

From DEPARTMENT OF MEDICAL BIOCHEMISTRY AND
BIOPHYSICS

Karolinska Institutet, Stockholm, Sweden

MOLECULAR MECHANISMS AND TARGETS OF NEW ANTICANCER TREATMENTS

Amirata Saei Dibavar



**Karolinska
Institutet**

Stockholm 2019

Cover: learning from molecules on protein regulation, co-regulation and networks, and returning the favor by providing information on drug targets, mechanisms and similarities. Designed by Amirata Saei Dibavar and Nazli Saeedi. Drug structures are taken from DrugBank.

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by E-Print AB 2019

© Amirata Saei Dibavar, 2019

ISBN 978-91-7831-360-0

Molecular Mechanisms and Targets of New Anticancer
Treatments
THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Amirata Saei Dibavar

Principal Supervisor:

Prof. Roman Zubarev
Karolinska Institutet
Department of Medical Biochemistry and
Biophysics
Division of Physiological Chemistry I

Co-supervisor:

Prof. Thomas Helleday
Karolinska Institutet
Department of Oncology-Pathology

Opponent:

Dr. Karl Mechtler
IMP/IMBA/GMI/VBCF, Vienna Bio-Campus 1
Protein Chemistry Facility

Examination Board:

Prof. Johan Malmström
Lund University
Department of Clinical Sciences
Division of Infection Medicine

Prof. Rolf Lewensohn
Karolinska Institutet
Department of Oncology-Pathology

Prof. Jonas Bergquist
Uppsala University
Department of Chemistry

To my family and those with genuine interest in science

“I draw from the absurd three consequences, which are my revolt, my freedom, and my passion.”

The Myth of Sisyphus, Albert Camus

ABSTRACT

The work presented in this thesis is an effort to decipher and understand the mechanism of action (MOA) of anticancer agents by building on and complementing chemical proteomics methods. The backbone of the thesis relies on a recent method called Functional Identification of Target by Expression Proteomics (FITeXP) developed in Zubarev lab, where drug induced proteomic signatures are analyzed in various cell lines and top differentially regulated proteins with consistent behavior are determined, among which the drug target and mechanistic proteins are usually present. FITeXP relies on the assumption that proteins most affected with a perturbation have a higher probability of being involved in that process.

In this regard, **Paper I** aimed to enhance the performance of FITeXP analysis by merging proteomic data from drug-treated matrix attached and detached cells. This is while the majority if not all proteomics and molecular biology experiments are performed in matrix attached cells, as the general belief is that detached cells lose their structural integrity and do not harbor valuable information. However, detached cells are those that are more sensitive to chemotherapeutics and might reflect the proteome changes better. The comparative proteomics of living and dying cells improved FITeXP performance with regards to identification of targets and provided insight about proteins involved in cellular life and death decisions. Furthermore, the orthogonal partial least squares-discriminant analysis (OPLS-DA) paradigm presented in this study, was used throughout the thesis for contrasting and visualizing the proteomic signature of a molecule against others, to reveal targets and specific proteins changing in response to the molecule of interest.

In **Paper II**, as a further development of FITeXP and to demonstrate its applicability in a broader context, we built a proteome signature library of 56 clinical and experimental anticancer agents in A549 lung adenocarcinoma cell line. This resource called ProTargetMiner can be used for different purposes. The proximity of compounds in hierarchical clustering or t-SNE could be used for prediction of the mechanism of new compounds. Contrasting each molecule against other treatments using the OPLS-DA scheme presented in **Paper I**, revealed drug targets, mechanistic proteins, resistance factors, drug metabolizing enzymes and effects on protein complexes. Representative examples were used to demonstrate that the specificity factors extracted from the OPLS-DA models can help identify subtle but biologically significant processes, even when such an effect is as low as 15% fold change. Furthermore, we showed that the inclusion of 8-10 contrasting molecules in the OPLS-DA models can produce enough specificity for drug target deconvolution, which offered a miniaturization opportunity. Therefore, we built three deeper datasets using 9 compounds that showed the most diverse proteome changes in the orthogonal space in three cell lines from major cancer types: A549 lung, MCF-7 breast and RKO colon cancers. These datasets provide a unique depth of 7398, 8735 and 8551 respectively, with no missing values. Subsequently, a Shiny package was created in R, which can employ these datasets as a resource and merge it with user data and provide OPLS-DA output and target deconvolution opportunity for new compounds. Finally, using the original ProTargetMiner data, we also built a first of its kind proteomic correlation

database which can find applications in deciphering the function of uncharacterized proteins. Moreover, the resource helped to identify a set of core or untouchable proteins with stable expression across all the treatments, revealing essential functions within the cells. Such proteins could be used as house-keeping controls in molecular biology experiments.

In **paper III**, we combined FITeXP with other chemical proteomics tools Thermal Proteome Profiling (TPP) and multiplexed redox proteomics, to study the target and mechanism space of auranofin. This would also allow to assess the power, orthogonality and complementarity of these techniques in the realm of chemical proteomics. TPP is a recently developed technique that can monitor changes in the stability of proteins upon binding to small molecules. Redox proteomics is a method by which the oxidation level of protein cysteinome can be quantitatively analyzed. Auranofin is an FDA-approved anti-inflammatory drug for treatment of rheumatoid arthritis, but due to its potent antitumor activity, it is currently in clinical trials against cancer. Although several MOAs have been suggested for auranofin, uncertainties exist regarding its cellular targets; therefore, this molecule was chosen as a challenging candidate to test the chemical proteomics tools. A combination of the above mentioned tools confirmed thioredoxin reductase 1 (TXNRD1) (ranking 3rd) as the cognate target of auranofin and demonstrated that perturbation of oxidoreductase pathway is the main route of auranofin cytotoxicity. We next showed that changes in the redox state of specific cysteines can be linked to protein stability in TPP. Some of these cysteines were mapped to the active sites of redox-active enzymes.

In **Paper IV**, using quantitative multiplexed proteomics, we helped to show that b-AP15, a bis-benzylidene piperidone compound inhibiting deubiquitinases USP14 and UCHL5, produces a similar perturbation signature as bortezomib in colon cancer cells. However, in comparison with bortezomib, b-AP15 induces chaperone expression to a significantly higher level and leads to a more extensive accumulation of polyubiquitinated proteins. The polyubiquitinated proteins co-localize with mitochondrial membrane and subsequently reduce oxidative phosphorylation. These results help define the atypical cell death induced by b-AP15 and describe why this molecule is effective against apoptosis resistant cells in variety of tumor models.

Finally, in **Paper V**, we extended the applications of TPP and combined it with specificity concept for proteome-wide discovery of specific protein substrates for enzymes. We developed a universal method called System-wide Identification of Enzyme Substrates by Thermal Analysis (SIESTA) that relies on the hypothesis that enzymatic post-translational modification of substrate proteins can potentially change their stability against thermal denaturation. Furthermore, we applied the concept of specificity similar to the above papers, to reveal potential substrates using OPLS-DA. SIESTA was applied to two enzyme systems, namely TXNRD1 and poly-(ADP-ribose) polymerase-10 (PARP10), identifying known and putative candidate substrates. A number of these candidate proteins were validated as PARP10 substrates by targeted mass spectrometry, chemiluminescence and other assays. SIESTA is an unbiased and system wide approach and its broad application can improve our understanding of enzyme function in homeostasis and disease. In turn, specific protein substrates can serve as readouts in high throughput screening and facilitate drug discovery.

Taken together, in this thesis, FITExP methodology was improved in two directions. In **paper I**, we improved the performance of FITExP by combining the proteomics data from detached and attached cells. In **Paper II**, we demonstrated how the proteomics data on a multitude of drugs in a single cell line enables the discovery of compound targets and MOA. Furthermore, we built an R Shiny package which can serve as a resource for the cancer community in target and MOA deconvolution. In **Papers III and IV**, we applied an arsenal of chemical proteomics tools for characterization of two anticancer compounds. In **Paper V**, we expanded the applications of TPP to identification of specific protein substrates for enzymes in a system-wide manner.

LIST OF SCIENTIFIC PAPERS

- I. **Amir Ata Saei**, Pierre Sabatier, Ülkü Güler Tokat, Alexey Chernobrovkin, Mohammad Pirmoradian and Roman A. Zubarev. 2018. Comparative proteomics of dying and surviving cancer cells improves the identification of drug targets and sheds light on cell life/death decisions. *Molecular and Cellular Proteomics*. <https://doi.org/10.1074/mcp.RA118.000610>
- II. **Amir Ata Saei**, Alexey Chernobrovkin, Pierre Sabatier, Bo Zhang, Christian Beusch, Ülkü Güler Tokat, Massimiliano Gaetani, Ákos Végvári, Roman A. Zubarev. ProTargetMiner: A proteome signature library of anticancer molecules for functional discovery [manuscript]. (<https://www.biorxiv.org/content/early/2018/09/18/421115>)
- III. **Amir Ata Saei**, Hjalmar Gullberg, Pierre Sabatier, Christian Beusch, Alexey Chernobrovkin, Katarina Johansson, Bo Lundgren, Per I Arvidsson, Elias Arner, Roman A. Zubarev. Comprehensive chemical proteomics predicts drug targets: Auranofin as example [manuscript].
- IV. Xiaonan Zhang, Paola Pellegrini, **Amir Ata Saei**, Ellin-Kristina Hillert, Magdalena Mazurkiewicz, Maria Hägg Olofsson, Roman A Zubarev, Pádraig D'Arcy, Stig Linder. 2018. The deubiquitinase inhibitor b-AP15 induces strong proteotoxic stress and mitochondrial damage. *Biochemical Pharmacology*. 156: 291-301.
- V. **Amir Ata Saei**, Juan Astorga Wells, Pierre Sabatier, Christian Beusch, Alexey Chernobrovkin, Sergey Rodin, Katja Näreoja, Ann-Gerd Thorsell, Tobias Karlberg, Qing Cheng, Ákos Végvári, Elias S.J. Arnér, Herwig Schüler, Roman A. Zubarev. System-wide Identification of Enzyme Substrates by Thermal Analysis (SIESTA) [manuscript]. (<https://www.biorxiv.org/content/early/2018/09/20/423418>)

ADDITIONAL PUBLICATIONS

- VI. Pierre Sabatier, Amir Ata Saei, Shiyu Wang, Roman A Zubarev. 2018. Dynamic proteomics reveals high plasticity of cellular proteome: growth-related and drug-induced changes in cancer cells are comparable. *Proteomics*. <https://doi.org/10.1002/pmic.201800118>
- VII. Xiaonan Zhang, Karthik Selvaraju, Amir Ata Saei, Padraig D'Arcy, Roman Zubarev, Elias Arner, Stig Linder. Repurposing of auranofin: thioredoxin reductase remains the primary target [manuscript].
- √III. Ellin-Kristina Hillert, Slavica Brjnic, Xiaonan Zhang, Magdalena Mazurkiewicz, Amir Ata Saei, Arjan Mofers, Roman Zubarev, Stig Linder, Padraig D'Arcy. Proteasome inhibitor b-AP15 induces enhanced proteotoxicity by inhibiting cytoprotective aggresome formation [manuscript].
- IX. Massimiliano Gaetani, Pierre Sabatier, Amir Ata Saei, Christian Beusch, Zhe Yang, Susanna Lundstrom, Roman A. Zubarev. Proteome Integral Stability Alteration assay dramatically increases throughput and sensitivity in profiling factor-induced proteome changes [manuscript]. (<https://www.biorxiv.org/content/early/2018/12/13/496398>)
- X. Amir Ata Saei, Christian Beusch, Alexey Chernobrovkin, Pierre Sabatier, Roman A. Zubarev. Untouchable proteome uncovers core cellular functions and unravels house-keeping proteins [manuscript].
- XI. Amir Ata Saei, Bo Zhang, Christian Beusch, Pierre Sabatier, Alexey Chernobrovkin, Roman A. Zubarev. The first human protein correlation database uncovers unexpected complexity in protein regulation [manuscript].
- XII. Xiaonan Zhang, Belén Espinosa, Amir Ata Saei, Padraig D'Arcy, Roman A. Zubarev, and Stig Linder. Oxidative stress induced by the deubiquitinase inhibitor b-AP15 is associated with mitochondrial impairment [manuscript].

CONTENTS

1	INTRODUCTION	1
1.1	Of anticancer drugs and targets: why phenotypic screening?	1
1.2	Advances in mass-spectrometry based proteomics and its unique applications	2
1.3	Proteomics based drug target deconvolution strategies	4
1.3.1	Small-molecule affinity proteomics	5
1.3.2	Functional expression based-approaches	7
1.3.3	Discovery of covalent binders	8
1.3.4	Methods based on target stability	8
1.3.5	Computational techniques and connectivity maps.....	11
2	AIMS	15
2.1	Broad aims of the thesis	15
2.2	Specific aims.....	15
3	METHODS.....	17
3.1	Chemical proteomics paradigms.....	17
3.2	Label free sample preparation.....	17
3.3	TMT10 multiplexing	18
3.4	Sequential iodoTMT labeling for redox proteomics	19
3.5	Proteomics	19
3.6	LC-MS/MS	20
3.7	Quantification	20
3.8	Data analysis.....	21
4	RESULTS AND DISCUSSION.....	22
4.1	Paper I. Comparative proteomics of dying and surviving cancer cells improves the identification of drug targets and sheds light on cell life/death decisions	22
4.2	Paper II. ProTargetMiner: A proteome signature library of anticancer molecules for functional discovery.....	28
4.2.1	Construction of ProTargetMiner	28
4.2.2	Proteomic dissection of cell death trajectories	30
4.2.3	Similar compounds produce similar proteome signatures.....	30
4.2.4	Functional discovery in a compendium of proteome signatures	33
4.2.5	The degree of drug-induced proteome changes	34
4.2.6	Assessing the miniaturization possibility of ProTargetMiner concept.....	36
4.2.7	Miniaturization of ProTargetMiner	37
4.2.8	Merging deep datasets to obtain a general picture of drug targets and MOA.....	39
4.2.9	ProTargetMiner R Shiny package as a public tool	41
4.2.10	The first human protein correlation database uncovers unexpected complexity in protein regulation.....	43

