. doi: 10.1038/nmeth.1322.

162. Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J. Ultrasensitive proteome analysis using paramagnetic bead technology. Mol Syst Biol. 2014; 10: 757.

163. Saveliev, S., Bratz, M., Zubarev, R., Szapacs, M., Budamgunta, H. & Urh M. Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. Nat Methods. 2013; : 87–114. doi: 10.1038/nmeth.f.371.

164. Dickhut C, Feldmann I, Lambert J, Zahedi RP. Impact of digestion conditions on phosphoproteomics. J Proteome Res. 2014; 13: 2761–70. doi: 10.1021/pr401181y.

165. Sharma K, D'Souza RCJ, Tyanova S, Schaab C, Wi??niewski J, Cox J, Mann M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. Cell Rep. 2014; 8: 1583–94. doi: 10.1016/j.celrep.2014.07.036.

166. Riley NM, Coon JJ. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. Anal Chem. 2016; 88: 74–94. doi: 10.1021/acs.analchem.5b04123.

167. Bian Y, Li L, Dong M, Liu X, Kaneko T, Cheng K, Liu H, Voss C, Cao X, Wang Y, Litchfield D, Ye

M, Li SS-C, et al. Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder. Nat Chem Biol. Nature Research; 2016; . doi: 10.1038/nchembio.2178.

168. Bekker-Jensen DB, Kelstrup CD, Batth TS, Larsen SC, Haldrup C, Bramsen JB, Sørensen KD, Høyer S, Ørntoft TF, Andersen CL, Nielsen ML, Olsen J V. An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. Cell Syst. 2017; 4: 587–599.e4. doi: 10.1016/j.cels.2017.05.009.

169. von Stechow L, editor. Phospho-Proteomics. New York, NY: Springer New York; 2016. doi: 10.1007/978-1-4939-3049-4.

170. Thurlkill RL, Grimsley GR, Scholtz JM, Pace CN. pK values of the ionizable groups of proteins. Protein Sci. Wiley-Blackwell; 2006; 15: 1214–8. doi: 10.1110/ps.051840806.

171. Li QR, Ning Z Bin, Tang JS, Nie S, Zeng R. Effect of peptide-to-TiO2 beads ratio on phosphopeptide enrichment selectivity. J Proteome Res. 2009; 8: 5375–81. doi: 10.1021/pr900659n.

172. Ong S-E, Mann M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat Protoc. Nature Publishing Group; 2006; 1: 2650–60. doi: 10.1038/nprot.2006.427.

173. Andrew Thompson †, Jürgen Schäfer ‡, Karsten Kuhn ‡, Stefan Kienle ‡, Josef Schwarz ‡, Günter Schmidt †, Thomas Neumann ‡ and, Christian Hamon* ‡. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. American Chemical Society ; 2003; . doi: 10.1021/AC0262560.

174. Zubarev RA. The challenge of the proteome dynamic range and its implications for in-depth proteomics. Proteomics. 2013; 13: 723–6. doi: 10.1002/pmic.201200451.

175. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villén J, Li J, Cohn MA, Cantley LC, Gygi SP. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A. 2004; 101: 12130–5. doi: 10.1073/pnas.0404720101.

176. Batth TS, Francavilla C, Olsen J V. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. J Proteome Res. 2014; 13: 6176–86. doi: 10.1021/pr500893m.

177. Branca RMM, Orre LM, Johansson HJ, Granholm V, Huss M, Pérez-Bercoff Å, Forshed J, Käll L, Lehtiö J. HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. Nat Methods. 2014; 11: 59–62. doi: 10.1038/nmeth.2732.

178. Hung C-W, Kübler D, Lehmann WD. pI-based phosphopeptide enrichment combined with nanoESI-MS. Electrophoresis. 2007; 28: 2044–52. doi: 10.1002/elps.200600678.