4.2.11	The database pulls out dense regions of protein interaction networks	44
4.2.12	Protein anti-correlation, true or false?	46
4.2.13	Adding complexity to protein regulation	48
4.2.14	Anti-correlation provides further information on protein regulation	49
4.2.15	Untouchable proteome reflects essential cell functions	49
4.2.16	Untouchable proteome unravels house-keeping proteins	50
4.3	Paper III. Comprehensive chemical proteomics predicts drug targets: Auranofin as example.....	52
4.3.1	Unbiased prediction of auranofin targets using a combination of chemical proteomics tools.....	52
4.3.2	Pathway analysis reveals the major auranofin mechanism.....	54
4.3.3	Redox proteomics links cysteine oxidation and protein stability	55
4.4	Paper IV. The deubiquitinase inhibitor b-AP15 induces strong proteotoxic stress and mitochondrial damage	57
4.5	Paper V. System-wide identification of enzyme substrates by thermal analysis (SIESTA)	58
4.5.1	SIESTA identifies known and putative TXNRD1 substrates.....	59
4.5.2	SIESTA identifies many novel putative substrates for PARP10.....	61
5	CONCLUSIONS AND FUTURE PERSPECTIVES	64
6	Acknowledgements	65
7	References	69

LIST OF ABBREVIATIONS

CETSA	CELLular Thermal Shift Assay
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
FDR	false discovery rate
FITExP	Functional Identification of Target by Expression Proteomics
GSH	glutathione
HKP	housekeeping proteins
IAA	iodoacetamide
LC-MS/MS	liquid chromatography combined with tandem mass spectrometry
LysC	endoproteinase Lys-C
m/z	mass to charge ratio
MOA	mechanism of action
MS	mass spectrometry
MS/MS	tandem mass spectrometry
NF- κ B	nuclear factor kappa-light-chain-enhancer of activated B cells
OPLS-DA	Orthogonal Partial Least Squares-Discriminant Analysis
PCA	Principal Component Analysis
PISA	Proteome Integral Stability Alteration (assay)
PLS	Partial Least Squares
PTM	post-translational modification
SDS	sodium dodecyl sulfate
SIESTA	System-wide Identification of Enzyme Substrates by Thermal Analysis
T _m	melting point
TPP	Thermal Proteome Profiling

1 INTRODUCTION

1.1 OF ANTICANCER DRUGS AND TARGETS: WHY PHENOTYPIC SCREENING?

Despite advances in cancer diagnosis and therapy, International Agency for Research on Cancer projected 18.1 million new cases of cancer and 9.6 million deaths from cancer in 2018 worldwide (Bray, Ferlay et al. 2018). Furthermore, although the total cancer death rate has dropped by 26% since 1991 till 2015, some types of cancers such as liver and brain have increased death rates (Siegel, Miller et al. 2018). With significant reduction in heart disease, according to WHO estimates in year 2015, in 91 of 172 countries studied, cancer has taken the first or second ranking as the leading cause of death and in an additional 22 countries, ranks third or fourth.

Although some cancer types such as acute lymphoblastic leukemia (ALL) in children can be mostly treated (around 98% remission) (Vora, Goulden et al. 2013), some cancer types are irresponsive or resistant to the available treatments or are prone to relapse. Many if not most existing therapies only extend the patient lifespan by a limited time (Weigmann 2016). Taken together, drugs and methodologies aiming at cancer treatment and eradication are highly desired and sought for.

For discovery and development of novel anticancer drugs, several paradigms exist, which mainly include targeted high throughput screening (Broach and Thorner 1996) and phenotyping screening (Moffat, Rudolph et al. 2014). On the contrary to targeted high throughput screening where a massive number of compounds are surveyed for binding to a given protein, in phenotypic screening, compounds inducing a certain biological effect are identified through cell-based assays. The other major advantage with phenotypic screening is that the chosen compounds are already bioactive, and have drug-like properties such as the potential to penetrate cell membrane (Moffat, Rudolph et al. 2014). Furthermore, the compounds are given a chance to induce the phenotype of interest by acting on multiple targets (Medina-Franco, Giulianotti et al. 2013), and therefore, they can in return, help identify novel drug targets, as “Determining the causal relationships between target inhibition and phenotypic effects may well open up new and unexpected avenues of cancer biology” (Moffat, Rudolph et al. 2014). The merits and disadvantages of each approach is summarized in **Figure 1**. Since these cell based screens are disease relevant and cover multiple targets in a single screening, they are believed to have higher success rates than the targeted approach. Statistically, during the period between years 1999-2008, compounds derived from phenotypic screening comprised about 37% of the first-in-class drugs approved by FDA, while targeted screening contributed 23%. However, targeted screening had a better success rate in follower drugs comprising 51% of approved drugs (vs. 18% from phenotypic screening) (Swinney and Anthony 2011; Lee, Uhlik et al. 2012). The higher success rate for the follower drugs is not surprising, since recognizing a feasible protein target facilitates the screening and drug development process. Although a recent study has noted a substantial increase in the approval of first-in-class drugs discovered by targeted screening covering the period of 2008-2013 (Eder,

Sedrani et al. 2014), phenotypic screening is still an essential and viable source of novel key compounds.

Sophisticated and parallel analytical methods are usually required for deconvolution of the targets for molecules discovered in phenotypic screening and off-targets of those from high-throughput screening. Target and MOA deconvolution can be a time- and labor intensive process, if not even fruitless. These challenges have left many promising experimental compounds

without a known MOA. Even among the FDA approved drugs, 7% have no known primary target, and around 18% lack a clear MOA (Gregori-Puigjané, Setola et al. 2012). In addition, some MOAs might have been wrongly associated with a number of molecules in the past (Somlyai, Collins et al. 2017) and off-targets are being identified on a routine basis (Klaeger, Gohlke et al. 2016; Van Esbroeck, Janssen et al. 2017). Thus, there is a huge unmet need for developing new methods for identification of the target and deconvolution of MOA in anticancer drug discovery and development (Schirle and Jenkins 2016). The majority of compounds act by binding and modulating the activity of target proteins; therefore, mass-spectrometry based proteomics has become an indispensable tool in drug development (Schirle, Bantscheff et al. 2012).

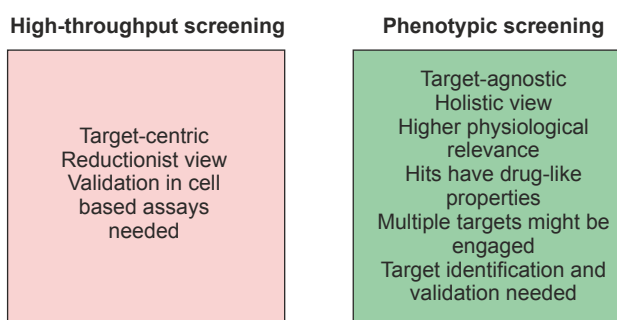


Figure 1. The attributes of high-throughput vs. phenotypic screening

1.2 ADVANCES IN MASS-SPECTROMETRY BASED PROTEOMICS AND ITS UNIQUE APPLICATIONS

Advances in mass spectrometry in the last decades have turned it into an unprecedented means to investigate the proteome (Aebersold and Mann 2003). Proteomics is now routinely used to study variations in protein abundances, modifications, stability and etc. between different conditions. The realm of applications will be vastly expanded even further with the advent of single cell proteomics (Budnik, Levy et al. 2018).

Milestone studies have extended the proteome coverage captured by mass spectrometry. Over a three-year period from 2011, the number of detected proteins in a single LC-MS analysis improved from 3000 yeast proteins in 8 hours (Thakur, Geiger et al. 2011) to 4000 proteins in 1.1 hours (Hebert, Richards et al. 2014). That significant improvement was mainly due to the introduction of new MS platforms with enhanced MS/MS acquisition rate as well as a higher resolution and mass accuracy. However, the advancements in sample preparation techniques, chromatographic separation, quantification algorithms and data processing also played a part. Currently, single-shot proteomics experiment can routinely sample half of the expressed mammalian cell proteome, i.e. exceeding 5,000 proteins (Pirmoradian, Budamgunta et al.

2013; Scheltema, Hauschild et al. 2014). Very recently, it has become possible to quantify 10,000 proteins in 100 minutes, although special set-ups and specific high resolution mass spectrometers are required (Meier, Geyer et al. 2018).

In parallel, various labeling technologies have made it possible to multiplex samples in a single analysis. Not only has this made direct comparison between the proteomes possible within the same LC-MS run, but it has also reduced the variation between experimental replicates. Labeling also provides an opportunity to fractionate the pooled samples to obtain higher proteome coverage. Orthogonal fractionation of a sample reduces the sample complexity, which allows one to take deeper snapshots of the proteome, reaching more than 10,000 proteins (Sabatier, Saei et al.) (paper not included in this thesis). Extensively used fractionation techniques include high pH reversed-phase chromatography (Batth, Francavilla et al. 2014) and isoelectric focusing (Branca, Orre et al. 2014; Pirmoradian, Astorga-Wells et al. 2015).

As a result of these advances, proteomics is now routinely used in biomarker discovery (Zhan, Li et al. 2018), protein network analysis (Bennett, Rush et al. 2010; Tan, Go et al. 2018), cell type comparison (Gholami, Hahne et al. 2013), analysis of protein stability (Leuenberger, Ganscha et al. 2017; Becher, Andres-Pons et al. 2018; Dai, Zhao et al. 2018) and turnover (Savitski, Zinn et al. 2018), as well as investigation of post-translational modifications (PTMs) (Christophorou, Castelo-Branco et al. 2014; Weinert, Narita et al. 2018), proteoforms (Ntai, Fornelli et al. 2018) and proteogenomics (Zhu, Orre et al. 2018). Arguably, one of the most important applications of mass spectrometry is drug target and MOA deconvolution, in line with the chemical proteomics paradigm (Savitski, Reinhard et al. 2014; Browne, Jiang et al. 2018).

A routine question from biologists is: “what can proteomics do in comparison with other high-throughput technologies e.g. RNA sequencing?” There are inherent differences in the type of information obtained from proteomics compared to RNA sequencing. First and foremost, proteomics is the only system-wide tool that provides information on both protein production and degradation. Such information is not reflected in genomics or transcriptomics data. Several studies have shown that the global mRNA levels are not strongly correlated with protein abundances even under steady state (Edfors, Danielsson et al. 2016; Liu, Beyer et al. 2016).

Furthermore, some particular proteomic applications exist that are by default not amenable to genomics and transcriptomics. For example, proteomics can be used to observe and quantify myriad of PTMs on proteins (Savitski, Nielsen et al. 2006; Chick, Kolippakkam et al. 2015). Metabolic labeling can be applied to quantify protein degradation, translation and overall turnover (Boisvert, Ahmad et al. 2012). Redox proteomics, on the other hand, utilizes cysteine labeling to quantify the oxidation state of the proteins (Checconi, Salzano et al. 2015). Hydrogen-deuterium exchange mass spectrometry (HDX-MS) can be employed to study the binding site of various molecules on proteins or to validate ligand-protein interactions (Visnes, Cázares-Körner et al. 2018).

A major recent development was combining Cellular Thermal Shift Assay (CETSA) (Molina, Jafari et al. 2013) with multiplexed proteomics to establish the TPP technique which is capable of proteome-wide monitoring of protein stability in response to different conditions or small molecule treatments in living cells (Savitski, Reinhard et al. 2014). Protein stability is thus a new dimension in proteomics, inaccessible to genomics and transcriptomics, which sparked new interest in studying protein thermal unfolding and structural parameters dictating thermal stability in different organisms (Leuenberger, Ganscha et al. 2017).

Overall, such parameters as protein presence and abundance, sequence, stability, mutations (proteogenomics), PTM type and occupancy, protein degradation/turnover rate, oxidized-reduced states, interactions, binding sites and even isotopic composition can be investigated by mass spectrometry.

1.3 PROTEOMICS BASED DRUG TARGET DECONVOLUTION STRATEGIES

Chemical proteomics can be used to study the relationship between the drug molecules and cellular phenotype (Kwon and Karuso 2018). Classically, chemical proteomics refers to mass spectrometry-based affinity chromatography approaches to identify the protein interactors of small molecules or probes in the system-wide scale (Rix and Superti-Furga 2009). Recently, chemical proteomics is extended to other techniques, where changes in the proteome state – such as stability, expression, oxidation state and covalent modification, can be connected to the targets or functions of small molecules (Savitski, Reinhard et al. 2014; Chernobrovkin, Marin-Vicente et al. 2015; Browne, Jiang et al. 2018; Piazza, Kochanowski et al. 2018). Although identification and characterization of the efficacy target, the protein through which the compound exerts the desired phenotype, is very challenging and at times fruitless (Kotz 2012), it is indeed of paramount importance. Characterization of target and MOA can help simplify the prediction of efficacy and side effects and later on contribute to optimization of the lead compound scaffold. In parallel, for already approved drugs with poorly known targets, drug target studies can facilitate the design of next generation compounds with reduced side effects and perhaps higher efficacy. Target characterization can also reduce the chance of failure during clinical trials, as many drug candidates fail due to lack of efficacy in humans (Bunnage, Gilbert et al. 2015).

Taken together, unbiased techniques are required for uncovering drug target and MOA, especially those that exert a phenotype of interest. Several technologies exist for characterizing the direct and indirect target space of a given anticancer compound. Individual methodologies harbor different resolutions, strengths and weaknesses (Schirle and Jenkins 2016). Since these methods usually provide complementary information, it is imperative to use them in concert (Schenone, Dančik et al. 2013). In this section, we will only focus on drug target deconvolution strategies based on mass spectrometry.

1.3.1 Small-molecule affinity proteomics

Affinity based approaches were first introduced in the early 1950s and aimed to identify protein-binding spectrum of a molecule (Campbell, Luescher et al. 1951). Today, chemical proteomics encompasses mass spectrometry-based affinity chromatography approaches for identification of ligand-protein interactions in the whole proteome background. The first step in affinity purification for target identification is the modification of compound of interest with a linker and its subsequent immobilization on a solid support to produce a compound display. Subsequently, the immobilized drug is exposed to the protein of interest or the disease relevant cell lysate. After several wash steps to remove the non-binding proteins, the binders are digested and subjected to LC-MS analysis and the enriched interactors are identified. In this paradigm, the advantage of using a cell lysate over purified protein is that the reaction will take place with the specific proteoforms with natural sequences, PTMs and potential protein interactors which might be important in compound binding. A drawback of this technique is that direct and indirect targets cannot be differentiated, as accessory proteins interacting with the real target might also be enriched. Another downside of the lysate-based non-covalent affinity pulldown methods for identification of drug targets is the incompatibility of integral membrane proteins, as such proteins might lose their native binding conformation through the experimental workflow (Basilana, Carlson et al. 2014; Akbulut, Gaunt et al. 2015).