179. Maccarrone G, Kolb N, Teplytska L, Birg I, Zollinger R, Holsboer F, Turck CW. Phosphopeptide enrichment by IEF. Electrophoresis. 2006; 27: 4585–95. doi: 10.1002/elps.200600145.

180. Deterding LJ, Moseley MA, Tomer KB, Jorgenson JW. Nanoscale separations combined with tandem mass spectrometry. J Chromatogr A. 1991; 554: 73–82. doi: 10.1016/S0021-9673(01)88438-7.

181. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. Science. 1989; 246: 64–71.

182. Makarov* A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. American Chemical Society ; 2000; . doi: 10.1021/AC991131P.

183. Michalski A, Damoc E, Hauschild J-P, Lange O, Wieghaus A, Makarov A, Nagaraj N, Cox J, Mann M, Horning S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol Cell Proteomics. American Society for Biochemistry and Molecular Biology; 2011; 10: M111.011015. doi: 10.1074/mcp.M111.011015.

184. Louris JN, Cooks RG, Syka JEP, Kelley PE, Stafford Jr. GC, Todd JFJ. Instrumentation, Applications, and Energy Deposition in Quadrupole Ion-Trap Tandem Mass Spectrometry. Anal Chem. 1987; 59: 1677–85. doi: 10.1021/ac00140a021.

185. Jedrychowski MP, Huttlin EL, Haas W, Sowa ME, Rad R, Gygi SP. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. Mol Cell

Proteomics. American Society for Biochemistry and Molecular Biology; 2011; 10: M111.009910. doi: 10.1074/mcp.M111.009910.

186. D'Haeseleer P. How does gene expression clustering work? Nat Biotechnol. 2005; 23: 1499–501. doi: 10.1038/nbt1205-1499.

187. Mooi E, Sarstedt M. A Concise Guide to Market Research. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. doi: 10.1007/978-3-642-12541-6.

188. Sarstedt M, Mooi E. Cluster Analysis. Springer, Berlin, Heidelberg; 2014. p. 273–324. doi: 10.1007/978-3-642-53965-7_9.

189. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, et al. Gene Ontology: tool for the unification of biology. Nat Genet. Nature Publishing Group; 2000; 25: 25–9. doi: 10.1038/75556.

190. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015; 43. doi: 10.1093/nar/gku1179.

191. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z, Mao M, Peterse H, Kooy K der, Marton M, Witteveen A, Schreiber G, Kerkhoven R, Roberts C, et al. GOrilla : a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinforma 2009 101. BioMed Central; 2009; 30: 530–6. doi: 10.1002/1097-024x(200009)30:11<1203::aid-spe338>3.0.co.

192. Munk S, Refsgaard JC, Olsen J V. Systems analysis for interpretation of phosphoproteomics data. Methods in Molecular Biology. 2016. p. 341–60. doi: 10.1007/978-1-4939-3049-4_23.

193. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. Cold Spring Harbor Laboratory Press; 2003; 13: 2498–504. doi: 10.1101/gr.1239303.

194. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res. Oxford University Press; 2004; 33: D433–7. doi: 10.1093/nar/gki005.

195. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. Oxford University Press; 2017; 45: D362–8. doi: 10.1093/nar/gkw937.

196. Franceschini A, Lin J, von Mering C, Jensen LJ. SVD-phy: improved prediction of protein functional associations through singular value decomposition of phylogenetic profiles. Bioinformatics. Oxford University Press; 2016; 32: 1085–7. doi: 10.1093/bioinformatics/btv696.

197. Raaijmakers LM, Giansanti P, Possik PA, Mueller J, Peeper DS, Heck AJR, Altelaar AFM. PhosphoPath: Visualization of Phosphosite-centric Dynamics in Temporal Molecular Networks. J Proteome Res. 2015; 14: 4332–41. doi: 10.1021/acs.jproteome.5b00529.

198. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, Reguly T, Nixon J, Ramage L, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res. 2015; 43: D470–8. doi: 10.1093/nar/gku1204.

199. Hornbeck P V., Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. Nucleic Acids Res. 2015; 43: D512–20. doi: 10.1093/nar/gku1267.

200. Linding R, Jensen LJ, Pasculescu A, Olhovsky M, Colwill K, Bork P, Yaffe MB, Pawson T. NetworKIN: A resource for exploring cellular phosphorylation networks. Nucleic Acids Res. 2008; 36. doi: 10.1093/nar/gkm902.

201. Xiao Q, Miao B, Bi J, Wang Z, Li Y. Prioritizing functional phosphorylation sites based on multiple feature integration. Sci Rep. The Author(s); 2016; 6: 24735.

202. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015; 43: D447-52. doi: 10.1093/nar/gku1003.

203. Fei F, Qu J, Zhang M, Li Y, Zhang S. S100A4 in cancer progression and metastasis: A systematic review. Oncotarget. 2015; . doi: 10.18632/oncotarget.18016.

204. Chen H, Fernig DG, Rudland PS, Sparks A, Wilkinson MC, Barraclough R. Binding to Intracellular Targets of the Metastasis-Inducing Protein, S100A4 (p9Ka). Biochem Biophys Res Commun. 2001; 286: 1212–7. doi: 10.1006/bbrc.2001.5517.

205. Grigorian M, Andresen S, Tulchinsky E, Kriajevska M, Carlberg C, Kruse C, Cohn M, Ambartsumian N, Christensen A, Selivanova G, Lukanidin E. Tumor Suppressor p53 Protein Is a New Target for the Metastasis-associated Mts1/S100A4 Protein. J Biol Chem. 2001; 276: 22699–708. doi: 10.1074/jbc.M010231200.

206. Ismail TM, Zhang S, Fernig DG, Gross S, Martin-Fernandez ML, See V, Tozawa K, Tynan CJ, Wang G, Wilkinson MC, Rudland PS, Barraclough R. Self-association of Calcium-binding Protein S100A4 and Metastasis. J Biol Chem. 2010; 285: 914–22. doi: 10.1074/jbc.M109.010892.

207. van Dieck J, Fernandez-Fernandez MR, Veprintsev DB, Fersht AR. Modulation of the Oligomerization State of p53 by Differential Binding of Proteins of the S100 Family to p53 Monomers and Tetramers. J Biol Chem. 2009; 284: 13804–11. doi: 10.1074/jbc.M901351200.

208. van Dieck J, Teufel DP, Jaulent AM, Fernandez-Fernandez MR, Rutherford TJ, Wyslouch-Cieszynska A, Fersht AR. Posttranslational Modifications Affect the Interaction of S100 Proteins with Tumor Suppressor p53. J Mol Biol. 2009; 394: 922–30. doi: 10.1016/j.jmb.2009.10.002.

209. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. Nature. 2016; 537: 347–55. doi: 10.1038/nature19949.

210. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. Annu Rev Biomed Eng. Annual Reviews ; 2009; 11: 49–79. doi: 10.1146/annurev-bioeng-061008-124934.

211. Breker M, Schuldiner M. The emergence of proteome-wide technologies: systematic analysis of proteins comes of age. Nat Rev Mol Cell Biol. 2014; 15: 453–64. doi: 10.1038/nrm3821.

212. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods. 2014; 11: 319–24. doi: 10.1038/nmeth.2834.

213. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. Mol Cell Proteomics. 2012; 11: O111.016717. doi: 10.1074/mcp.O111.016717.

214. Olsen J V, Mann M. Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol Cell Proteomics. American Society for Biochemistry and Molecular Biology; 2013; 12: 3444–52. doi: 10.1074/mcp.O113.034181.

215. Mertins P, Qiao JW, Patel J, Udeshi ND, Clauser KR, Mani DR, Burgess MW, Gillette MA, Jaffe JD, Carr SA. Integrated proteomic analysis of post-translational modifications by serial enrichment. Nat Methods. 2013; 10: 634–7. doi: 10.1038/nmeth.2518.