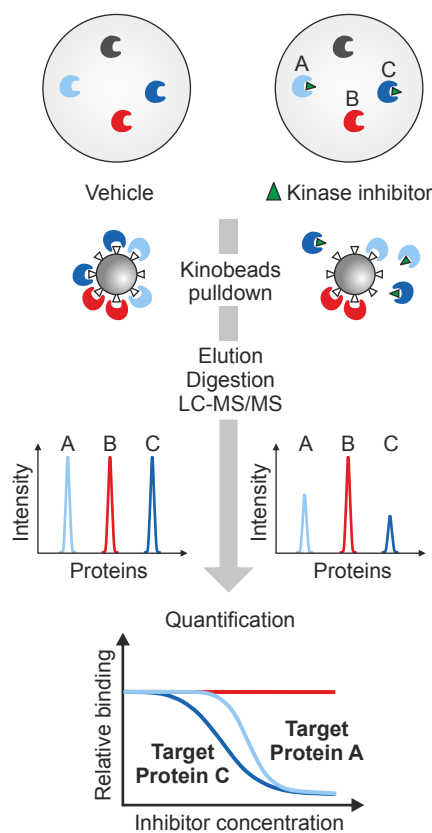
Since the linker should be attached on a permissive site to retain compound activity, the synthesis of these tool compounds can be challenging. For some compounds, it might even be impossible to introduce such a linker. To ensure that the compound is still specific after the modification, its activity must be investigated in comparison with the parent compound. Alternatively, photo-crosslinker matrices can be irradiated together with the parent compound to produce a random compound display, which can presumably include orientations compatible with target protein binding (Kano, Honda et al. 2005). Furthermore, assessing active and inactive probes sharing the same scaffold can help to identify the most phenotypically-relevant interactors (Oda, Owa et al. 2003). To rule out false interactions, competitive-binding experiment can be performed, in which the free probe competes with immobilized probe for binding to the target proteins. Affinity purification has been used to identify the complexes targeted by histone deacetylase inhibitors (Bantscheff, Hopf et al. 2011). Such a methodology has also been used to characterize oxysterol binding proteins as the cellular carriers of OSW-1 -a highly potent natural anticancer lead compound, and to show that these proteins are essential for cell survival (Burgett, Poulsen et al. 2011). A miniaturized version of these affinity pulldowns has been optimized for profiling of the targets of clinical kinase inhibitors in tumor biopsies (Chamrád, Rix et al. 2013).

On the contrary to non-covalent methods, covalent approaches can freeze the interaction of an affinity probe with its target. In this approach, a reactive group and an affinity handle (e.g. biotin) are conjugated to the compound. This strategy has been used for validation of Sec61a as a target of cyclodepsipeptide (MacKinnon, Garrison et al. 2007). Covalent approaches also allow for using denaturing conditions, since the preservation of protein binding conformation is not necessary. As the interactors are identified through proximity-driven labeling events, this

methodology can be used to eliminate indirect accessory proteins enriched with the target proteins in complexes. For example, while a non-covalent suberanilohydroxamic acid (SAHA, a class I/IIb HDAC active-site inhibitor) probe gave several HDAC-containing complexes as specific targets (Bantscheff, Hopf et al. 2011), the covalent SAHA–BPyne probe was more specific in finding core HDACs as direct binders (Salisbury and Cravatt 2007). The covalent approaches sometimes suffer from low affinity of interactions and low efficiency of labeling, which can be ameliorated by performing further assays such as competition with free compound or probe titration, to enhance the specificity of detected interactions. A challenge in the latter technique is its uncertain applicability to live cells and membrane fractions.

Browne et al. recently developed a method called CITE-Id (Covalent Inhibitor Target site IDentification) for proteome-wide target-site identification of covalent inhibitors (Browne, Jiang et al. 2018). This chemoproteomic approach employs covalent inhibitors for enriching cysteine-thiols across the proteome, and therefore provides direct, amino acid level readout in a dose-response manner. Using this method, they discovered that PKN3 is covalently inhibited by JZ128.

Kinobead technology for target profiling might also be classified as an affinity-based technique, but is based on a different methodology. In this method (**Figure 2**), broad spectrum kinase inhibitors or ATP mimetics are immobilized on beads and incubated with the lysate of interest. The addition of a desired kinase inhibitor (new compound) to the experiment, will compete with beads for binding to the present kinases. After capturing the tissue kinome on matrix, the enriched proteins are identified by proteomics (Bantscheff, Eberhard et al. 2007). The kinobead methodology was recently applied to 243 kinase inhibitors (Klaeger, Heinzlmeir et al. 2017). One of the limitations of this technology is that it currently does not apply to the full human kinome. Furthermore, the potential non-kinase targets of the compound are ignored. Moreover, since *a priori* knowledge about the compound – being a kinase inhibitor is assumed, kinobead technology cannot be categorized as a target deconvolution strategy.



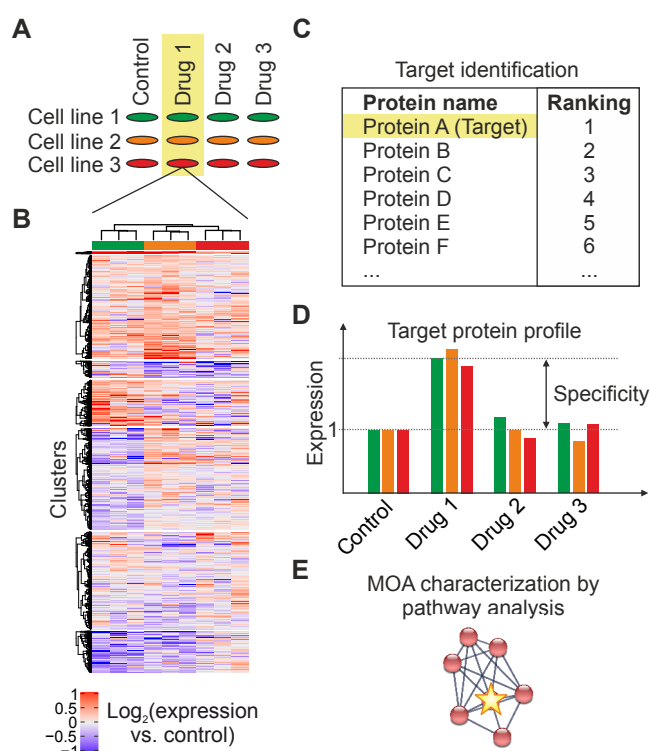
After capturing the tissue kinome on matrix, the enriched proteins are identified by proteomics (Bantscheff, Eberhard et al. 2007). The kinobead methodology was recently applied to 243 kinase inhibitors (Klaeger, Heinzlmeir et al. 2017). One of the limitations of this technology is that it currently does not apply to the full human kinome. Furthermore, the potential non-kinase targets of the compound are ignored. Moreover, since *a priori* knowledge about the compound – being a kinase inhibitor is assumed, kinobead technology cannot be categorized as a target deconvolution strategy.

Figure 2. Kinobead technology. Broad spectrum kinase inhibitors or ATP mimetics are immobilized on beads and incubated with the lysate. The added kinase inhibitor will compete with kinase binding to beads. After capturing the tissue kinome on matrix, the enriched proteins are identified by proteomics. A reduction of kinase binding to beads is a measure of compound binding to the respective kinase. Redrawn from (Klaeger, Heinzlmeir et al. 2017), with permission from The American Association for the Advancement of Science.

1.3.2 Functional expression based-approaches

Chernobrovkin et al. developed a method in 2015, called Functional Identification of Target by Expression Proteomics (FITeXP) (Chernobrovkin, Marin-Vicente et al. 2015). FITeXP is an expression proteomics-based strategy for deconvolution of potential drug targets and MOA. Unlike other methods, expression proteomics takes protein degradation into account, and it can prove highly useful for investigating drug target and MOA. In FITeXP, at least three cell lines are treated with the test compound at LC50 concentrations for about two doubling times and subsequently, differentially expressed proteins with consistent behavior in all tested cell lines are deciphered (**Figure 3**). In brief, all proteins are sorted by their regulation upon treatment in different cell lines and the rankings are combined. To increase the specificity, and to filter out the generic and/or stochastically behaving proteins involved in cell death or detoxification, a few known drugs are added to the compound panel for contrast. A collateral advantage of FITeXP is that co-regulated proteins with biggest abundance changes can be subjected to pathway analysis to characterize the compound MOA. Another advantage of FITeXP is that chemical modification of the parent molecule is not necessary and no *a priori* knowledge about the drug MOA is required. Furthermore, FITeXP allows for rank ordering of the putative targets. In the proof of principle experiments, FITeXP could identify the target of several known small anticancer molecules among more than 4000 proteins (Chernobrovkin, Marin-Vicente et al. 2015). Besides the analysis of the target and MOA of chemotherapeutics (Chernobrovkin, Marin-Vicente et al. 2015), FITeXP has been also used to deconvolute the target and MOA of metallodrugs to help finding optimal drug combination (Lee, Chernobrovkin et al. 2017). The approach was also applied to nanoparticles to investigate their cytotoxic and pro-inflammatory effects (Tarasova, Gallud et al. 2016).

Figure 3. FITeXP methodology. **A)** At least three cell lines are treated at LC50 concentrations with a compound panel including the molecule of interest and **B)** the proteomes are analyzed. **C)** Proteins are ranked based on their fold change in different cell lines, culminating in a combined ranking. The top proteins are potential target candidates. **D)** The target protein profile for compound of interest (Drug 1 here) usually has a higher regulation in majority of cell lines. **E)** Finally, a list of top ranking proteins can be subjected to pathways analysis in StringDB or similar databases to reveal compound MOA.



The underlying principle in FITeXP is based on the exceptional abundance change of the drug target and compound-specific mechanistic proteins upon treatment. Various phenomena can underlie the exceptional up-regulation (or in rare cases down-regulation) of the target protein in response to drug binding. First and perhaps the most important, are the feedback mechanisms

which induce the up-regulation of proteins with suppressed activity (Legewie, Herzel et al. 2008). Second, drug binding can alter the stability of protein and accelerate or decelerate its turnover (Cohen, Geva-Zatorsky et al. 2008). Such feedback loops are not limited to higher organisms; for example, inhibition of the cholesterol pathway in yeast transcriptionally induces the expression of several genes in this pathway (Daum, Lees et al. 1998). Such level of protein expression control have also been observed in more primitive organisms (Smits, Kuipers et al. 2006; Palmer and Kishony 2014).

1.3.3 Discovery of covalent binders

Proteomics might also be used for direct discovery of covalent protein-compound adducts. Such a strategy has been used to study covalent modifications of β -lactoglobulin A and human serum albumin by reactive metabolites from paracetamol, amodiaquine, and clozapine (Lohmann, Hayen et al. 2008). Furthermore, mass spectrometry has been used in a high-throughput format to identify protein-inhibitor adducts for discovery of irreversible inhibitors (Campuzano, San Miguel et al. 2016).

1.3.4 Methods based on target stability

Stability is a new dimension in proteomics. A number of methods exist for monitoring protein stability changes upon ligand binding. An important advantage of these techniques, similar to expression proteomics based strategies such as FITeXP, is that the modification of the parent compound is not required.

One of the early stability-based methods called Pulse Proteolysis is based on the fact that urea-induced unfolding enhances protein susceptibility to proteolysis. This feature can be exploited to monitor the level of protein unfolding in the presence and absence of a ligand, which can be detected by SDS-PAGE and mass spectrometry. In the proof-of-principle experiments, the authors showed the applicability of this method to confirm the binding of maltose and other cognate ligands to maltose-binding protein (Park and Marqusee 2005).

The first technique that could be utilized for drug target identification in the whole proteome was drug affinity responsive target stability (DARTS) (Lomenick, Hao et al. 2009). The underlying concept in DARTS is that binding of a ligand can locally or globally stabilize the target protein, which can also be accompanied by covering protease recognition sites, and diminish the proteolysis of the target protein. This is highly similar to, for example, the resistance of transcription factor-bound DNA sites to DNAases (Maniatis and Ptashne 1973). In proof of principle studies, DARTS could confirm the binding of several molecules to their target proteins in the cell lysate background and establish proteins in translation machinery as targets for resveratrol. However, the applicability of DARTS for monitoring dynamic interactions in intact cells is yet to be proven.

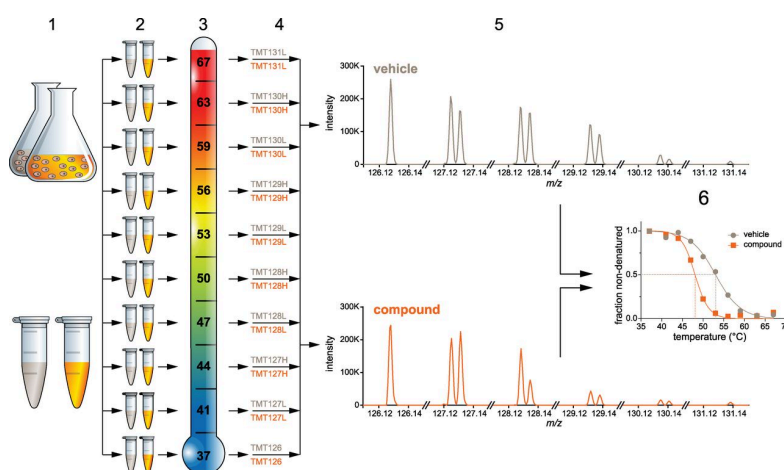
A highly similar technique called Limited Proteolysis (LiP-MS) has been used to investigate the effects of metabolite binding on protein structure in *Escherichia coli* (Piazza, Kochanowski et al. 2018). This method relies on limited proteolytic digestion of proteins in the bound and unbound state and does not require the modification of small molecules. Although this method can only detect the conformational change of a fraction of the proteome and has never been applied to deconvolution of targets for drug molecules, it is envisioned that with further advances it can potentially find applications in such studies.

Cellular Thermal Shift Assay (CETSA) was recently developed to monitor ligand binding to potential target proteins in living cells and lysate (Molina, Jafari et al. 2013). This methodology is using the well-established fact of the thermal stabilization (or occasional destabilization) of a target protein upon ligand binding, which results in a shift in the melting curve of the protein (Pantoliano, Petrella et al. 2001). CETSA experiments are performed in a Western blot format, in which antibodies are used to individually detect soluble protein content after heating cells to different temperatures and removing the aggregated proteins by centrifugation (Molina, Jafari et al. 2013). The advantage of CETSA over other target engagement methods is that it can be performed in living cells and tissues and as such, confirms the drug target engagement in a physiological context. Such aspect can be particularly important for drug metabolites or prodrugs that are activated after cellular entry. However, *a priori* knowledge about the target is needed and thus CETSA cannot be classified as a target deconvolution technique, but CETSA can still be used for screening. For example, CETSA screening was used to identify novel thymidylate synthase (TYMS) inhibitors from a compound library (Almqvist, Axelsson et al. 2016).