216. von Stechow L, Francavilla C, Olsen J V. Recent findings and technological advances in phosphoproteomics for cells and tissues. Expert Rev Proteomics. 2015; 12: 469–87. doi: 10.1586/14789450.2015.1078730.

217. Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen J V., Mann M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. Science (80-). 2009; 325.

218. Chuh KN, Pratt MR. Chemical methods for the proteome-wide identification of posttranslationally modified proteins. Curr Opin Chem Biol. NIH Public Access; 2015; 24: 27–37. doi: 10.1016/j.cbpa.2014.10.020.

219. Hahne H, Sobotzki N, Nyberg T, Helm D, Borodkin VS, van Aalten DMF, Agnew B, Kuster B. Proteome Wide Purification and Identification of *O* -GlcNAc-Modified Proteins Using Click Chemistry

and Mass Spectrometry. J Proteome Res. American Chemical Society; 2013; 12: 927–36. doi: 10.1021/pr300967y.

220. Li X, Wang W, Chen J. From pathways to networks: Connecting dots by establishing protein-protein interaction networks in signaling pathways using affinity purification and mass spectrometry. Proteomics. 2015; 15: 188–202. doi: 10.1002/pmic.201400147.

221. Rhee H-W, Zou P, Udeshi ND, Martell JD, Mootha VK, Carr SA, Ting AY. Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. Science (80-). 2013; 339.

222. Guruharsha KG, Rual J-F, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY, Cenaj O, McKillip E, Shah S, Stapleton M, et al. A Protein Complex Network of Drosophila melanogaster. Cell. 2011; 147: 690–703. doi: 10.1016/j.cell.2011.08.047.

223. Hubner NC, Bird AW, Cox J, Splettstoesser B, Bandilla P, Poser I, Hyman A, Mann M. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J Cell Biol. 2010; 189.

224. Kirkwood KJ, Ahmad Y, Larance M, Lamond AI. Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. Mol Cell Proteomics. American Society for Biochemistry and Molecular Biology; 2013; 12: 3851–73. doi: 10.1074/mcp.M113.032367.

225. Havugimana PC, Hart GT, Nepusz T, Yang H, Turinsky AL, Li Z, Wang PI, Boutz DR, Fong V, Phanse S, Babu M, Craig SA, Hu P, et al. A Census of Human Soluble Protein Complexes. Cell. 2012; 150: 1068–81. doi: 10.1016/j.cell.2012.08.011.

226. Kristensen AR, Gsponer J, Foster LJ. A high-throughput approach for measuring temporal changes in the interactome. Nat Methods. 2012; 9: 907–9. doi: 10.1038/nmeth.2131.

227. Leitner A, Walzthoeni T, Aebersold R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. Nat Protoc. 2014; 9: 120–37. doi: 10.1038/nprot.2013.168.

228. Fischer L, Chen ZA, Rappsilber J. Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. J Proteomics. 2013; 88: 120–8. doi: 10.1016/j.jprot.2013.03.005.

229. Weisbrod CR, Chavez JD, Eng JK, Yang L, Zheng C, Bruce JE. *In Vivo* Protein Interaction Network Identified with a Novel Real-Time Cross-Linked Peptide Identification Strategy. J Proteome Res. American Chemical Society; 2013; 12: 1569–79. doi: 10.1021/pr3011638.

230. Boisvert F-M, Ahmad Y, Gierliński M, Charrière F, Lamont D, Scott M, Barton G, Lamond AI. A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. Mol Cell Proteomics. 2012; 11: M111.011429. doi: 10.1074/mcp.M111.011429.