More recently, Savitski et al. (Savitski, Reinhard et al. 2014) have coupled CETSA with quantitative multiplexed proteomics to study potential drug binding in the whole proteome, in a method called TPP. Since TPP measures the concentration of thousands of proteins simultaneously, it can also be used for drug target deconvolution. Different modes of TPP can be used to detect targets, off-targets and proteins which are affected down-stream of the target in the cells. In the TR-TPP (temperature range) experiments (**Figure 4**), TPP is performed in a fixed biologically relevant concentration of the compound over a temperature range. The readout is the melting point (T_m) of proteins and the shift in melting curves of a potential target protein between treatment and control (ΔT_m). In most cases, TPP can identify the direct and indirect targets of a drug in live cells. In addition, TR-TPP experiments in cell lysate can help differentiate the proteins which are directly binding the compound, as the majority of downstream pathways are absent in the lysate. Furthermore, lysate experiments might also help to differentiate the compounds which engage their targets after metabolic transformation. Alternatively, CCR-TPP (compound concentration range) experiments, are performed at a fixed temperature (T_m of the target proteins) over a compound concentration range (Franken, Mathieson et al. 2015). The output of CCR-TPP is usually used to confirm the TR-TPP data, and the approach touches upon compound potency when comparing a panel of molecules within the same experiment (Franken, Mathieson et al. 2015).

Figure 4. TPP methodology.

Living cells or lysate aliquots are heated step-wise within 37-67 °C, in the presence and absence of compound. Ultracentrifugation of the samples after heating leaves only the soluble non-aggregated fractions, which are multiplexed by TMT10 and quantitatively analyzed by mass spectrometry. The change in stability of a protein can be detected by tracking ΔT_m . Reused from (Savitski, Reinhard et al. 2014), with permission from The American Association for the Advancement of Science.



A shortcoming of CETSA and TPP is that they are only applied to a) proteins soluble at normal conditions or non-denaturing buffers (this excludes, e.g., the majority of membrane proteins – although this issue can be partially ameliorated with mild detergents (Reinhard, Eberhard et al. 2015)), and b) proteins whose stability decreases with temperature. Also, the magnitude of shift in thermal stability is not predictable and is largely specific to the target protein and the binding event. Not all bindings lead to detectable changes in protein stability. Overall, these issues make the rank ordering of putative targets difficult.

Very recently, a similar methodology, based on the same principles of thermal stability, has identified the target for small molecule compounds by in-gel fluorescence. The authors were able to identify and validate nucleophosmin as the target of hordenine, which is a natural compound known to up-regulate *in vitro* translation (Park, Ha et al. 2017).

Despite the fact that TPP provides valuable information on protein state in a system wide level, low throughput and cost are major drawbacks of this technique, limiting its applications to labs specializing in mass spectrometry or core facilities. Our group recently developed a technique called Proteome Integral Stability Alteration (PISA) assay with 10x higher throughput, where TPP is reformatted by concatenation of the samples over the temperature range (Gaetani, Sabatier et al. 2018) (paper not included in this thesis). PISA circumvents the need to perform fitting of melting curves. The simplicity, lower expense and significantly higher throughput of PISA guarantee its widespread use in chemical biology and drug development. Furthermore, since samples are combined, PISA can be easily performed in multiple replicates (unlike TPP which is mostly performed in 2 replicates), allowing for more rigorous statistics. The rationale, schematics and workflow of PISA are shown in **Figure 5**.

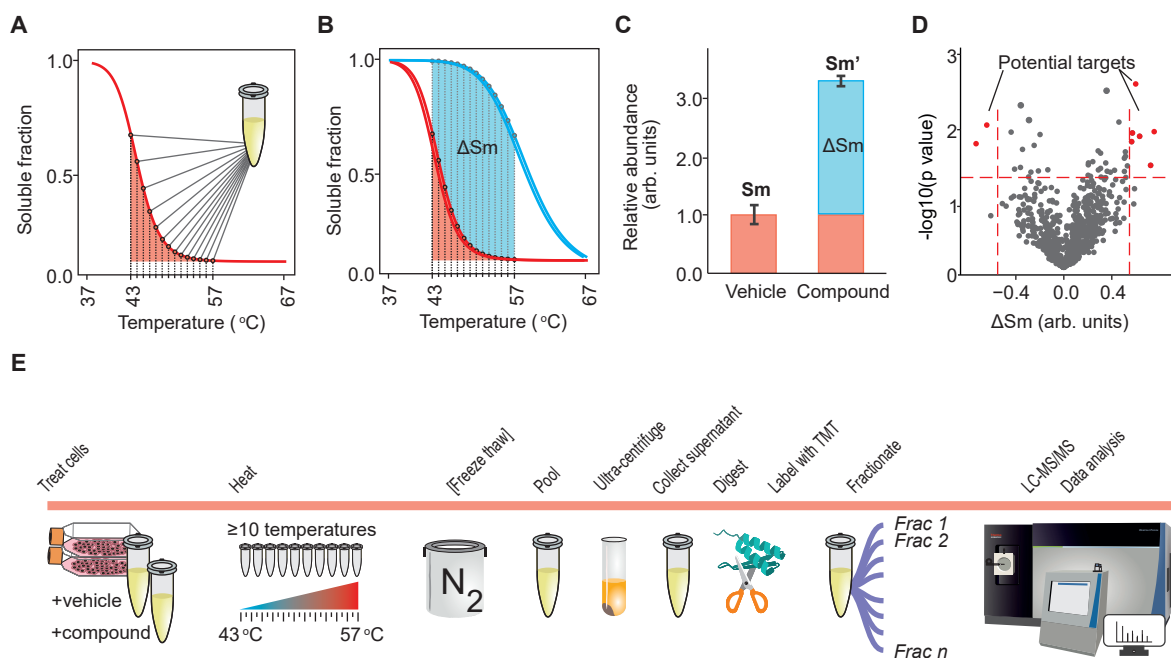


Figure 5. Schematics of PISA. **A)** In PISA, individual samples heated to different temperature points are pooled, circumventing the need for curve fitting. **B** and **C)** ΔSm can be easily calculated as the difference of the integral abundances of the protein between treated and untreated samples. **D)** Presentation of PISA results in a simple volcano plot allowing for a test of significance and highlighting potential targets of the compound. **E)** PISA workflow in living cells: incubation with compound, aliquoting, heating, freeze-thaw for lysing cells, pooling of samples, ultracentrifugation, collection of supernatant, digestion, TMT labeling, fractionation, high resolution LC-MS/MS analysis, and subsequent data processing. Figure adapted with permission from (Gaetani, Sabatier et al. 2018).

1.3.5 Computational techniques and connectivity maps

There have been several attempts to create computational *in silico* target identification methods, based on chemical similarity database searches and/or monitoring bioactivity fingerprint similarity. For example, Chemical Similarity Network Analysis Pulldown (CSNAP) is a computational drug target identification algorithm that exploits chemical similarity for recognition of chemotypes and identification of drug targets (Lo, Senese et al. 2015). Using this method, the authors were able to identify novel compounds targeting microtubules.

Reference genomic, transcriptomic, metabolomic and proteomic profiles can be potentially employed to discover functional connections among diseases, perturbations and drug action. Although proteomics databases are yet to be built, connectivity maps have already been created for gene expression signatures to connect small molecules sharing a MOA (Hughes, Marton et al. 2000; Lamb, Crawford et al. 2006; Rees, Seashore-Ludlow et al. 2016; Subramanian, Narayan et al. 2017; Readhead, Hartley et al. 2018; Ye, Ho et al. 2018). Such databases, in ideal case, must be able to cover all biological states that can be induced by chemical entities, in terms of genomic and/or proteomics profiles. Pattern-matching tools can be subsequently used to detect similarities among compound signatures. The assigned similarity scores can be used for rank ordering of the fingerprints. In a landmark study, such a gene expression database was built for 300 diverse mutations and chemical treatments in *S. cerevisiae* (Hughes, Marton

et al. 2000). The authors showed that the affected cellular pathways can be determined by pattern matching. Investigation of the gene expression profiles caused by deletion of uncharacterized genes led to functional annotation of eight previously uncharacterized open reading frames. Moreover, the authors could identify the dyclonine target.

A similar attempt has led to the generation of commercial databases embedded with expression profiles from rat tissues upon systemic drug administration (Ganter, Tugendreich et al. 2005). This database, in 2005, had gene expression profiles for approximately 600 different compounds in 7 rat tissues (approximately 3200 different drug-dose-time-tissue combinations). In this study, combining the clinical pathology assessments with gene expression profiles could explain why carmustine, methotrexate, and thioguanine had similar effects on the hematopoietic tissues, while they induced diverse hepatotoxicity.

Lamb et al. created the first connectivity map database in human cells (Lamb, Crawford et al. 2006). They created a gene expression profile database of 164 distinct small-molecule perturbagens, representing diverse functions, including FDA-approved drugs and non-drug bioactive tool compounds. The database was used to show that drugs with similar MOA will have similar genomic fingerprints (**Figure 6**). They reported that drugs with similar MOAs (estrogen receptor modulators and genistein) showed positive connectivities, while an antagonist (fulvestrant) demonstrated negative connectivity. The dataset could also be used as a discovery tool, for example by identification of gedunin as an HSP90 inhibitor. Ever since, connectivity map has been successfully used for deconvolution of compound MOA (D'arcy, Brnjic et al. 2011), lead discovery (Hieronymus, Lamb et al. 2006) and drug repurposing (Sirota, Dudley et al. 2011).

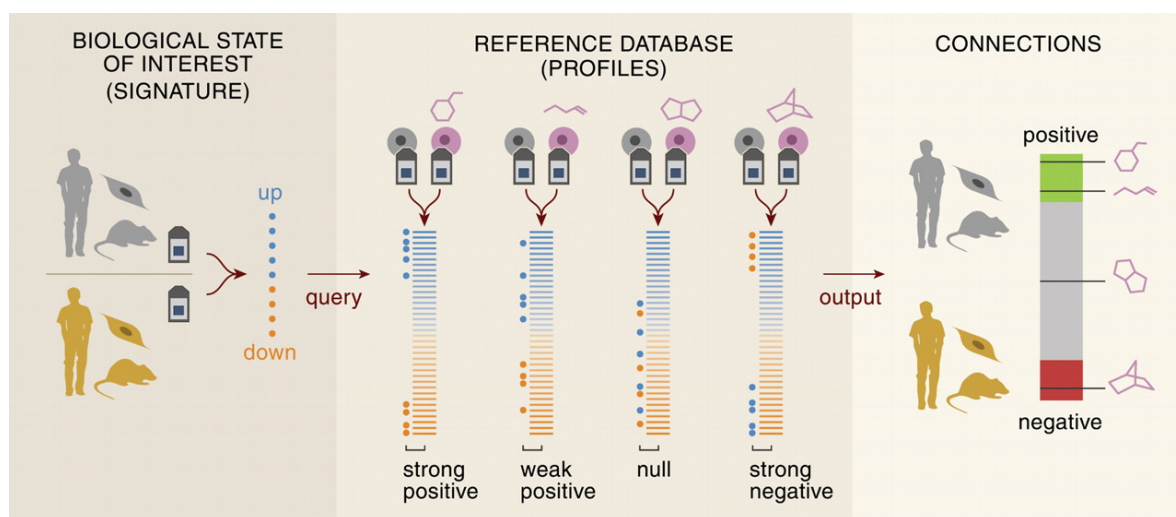


Figure 6. Connectivity map is a reference database of gene expression profiles built for hundreds of perturbagens in cultured human cells. A query gene expression signature (for a new compound) is scored for similarity/dissimilarity to/against each reference profile in the database using pattern matching algorithms. The “connectivity score”, as a measure of similarity, is used to rank the profiles that are connected to the query. Figure reused from (Lamb, Crawford et al. 2006), with permission from The American Association for the Advancement of Science.

Very recently, a new generation connectivity map (CMap-L1000v1) was introduced with 400,000 signatures using the L1000 platform (Subramanian, Narayan et al. 2017). Since standard library construction with RNA sequencing of many compounds can be costly, miniaturized approaches are desired. L1000 is a strategy involving ligation-mediated amplification of 1,000 landmark transcripts followed by capturing the amplified products on fluorescently addressed microspheres (Peck, Crawford et al. 2006). The data on the 1,000 landmark transcripts was used to computationally infer the expression level of non-measured transcripts. Since the new connectivity map is relatively new, its efficiency and applicability is yet to be determined.

Ye et al. published another strategy called DRUG-seq (Digital RNA with perturbation of Genes) for miniaturized high-throughput transcriptome profiling that can be used in drug discovery (Ye, Ho et al. 2018). DRUG-seq reduces the cost of RNA sequencing by 100 fold. In the proof-of-concept experiments, the authors profiled the transcriptional changes of 433 compounds across 8 doses. The transcriptional signatures could group the compounds into MOA clusters based on known targets.

However, all connectivity map efforts have been based on gene expression (mRNA) profiles, despite the fact that the cellular levels of proteins often differ significantly from those of mRNAs (Edfors, Danielsson et al. 2016; Liu, Beyer et al. 2016). A major reason for the lack of proteomics connectivity maps is the cost and labor of the analysis with respect to the required depth, which can be significantly higher in proteomics. Only recently, a proteomics connectivity map was built comprising the phosphoproteomic and chromatin signatures, quantitating 100 phospho-peptides and 59 histone modifications for 90 drugs in 6 human cancer cell lines (Litichevskiy, Peckner et al. 2017). However, such signatures might be irrelevant for most compounds. Due to the fact that proteins are the factual functioning biological entities in the cells, and most drug targets are proteins, analyzing proteomes can furnish further information and be more specific to drug MOA. After all, the biological systems are defined by protein expression and degradation collectively (Li and Biggin 2015; Savitski, Zinn et al. 2018). The balance between protein production and degradation is only reflected in proteomics data and cannot be assessed by genomics and transcriptomics in principle. Also, mRNA levels are not strongly correlated with protein concentrations even under steady state conditions (Edfors, Danielsson et al. 2016; Liu, Beyer et al. 2016), and this correlation must be even worse in dynamic situations, for example during programmed death, where many caspases-enhanced degradation pathways are activated (Nuñez, Benedict et al. 1998). Therefore, cataloging the proteomics fingerprints in response to a large number of diverse perturbations (compounds and treatments) can be the next step in building connectivity maps, by which drug targets and MOA of drugs can be studied in high detail, and new drugs can be characterized.

It should be noted that any mechanistic prediction resulting from the above computer-based analyses such as connectivity maps and other pattern-matching tools is just that – a prediction. Such predictions must be tested and validated by follow-up experimental approaches.

Furthermore, a major issue with these tools is that all imaginable perturbations must be recorded to build an ideal system. This means that specific inhibitors of all cellular proteins must be profiled, which can be practically impossible. In majority of cases, these inhibitors do not exist or if so, they are unspecific. Another disadvantage is that a large perturbation database would make it difficult to interpret the results, as a multitude of perturbations might be scored as similar to the query signature.