231. Toyama BH, Savas JN, Park SK, Harris MS, Ingolia NT, Yates JR, Hetzer MW. Identification of Long-Lived Proteins Reveals Exceptional Stability of Essential Cellular Structures. Cell. 2013; 154: 971–82. doi: 10.1016/j.cell.2013.07.037.

232. Claydon AJ, Thom MD, Hurst JL, Beynon RJ. Protein turnover: Measurement of proteome dynamics by whole animal metabolic labelling with stable isotope labelled amino acids. Proteomics. 2012; 12: 1194–206. doi: 10.1002/pmic.201100556.

233. Baboo S, Bhushan B, Jiang H, Grovenor CRM, Pierre P, Davis BG, Cook PR. Most Human Proteins Made in Both Nucleus and Cytoplasm Turn Over within Minutes. Preiss T, editor. PLoS One. Public Library of Science; 2014; 9: e99346. doi: 10.1371/journal.pone.0099346.

234. Larance M, Ahmad Y, Kirkwood KJ, Ly T, Lamond AI. Global subcellular characterization of protein degradation using quantitative proteomics. Mol Cell Proteomics. American Society for Biochemistry and Molecular Biology; 2013; 12: 638–50. doi: 10.1074/mcp.M112.024547.

235. Christoforou A, Arias AM, Lilley KS. Determining Protein Subcellular Localization in Mammalian Cell Culture with Biochemical Fractionation and iTRAQ 8-Plex Quantification. Methods in molecular biology (Clifton, NJ). 2014. p. 157–74. doi: 10.1007/978-1-4939-0685-7_10.

236. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, et al. Initial sequencing and analysis of the human genome. Nature. Nature Publishing Group; 2001; 409: 860–921. doi: 10.1038/35057062.

237. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, et al. The Sequence of the Human Genome. Science (80-). 2001; 291.

238. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. Cold Spring Harbor Laboratory Press; 2002; 12: 996–1006. doi: 10.1101/gr.229102. Article published online before print in May 2002.

239. The future of cancer genomics. Nat Med. Nature Research; 2015; 21: 99–99. doi: 10.1038/nm.3801.

240. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Poznan, Poland). Termedia Publishing; 2015; 19: A68-77. doi: 10.5114/wo.2014.47136.

241. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, et al. A draft map of the human proteome. Nature. 2014; 509: 575–81. doi: 10.1038/nature13302.

242. Ai .-L Jpo@benevolent, Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, Overington JP. A comprehensive map of molecular drug targets. Nat Publ Gr. 2016; : 2016. doi: 10.1038/nrd.2016.230.

243. Workman P, Al-Lazikani B. Drugging cancer genomes. Nat Rev Drug Discov. 2013; 12: 889–90. doi: 10.1038/nrd4184.

244. Chowdhury F, Williams A, Johnson P. Validation and comparison of two multiplex technologies, Luminex® and Mesoscale Discovery, for human cytokine profiling. J Immunol Methods. 2009; 340: 55–64. doi: 10.1016/j.jim.2008.10.002.

245. Greenwood C, Ruff D, Kirvell S, Johnson G, Dhillon HS, Bustin SA. Proximity assays for sensitive quantification of proteins. Biomol Detect Quantif. 2015; 4: 10–6. doi: 10.1016/j.bdq.2015.04.002.

246. Harlan R, Zhang H. Targeted proteomics: a bridge between discovery and validation. Expert Rev Proteomics. 2014; 11: 657–61. doi: 10.1586/14789450.2014.976558.

247. Attwood P V., Piggott MJ, Zu XL, Besant PG. Focus on phosphohistidine. Amino Acids. 2007. p. 145–56. doi: 10.1007/s00726-006-0443-6.

248. Thavarajah R, Mudimbaimannar VK, Elizabeth J, Rao UK, Ranganathan K. Chemical and physical basics of routine formaldehyde fixation. J Oral Maxillofac Pathol. Wolters Kluwer -- Medknow Publications; 2012; 16: 400–5. doi: 10.4103/0973-029X.102496.