One of the main purposes of the current investigation was to build a proteomic signature resource of anticancer compounds to assess the feasibility of target and MOA deconvolution. We have tried to take a step back and rather focus on miniaturization, where a minimal number of experiments would be enough for overall assessment of compound MOA and identification of its potential targets. We consider this resource called ProTargetMiner, not a rival, but a complement for preceding drug target deconvolution databases such as connectivity maps. The modeling paradigm used in ProTargetMiner can be also employed in transcriptomics, as long as the biological endpoint used is identical, as in ProTargetMiner.

2 AIMS

2.1 BROAD AIMS OF THE THESIS

A major part of this project was devoted to further development of FITeXP methodology and other chemical proteomics techniques. Subsequently, FITeXP was coupled with other drug target deconvolution techniques such as TPP and redox proteomics to reveal the target and MOA of novel and interesting anticancer compounds. Where needed, complementary techniques were used to validate the findings of the above-mentioned techniques. We tried to further enhance the FITeXP performance by inclusion of matrix-detached cells in the proteomics analysis (Saei, Sabatier et al. 2018). This is while previously in FITeXP and the majority of molecular biology experiments, only matrix-attached cells were included and detached cells were usually ignored.

The main building block of this project was on making a deep proteomic database of model major cancer cell line response to a library of FDA-approved drugs with known targets and MOA, as well as several novel anticancer compounds with unknown mechanisms. Features were adopted from original FITeXP method (Chernobrovkin, Marin-Vicente et al. 2015) to identify drug targets and deconvolute the MOA of the compounds in the library. The projection of proteome signatures in the multidimensional space or hierarchical clustering revealed the drugs which had similar gross mechanisms. OPLS-DA methodology was used to determine compound-specific proteins. Furthermore, a public R shiny package was built for merging with user queries for novel compound.

Along with building this database, we opted to apply chemical proteomics tools to investigate in detail the mechanism of interesting compounds auranofin and b-AP15. For this purpose TPP and FITeXP were used in parallel. Redox proteomics as well as time and concentration series experiments were also used to further look in post-treatment proteome dynamics.

Finally, we developed a method called System-wide Identification of Enzyme Substrates by Thermal Analysis (SIESTA) for extending the applications of TPP to discover novel specific protein substrates for enzymes. The underlying assumption in SIESTA is that the enzymatic post-translational modification of a substrate protein can change its stability profile. The substrates identified in SIESTA can serve as building blocks of high throughput screenings.

2.2 SPECIFIC AIMS

The specific aims of the current thesis, as presented in five scientific publications (**Papers I-V**), are the following:

- Objective 1 (**Paper I**): To finalize the development of a general proteomics-based method for identification of targets and molecular processes triggered by anticancer treatments.

- Objective 3 (**Paper II**): To make a deep proteomic database of model cancer cell line response to a library of FDA-approved anticancer library and novel anticancer compounds with unknown or complicated mechanisms. Develop tools to predict the target and the MOA of the compounds in the library. Finally, make a package by which user data can be merged and analyzed.
- Objective 3 (**Papers III and IV**): To apply the developed method along with other available target deconvolution strategies to a number of interesting anticancer compounds to unravel their targets, off-targets and cellular effects in detail.
- Objective 4 (**Paper V**): To extend TPP for identifications of specific substrate proteins for enzymes. As many enzyme are potential drug targets, information on specific protein substrates can be used to design and develop compound screening platforms for drug discovery.

3 METHODS

3.1 CHEMICAL PROTEOMICS PARADIGMS

Routine chemical proteomics methods used in the current thesis and their applicability areas have been covered in **Figure 7**. Apart from FITExP and TPP which were discussed in the introduction, other complementary methods such as redox proteomics as well as time- and concentration-series proteomics were used in this thesis to further dig into compound-induced cellular effects.

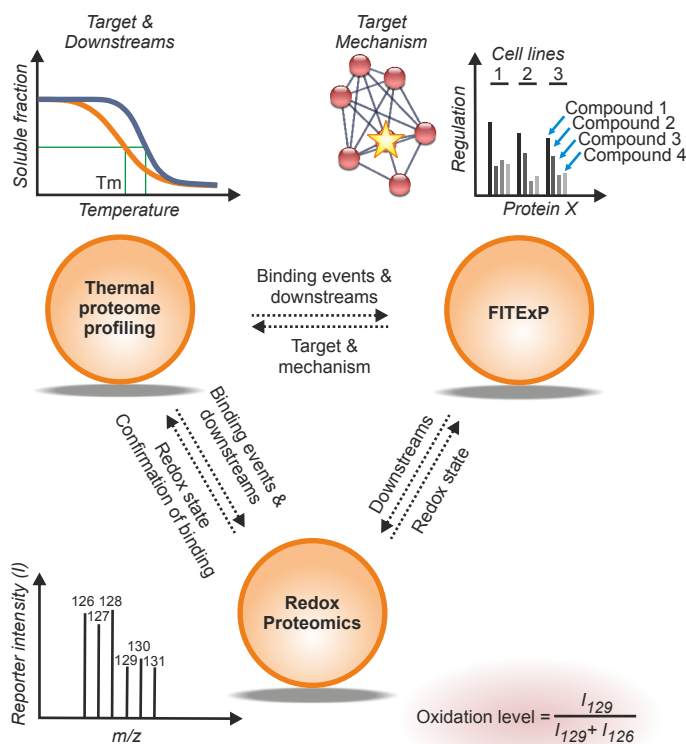


Figure 7. Complementary chemical proteomics strategies for deconvolution of targets and characterization of cellular downstream effects. While TPP provides information related to stability change of the target and downstream proteins, FITExP reveals disturbed cellular pathways and shortlists the target and mechanistic proteins. Redox proteomics reveals the changes in the oxidation state of cysteine, which can correlate to stability change of the corresponding proteins in TPP.

3.2 LABEL FREE SAMPLE PREPARATION

Label-free quantification is used to determine the relative amount of proteins in different samples and is thus mostly comparative. On the contrary to other protein quantification schemes, label-free strategy does not involve labeling of the proteins or peptides prior to quantification. In this approach, proteins are extracted, reduced with DTT, and subsequently alkylated with IAA. The proteins are then digested to peptides with specific enzymes such as trypsin and LysC, which are then cleaned and subjected to analysis (**Figure 8**). The classic label-free approach was used for all the experiments in **Paper I**, in FITExP experiment for auranofin (**Paper III**), and in the validation of ADP-ribosylation for SIESTA (**Paper V**).

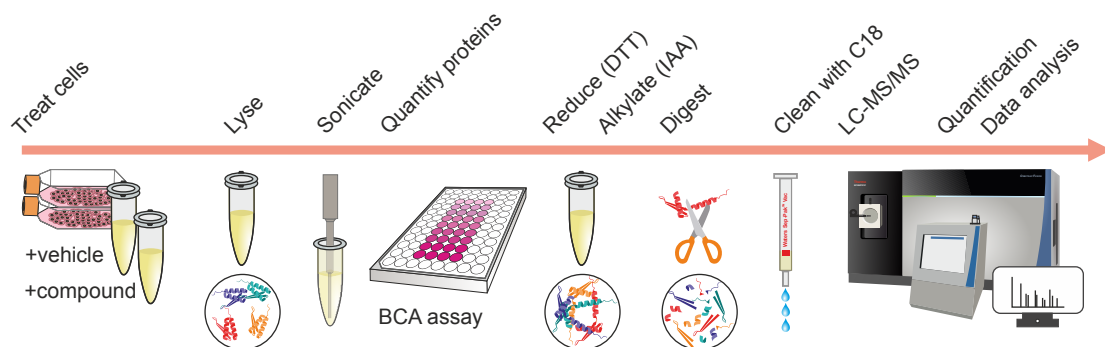


Figure 8. Label-free scheme for sample preparation for LC-MS/MS analysis.

3.3 TMT10 MULTIPLEXING

TMT10 multiplexing was the main proteomics scheme used throughout this work. The protocol is similar to label-free technique till digestion. Afterwards, the peptides are labeled with amine-reactive isobaric mass tags, pooled, cleaned and analyzed. TMT10 reagents consist of ten different isobaric labels with the same mass and chemical structure and are therefore isotopomeric (the general structure in shown in **Figure 9A**). Each tag has three constituents: amine-reactive NHS-ester group, a spacer arm and a mass reporter. In the mass reporter section, these labels contain a unique number and combination of ^{13}C and ^{15}N isotopes (TMT10 126-131Da), and can therefore be measured in the low-mass region of a high-resolution MS/MS spectrum upon peptide fragmentation (**Figure 9B**).

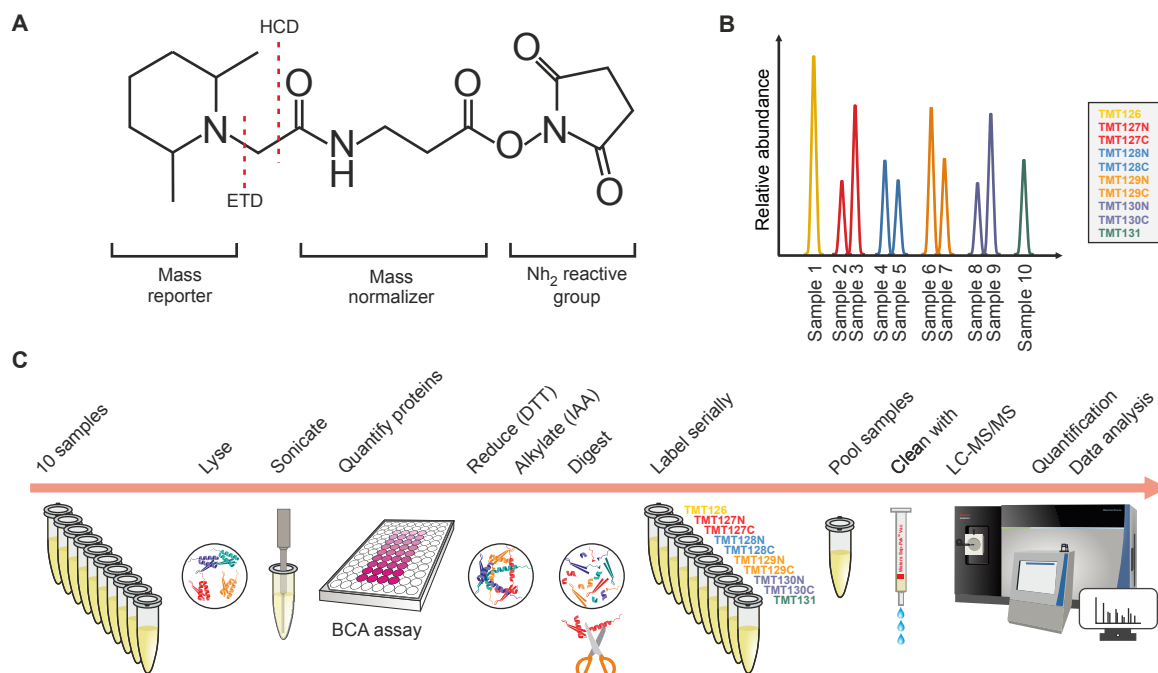


Figure 9. TMT10 labeling concept and workflow. **A)** Structural design of the TMT10 reagents. **B)** The relative abundance of peptide fragment in 10 different samples can be quantitated by comparing the reporter ion intensities which are generated by MS/MS fragmentation of the mass tags. **C)** Sample preparation with TMT10 multiplexing for LC-MS/MS analysis. HCD = higher-energy collisional dissociation and ETD = electron transfer dissociation.

Peptides from different samples are labeled with different tags, and therefore, the samples can be multiplexed and combined prior to analysis. Due to the multiplexed nature of samples, they can also be fractionated into peptide populations. This separation significantly simplifies the peptide mixture and enhances the proteome coverage captured by mass spectrometry. The overall workflow is given in **Figure 9C**.

3.4 SEQUENTIAL IODOTMT LABELING FOR REDOX PROTEOMICS

Within this thesis, redox proteomics was only applied in **Paper III**, where the changes in the redox state of proteins was measured in presence of auranofin in living cells. IodoTMT labels are currently six plex isobaric reagents which are used for covalent and irreversible labeling of sulfhydryl (—SH) groups. The concept is similar to TMT10, but the iodoTMT labels only react with reduced cysteines. The major application of these labels is in relative quantitation of cysteine modifications such as oxidation and disulfide bonds. In this approach, after extraction of cellular proteins, the already reduced cysteine are labeled with the first set of iodoTMT reagents and after reduction of disulfide bonds, the remaining cysteines are labeled with the second set of iodoTMT. The samples are then combined, digested, fractionated and subjected to mass spectrometry analysis (workflow in **Figure 10**). The oxidation of a peptide can thus be measured within the same sample by dividing the intensity of second label to the sum of the first and second label intensity.

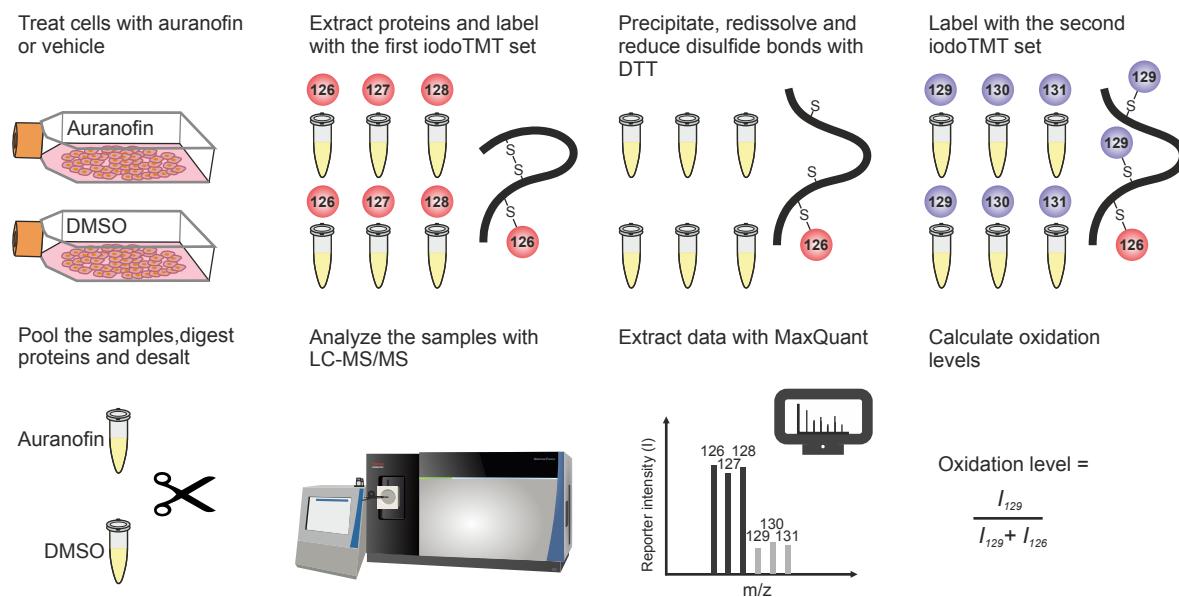


Figure 10. The redox proteomics workflow based on sequential labeling with iodoTMT reagents.

3.5 PROTEOMICS

All samples analyzed in this thesis were digested in solution. Proteins were reduced with DTT and then alkylated with IAA. Proteins were sequentially digested with LysC and trypsin. The obtained peptides were then cleaned using C18 columns. In experiments with TMT

multiplexing, the peptides were fractionated either using Pierce High pH Reversed-Phase Peptide Fractionation Kit (Thermo) or by a XBrigde BEH C18 2.1x150 mm column (Waters), using a Dionex Ultimate 3000 2DLC system (Thermo) for higher coverage (96 fractions concatenated to 16-24 based on the desired depth).

3.6 LC-MS/MS

Samples were loaded (in random order) with buffer A (0.1% formic acid (FA) in water) onto an EASY-Spray column (75 μ m internal diameter, packed with PepMap C18, 2 μ m beads, 100 Å pore size, 50 cm) connected to either the EASY-nLC 1000 (Thermo; Cat#LC120) or to a nanoflow Dionex UltiMate 3000 UPLC system (Thermo) and eluted with a buffer B (98% ACN, 0.1% FA, 2% H₂O) gradient from 2% to 26-35% of at a flow rate of 250-300 nL/min. Mass spectra were acquired with Orbitrap Elite, Orbitrap Q Exactive, Orbitrap Q Exactive Plus, Orbitrap HF or Fusion mass spectrometers (Thermo) in the data-dependent mode (various instruments were used in different studies and details can be found in relevant papers). Peptide fragmentation was mostly performed via higher-energy collision dissociation (HCD) with energy set at 33-35 NCE.

3.7 QUANTIFICATION

The raw data from LC-MS were analyzed by MaxQuant (Cox and Mann 2008) in all the papers. The Andromeda engine (Cox, Neuhauser et al. 2011) searched MS/MS data against Uniprot complete proteome database. Cysteine carbamidomethylation was set as a fixed modification, and methionine oxidation was used as a variable modification. Trypsin/P was selected as the enzyme and less than two missed cleavages were allowed. A 1% false discovery rate (FDR) was applied to filter the results at both protein and peptide levels. Match between runs was enabled in all the analyses, when applicable. Default settings were used for all the other parameters.

In **Paper V**, for identification of ADP-ribosylation, mass spectra were converted to Mascot generic format (MGF) files using in-house written RAWtoMGF v. 2.1.3 package. The resulting MGFs files were searched against the UniProtKB human database (v. 201806) including 71,434 sequences. Mascot 2.5.1 (Matrix Science) was used for identification of peptide sequences. Enzyme specificity was set to trypsin, and only two missed cleavages or less were allowed. ADP-ribosylation of C, D, E, K, N, R and S residues as well as methionine oxidation were chosen as the variable modification, while carbamidomethylation was set as a fixed modification on cysteine.

3.8 DATA ANALYSIS

After removing all the contaminants, only proteins with at least two peptides were included in the final datasets in all the studies. Proteins with missing values were ignored in most of the analyses. Protein abundances were normalized by the total protein abundance (total intensity) in each sample. For multidimensional data analysis, a second normalization was performed by average protein abundance in different samples. Data were processed by Excel, R, Python, and SIMCA (Version 15, UMetrics, Sweden). All reported p values throughout the studies were from two sided student t-test.

OPLS-DA models were used throughout this thesis and were built either in SIMCA 14-15 or R. OPLS-DA method allows for supervised grouping of the compounds before analysis. Using these model, the signature of a given compound or set of compounds can be contrasted against the other molecules in the dataset. In these models, the compound-induced specifically up- or down-regulated proteins can be found on the extremities of the loading plot, where each protein is represented by a dot. All the other components are cast off to the second dimension. A tutorial example is detailed in **Figure 11**.

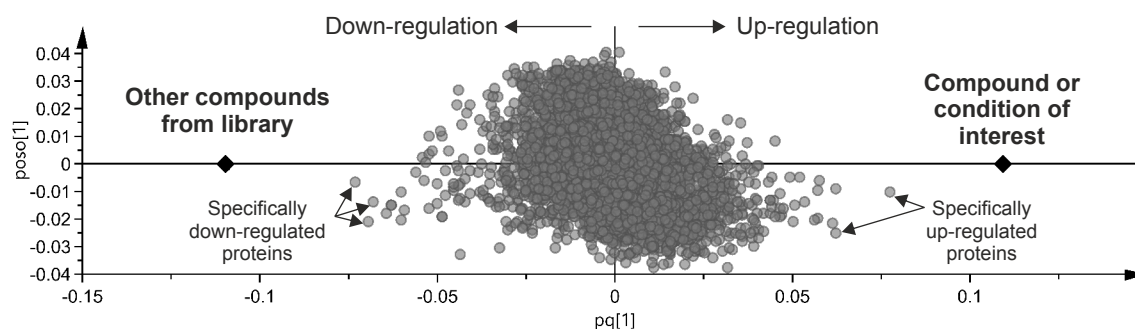


Figure 11. OPLS-DA modeling scheme. The proteomic signature of the compound (or any other perturbation) of interest (CLASS I) is contrasted against those of all other compounds (or a desired subset of compounds) (CLASS II). Proteins are represented by gray dots on the plot. Proteins on the right and left extremities of the x axis are most specifically up- and down-regulated proteins, respectively. Note that besides the magnitude of the effect, OPLS-DA takes reproducibility (significance) into account. Therefore, even proteins with minimal but reproducible changes can be identified using this strategy. Figure reused with permission from (Saei, Chernobrovkin et al. 2018).

4 RESULTS AND DISCUSSION

4.1 PAPER I. COMPARATIVE PROTEOMICS OF DYING AND SURVIVING CANCER CELLS IMPROVES THE IDENTIFICATION OF DRUG TARGETS AND SHEDS LIGHT ON CELL LIFE/DEATH DECISIONS

In **Paper I**, we focused on improving FITeXP by inclusion of proteomics data obtained from matrix-detached cells in the analyses. Chemotherapeutics usually cause the detachment of adherent cancer cells, culminating in cell death. The original FITeXP methodology and most molecular biology studies are based on the proteomic analysis of only matrix-attached cells after treatment and usually ignore matrix-detached cells. We hypothesized that since detached cells are more sensitive, they might reflect the chemotherapeutic-induced proteome changes in a better way. Furthermore, the proteomes of detached cells would possibly provide additional information on cell life and death decisions as well as sensitivity or resistance mechanisms to anticancer compounds. To test these hypotheses, we treated HCT-116, A375 and RKO cells for 48 h with 5-fluorouracil, methotrexate and paclitaxel and separately analyzed the proteomes of attached and detached cells.

Interestingly, the most obvious finding was that the proteomes of detached cells showed a higher dispersity than attached cells in the principal component analysis (PCA) (**Figure 12**), indicating that the extent of proteome change during cell death is higher than even cell type or the applied treatments. The bottom-line of such an observation could be that life exist within specific proteome boundaries and harsh distortion of the proteome space can lead to cell death.

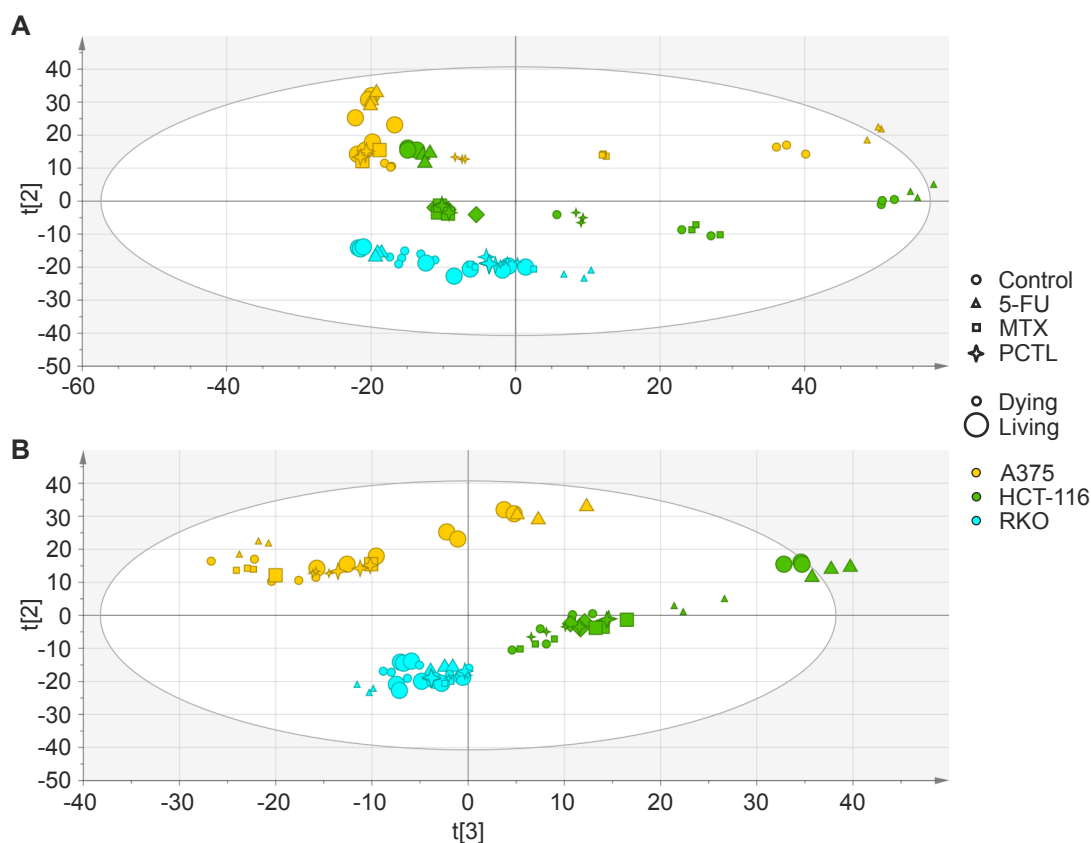


Figure 12. PCA of the whole proteome dataset of attached and detached cells ($Q2 = 0.705$). **A)** The first component separated the different states based on their attachment status, while the second component separated the tested cell lines. **B)** The third component apparently separated the proteomes based on carbohydrate metabolic parameters. Reused from (Saei, Sabatier et al. 2018) with permission from American Society for Biochemistry and Biophysics.

The separate proteomics datasets on attached and detached cells performed comparably in target and drug MOA deconvolution, and when used in combination improved the target ranking for paclitaxel significantly and for 5-fluorouracil marginally.

To discover proteins that specifically respond to a treatment of interest, we borrowed the specificity concept from FITE_xP method (Chernobrovkin, Marin-Vicente et al. 2015). It is known that the molecular components of a pathway targeted by a compound present differential regulation. However, simple differential regulation (ratio of protein abundances in cell treated with drug divided by vehicle control) is not an efficient way to approach drug target deconvolution, as similar generic effects such as stress response, detoxification or death pathways can be triggered by different molecules. Therefore, we introduced specific regulation as: “the ratio of protein regulation in response to the drug of interest, to the median regulation in all other drugs or treatments in the library” (Chernobrovkin, Marin-Vicente et al. 2015). This approach can thus highlight the proteins that are specifically responding to a particular molecule. Although specificity values can be easily calculated, for visualization purposes, we also employ a sophisticated supervised classification method called OPLS-DA for characterization of specific proteins (Bylesjö, Rantalainen et al. 2006).

Since these models were very useful in visualization of proteome responses, they were used throughout the whole project. OPLS-DA models were therefore built for the compounds in the study and their targets could be identified in top positions (**Figure 13**).

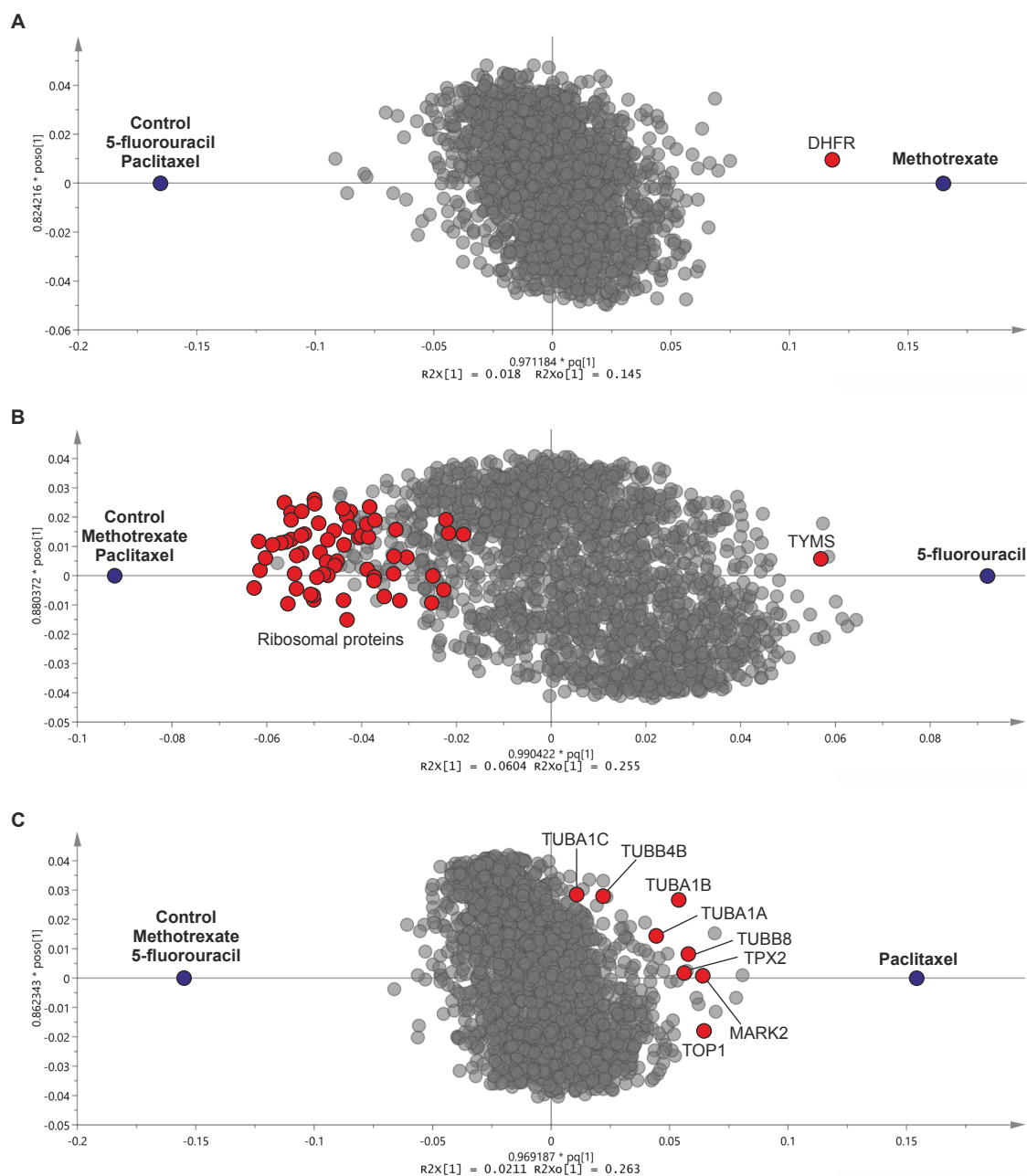
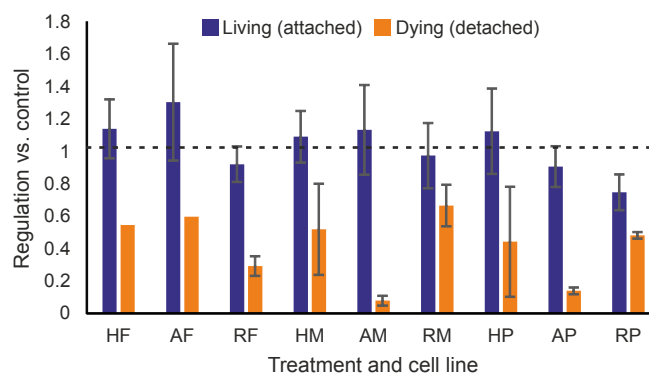


Figure 13. OPLS-DA loading enables the (discovery and) visualization of drug targets and mechanistic proteins in FITExP. OPLS-DA loadings for **A**) methotrexate, **B**) 5-fluorouracil and **C**) paclitaxel vs. all other states in different cell lines. The drug targets and other proteins directly involved in drug MOA are shown in red. Adapted from (Saei, Sabatier et al. 2018) with permission from American Society for Biochemistry and Biophysics.

We further analyzed the data to potentially discover features differentiating life from death in attached vs. detached cells. There were several proteins that showed consistent up- or down-

regulation in either class, regardless of the cell line or the drug used. Six proteins consistently up- or down-regulated in the attached vs. detached cells were selected (a representative example has been shown in **Figure 14**) and knocked down each by two specific (functionally verified or predesigned) siRNAs in the presence and absence of compounds to decipher their roles in cell death/survival. The rationale and workflow of the siRNA experiments is summarized in **Figure 15A-B** and the results of siRNA experiments in three cell lines in presence and absence of the test compounds is given in **Figure 15C**.

Figure 14. EIF4H as a representative protein which is down-regulated in all types of dying treated cells with all the treatments. Error bars not available for detached AF and HF, as EIF4H was not quantified in 1-2 replicates (H = HCT116 cells, A = A375 cells, R = RKO cells, F = 5-fluorouracil, M = methotrexate and P = paclitaxel). Reused from (Saei, Sabatier et al. 2018) with permission from American Society for Biochemistry and Biophysics.



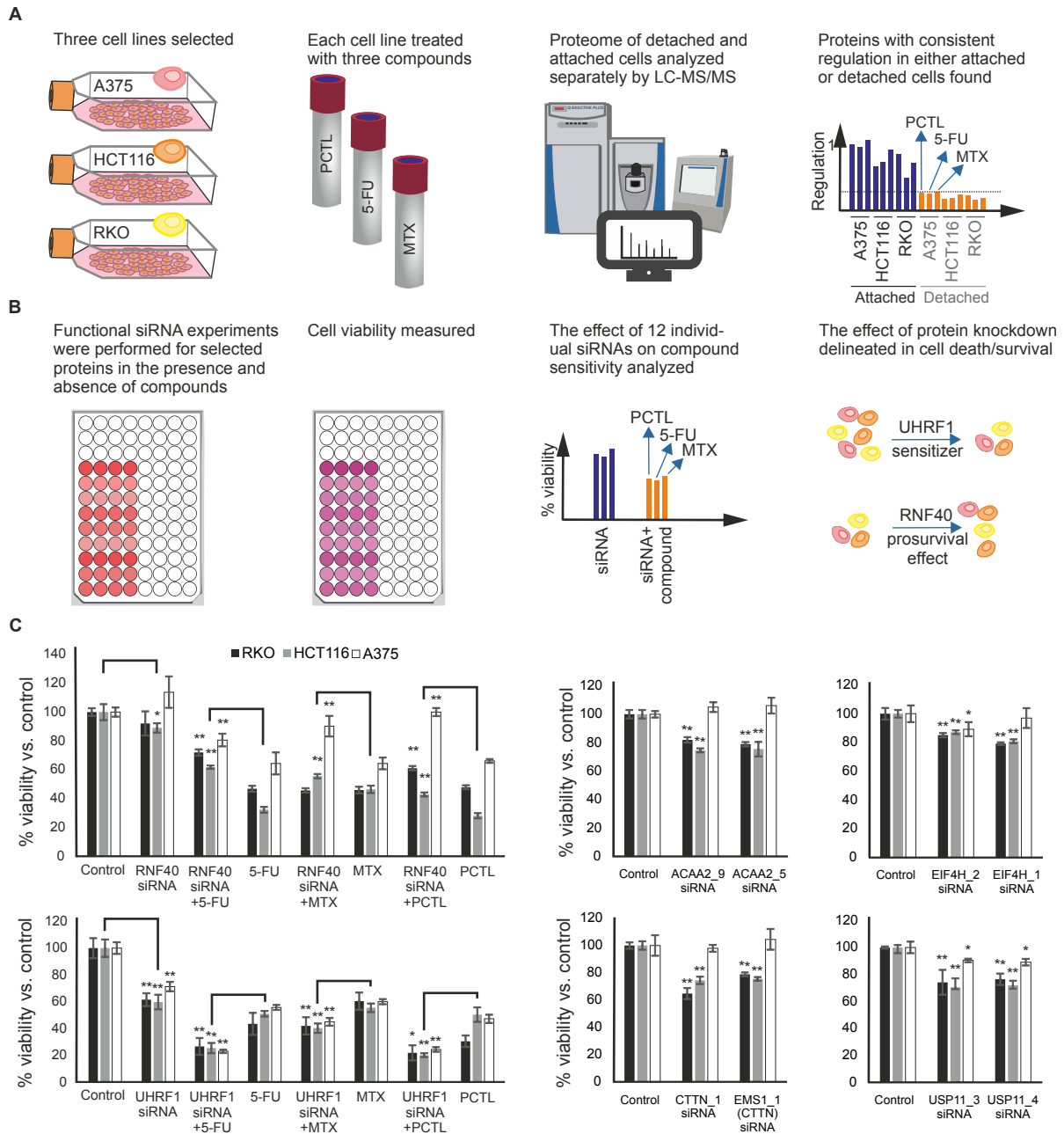


Figure 15. The rationale and design of follow-up siRNA experiments. **A)** The rationale and methodology for selection of 6 proteins potentially differentiating cell death vs. survival. **B)** siRNA experiments workflow for functional validation. **C)** The effect of different siRNA knockdowns on the viability of cells in the presence or absence of compounds. Silencing CTTN, USP11, ACAA2 and EIF4H did not affect the cell viability in the presence of drugs (error bars represent the SD in 4 replicates). Reused from (Saei, Sabatier et al. 2018) with permission from American Society for Biochemistry and Biophysics.

Conclusions

In summary, this study showed for the first time, that on the contrary to the common line of thought, the proteomes of matrix-detached cells can be studied. As shown in this work, the transition to cell death had a larger impact on cell proteome than different cell type or changes exerted by anticancer compounds. It was shown that data from matrix-detached cells are a

valuable companion to attached cells in studying drug target behavior and MOA. The study also led to the discovery of proteins characteristic of cell death or survival irrespective of the cell line and type of treatment. Some of these proteins could be potential drug targets.

However, there are a number of inherent limitations. This method can only be applied to adherent cells. Moreover, although this approach provides more information, it also doubles the number of samples in comparison with classic FITExP. Finally, some anticancer agents or treatments might disrupt the integrity of cell membrane and cannot be studied through this method.

4.2 PAPER II. PROTARGETMINER: A PROTEOME SIGNATURE LIBRARY OF ANTICANCER MOLECULES FOR FUNCTIONAL DISCOVERY

As a further development of FITExP method, we built a publicly available expandable proteome signature library of anticancer molecules in A549 adenocarcinoma cells, called ProTargetMiner. The objectives of this study was to demonstrate the general applicability of FITExP methodology for different classes of compounds and to provide a public platform where the MOA of new compounds can be analyzed at the protein level. The main objective of ProTargetMiner is depicted in **Figure 16**. This resource contains the data on 287 proteome signatures for 56 compounds. The main dataset contains 7,328 proteins and 1,307,859 refined protein-drug data points.

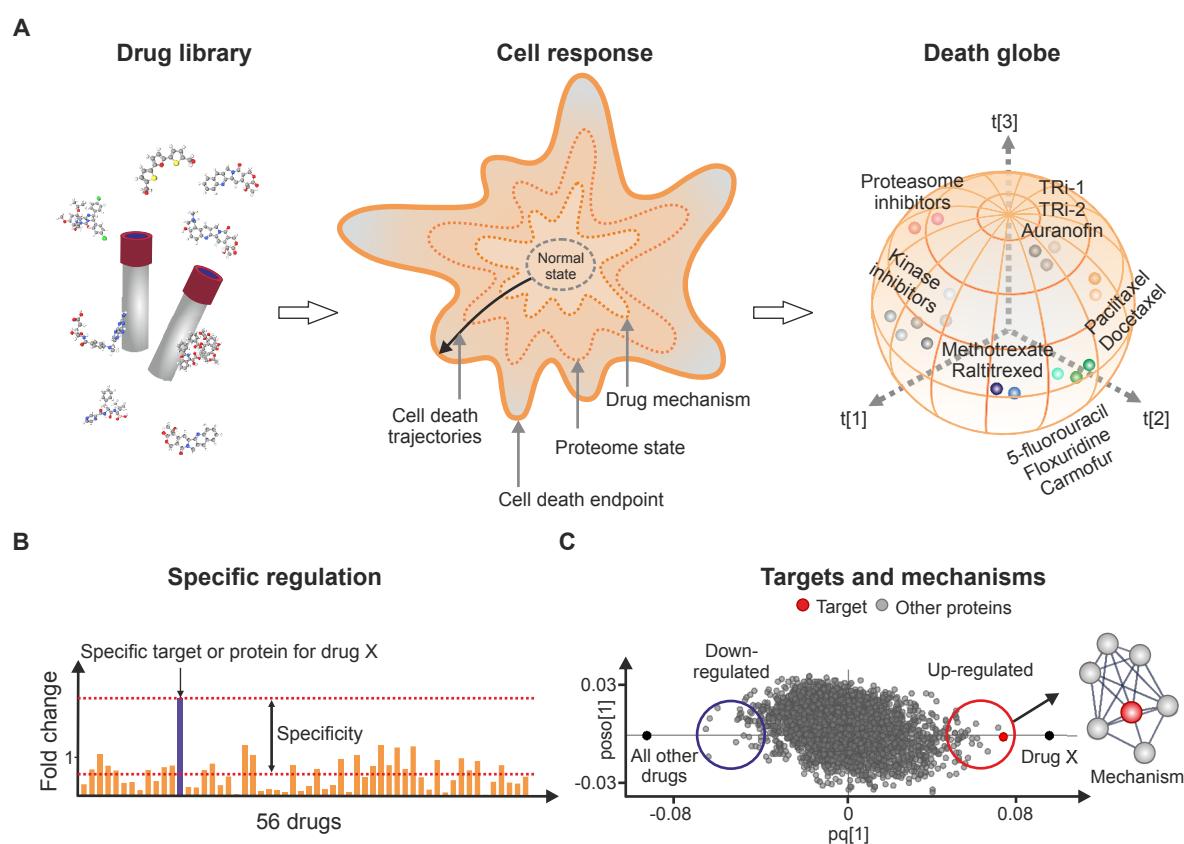


Figure 16. Objectives of ProTargetMiner. **A)** ProTargetMiner task, among other things, is to identify compounds with similar MOA in multidimensional space and provide specific information on drug targets, and protein involved in the cellular processes connected to the compound under study. **B)** The basic assumption is that drug targets and mechanistic proteins are specifically regulated. **C)** Compound-specific targets and proteins are pulled out using OPLS-DA-enabled specificity analysis. Figure adapted with permission from (Saei, Chernobrovkin et al. 2018).

4.2.1 Construction of ProTargetMiner

For the original dataset comprising 56 molecules, human A549 lung cancer cells were used as a model system for two reasons. First, A549 is a popular cell line harboring wild-type p53 and

is extensively characterized in the literature; and second, in the viability experiments this cell line showed a higher sensitivity to the compound library than the MCF-7 and RKO cells. The compound library comprised of 118 molecules cherry picked for cancer indication from Selleckchem FDA-approved drug library. Having compounds with known targets and MOAs was a prerequisite to successful establishment and benchmarking of the ProTargetMiner methodology. Filtered by sensitivity, 56 compounds were finally chosen (plus some compounds from collaborators with undefined or less defined targets or MOAs). The compounds grossly belonged to 19 different classes with versatile targets (112 known targets in total) and mechanisms, according to available information curated from DrugBank (<https://www.drugbank.ca/>) in January 2019. For each molecule, a concentration was used killing 50% of the cells (LC50) after 48 h and the respective proteomes were analyzed in at least three replicates. Methotrexate, paclitaxel, and camptothecin were included in each multiplexed experiment for normalization and quality control purposes (ProTargetMiner workflow in **Figure 17**).

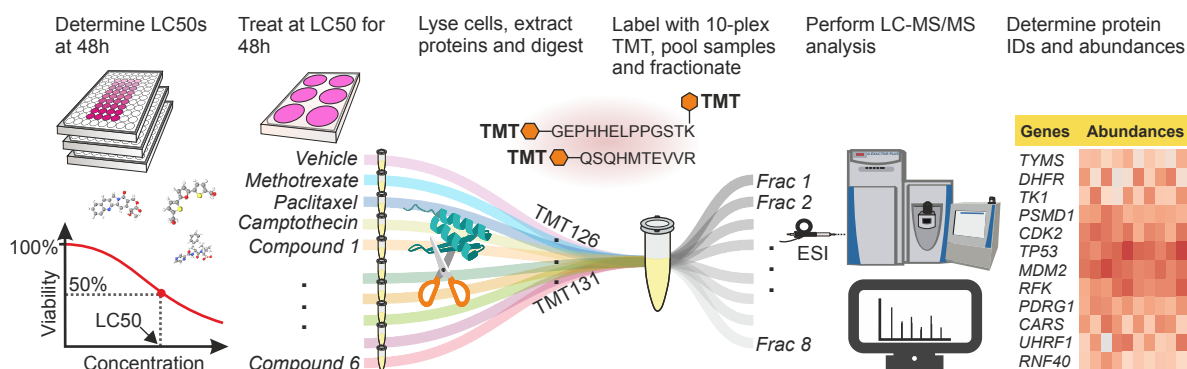


Figure 17. ProTargetMiner workflow. LC50 values were determined for the compound library (n=118) plus some other compounds in three cell lines; A549 cells were chosen based on their higher sensitivity; compounds were selected based on their cytotoxicity; cells were treated with 56 compounds and samples were prepared for shotgun proteomics in at least three independent biological replicates; samples were multiplexed in each experiment (methotrexate, paclitaxel and camptothecin were included in all 9 experiments); samples were lysed, digested, and labeled with TMT-10plex; pooled within each experiment and fractionated to increase the proteome coverage; individual fractions were analyzed by LC-MS/MS, followed by protein identification, quantification and data post-processing. Figure reused with permission from (Saei, Chernobrovkin et al. 2018).

Having an identical biological endpoint (here 50% cytotoxicity) was essential, as it would make the comparison of the proteome signatures possible. Any differences in proteome signatures was thus related to dissimilarities in targets, MOA and cell death. Therefore, a special attribute of ProTargetMiner vs. connectivity map efforts, is the normalization of the biological endpoint. In other connectivity map efforts, fixed or random concentrations of compounds have been used which might not be biologically relevant.

4.2.2 Proteomic dissection of cell death trajectories

It is known that compound signatures can be successfully separated by mapping the whole proteome or using even a subset of 1000 most abundant proteins (Chernobrovkin and Zubarev 2016). Since the ProTargetMiner library is composed of >50 molecules with diverse MOA, it can potentially probe most cell death pathways. We would like to envision that these pathways are represented by cell death trajectories in the proteome space, which encompasses all possible cell states. We define cell death trajectory as a track in the proteome space, which is caused by a cytotoxic agent and passes from a normal living state to a death state (Saei, Chernobrovkin et al. 2018). The proteome space and death trajectories are shown in schematics in **Figure 16**.

In the cell death community, there is a long-term debate and controversy on the number, nature and the molecular characteristics of cell death modalities (Galluzzi, Bravo-San Pedro et al. 2015). For approaching this problem statistically with proteomics, we set to determine the number of orthogonal dimensions in the dataset. At least 11 independent dimensions were discovered in the original ProTargetMiner dataset using factor analysis. Cyclin-dependent kinase inhibitor 1 (CDKN1A) and PCNA-associated factor (PCLAF) were the most contributing proteins with opposite signs to the first dimension. Interestingly, CDKN1A plays a role in p53 mediated suppression of cell proliferation (Harper, Adami et al. 1993), and PCLAF is a cell cycle-regulated protein that regulates DNA repair (Emanuele, Ciccia et al. 2011). The top 30 proteins of the first three dimensions mapped to “p53 signaling pathway and cell cycle”, “focal adhesion and angiogenesis” and “chromatin assembly and fatty acid metabolism”. The dimensions discovered in this dataset represent cell death trajectories as orthogonal pathways, or in other words theoretical constructs, and some of them might have little or no resemblance to the classical death modalities from textbooks (Zubarev, Nielsen et al. 2008). Expectedly though, some dimensions already corresponded to classic cell death modes, such as dimension 1 (p53-dependent apoptosis), dimension 4 (autophagy), and dimension 6 (macromitophagy). However, the other dimensions were hard to assign to known death modalities. Attributing the top proteins defining each dimension with the classic or novel cell death modalities can be an interesting subject for future research. Therefore, these dimensions and at least their top contributing proteins deserve detailed follow-up bioinformatics analysis and experimental validation.

4.2.3 Similar compounds produce similar proteome signatures

The nonlinear dimension reduction method t-SNE was used for dimension reduction of the data (Maaten and Hinton 2008). On this t-SNE plot that we call “death map”, all compound proteome signatures are projected as points. The proximity of signatures on this plot can be used to assess compound similarities. The compounds grouped by their proteomic signatures into expected clusters based on known targets and MOAs (**Figure 18**). As anticipated, well-known compounds with similar MOAs (e.g., tubulin depolymerization inhibitors, pyrimidine

analogues, thioredoxin reductase inhibitors auranofin, TRi-1 and TRi-2 (Stafford, Peng et al. 2018) and topoisomerase inhibitors) were close on the t-SNE plot.

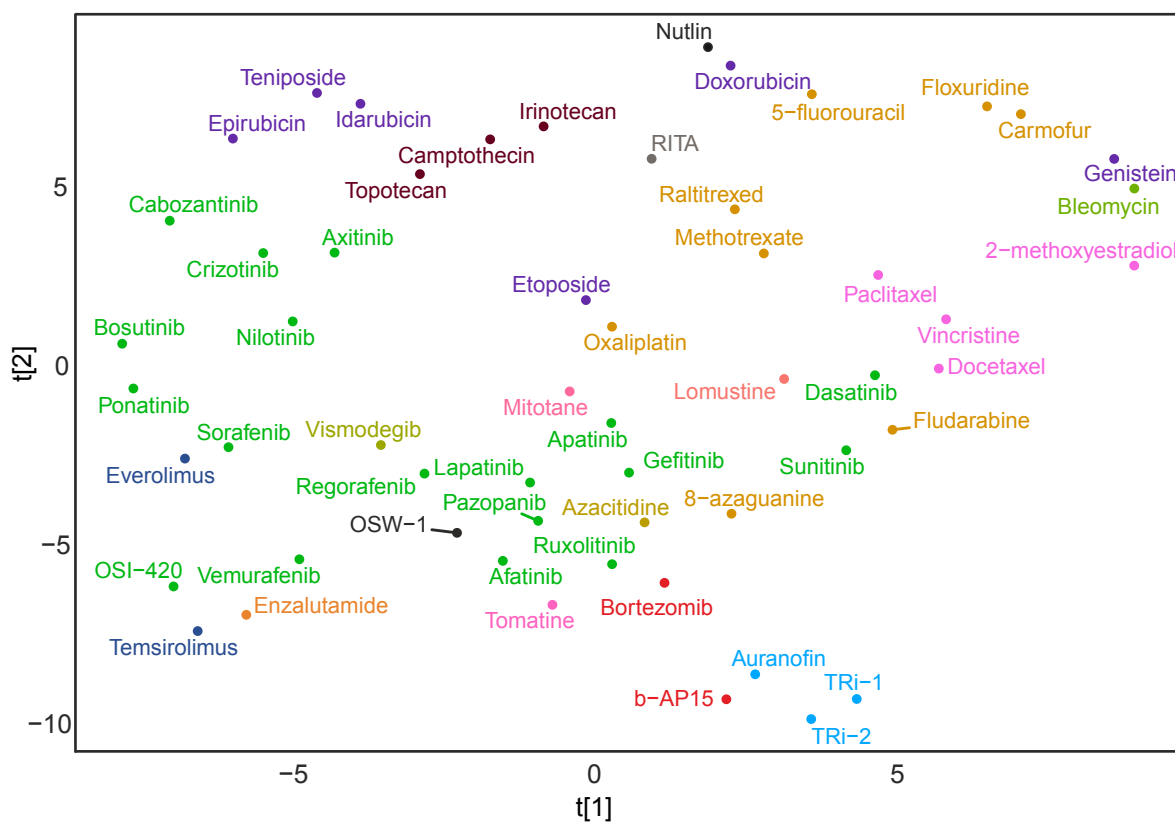
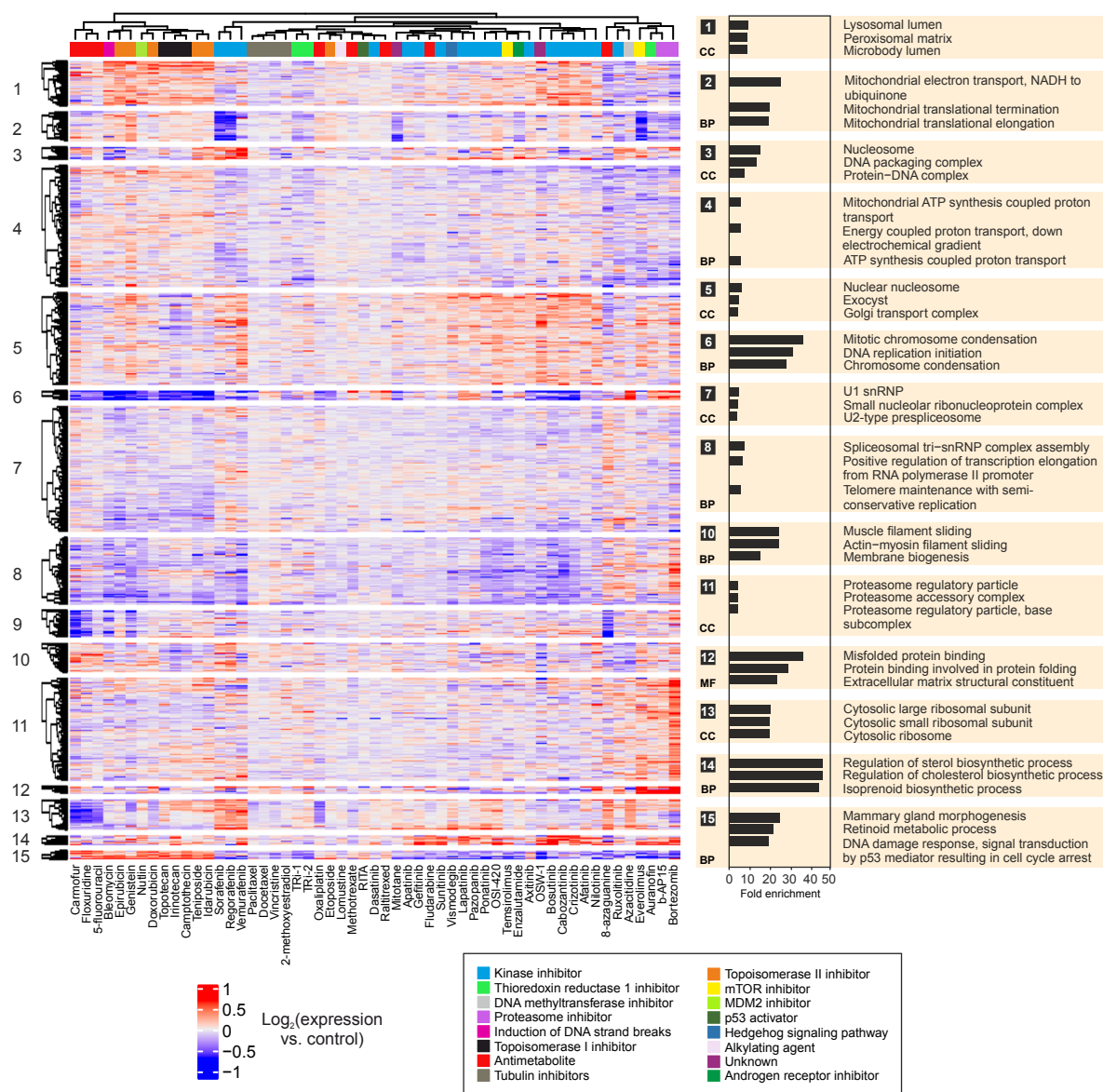


Figure 18. Drugs with similar MOAs have similar proteomics fingerprints and were mostly adjacent on the t-SNE plot (“death map”). Compound classes are shown with different colors. Figure reused with permission from (Saei, Chernobrovkin et al. 2018).

Since tomatine was an obvious outlier in t-SNE and PCA, its data was not included in the following analyses. The extraordinary proteome changes in response to tomatine happen in both directions (up- and down-regulation) and are likely associated with proteasome inhibition (da Silva, Andrade et al. 2017) and unspecific membrane damage (Roddick and Drysdale 1984), as reported before.

Hierarchical clustering was also used to separate the compounds into clusters. As shown in **Figure 19**, the compounds were mostly separated in mechanistic clusters based on known targets/MOAs. Compounds known to directly or indirectly induce DNA damage were separated from the rest of the molecules. Tubulin inhibitors paclitaxel, docetaxel, vincristine and 2-methoxyestradiol, proteasome inhibitors b-AP15 and bortezomib, as well as thioredoxin reductase inhibitors TRi-1 and TRi-2, grouped together, indirectly validating the approach. The proteins in each cluster were subjected to GO enrichment. Interestingly, some of these clusters map to high density protein networks. For example, cluster 13 mapped to ribosome. Four compounds showed down-regulation of the ribosomal proteins, which is an indicator of ribosomal stress. Three of those molecules were pyrimidine analogues 5-fluorouracil,

floxuridine and carmofur. Therefore, floxuridine and carmofur potentially induce ribosomal stress, similar to 5-fluorouracil (Marin-Vicente, Lyutvinskiy et al. 2013). The fourth compound was an alkylating agent oxaliplatin. A recent paper actually showed that oxaliplatin mechanism is peculiar and not similar to other platinum analogues and that it involves induction of ribosome biogenesis stress (Bruno, Liu et al. 2017). Cluster 12 mapped to pathways related to protein folding and mainly shows up-regulation with proteasome inhibitors bortezomib and b-AP15. Interestingly, a number of compounds particularly kinase inhibitors such as lapatinib, bosutinib and gefitinib showed an up-regulation of (chole)sterol synthesis pathways in cluster 14.



4.2.4 Functional discovery in a compendium of proteome signatures

As shown in **Figure 19**, there are specific features for every compound that must be pulled out using statistical approaches. We employed OPLS-DA modeling scheme presented in **Paper I** to identify specifically regulated proteins, among which drug target and mechanistic proteins are found. In the original FITeXP method and **Paper I**, a panel of cell lines and a number of contrasting compounds were used to increase the specificity for deconvolution of targets and mechanistic proteins. The hypothesis of **Paper II** was to investigate if an equivalent increase in specificity can be achieved by a multitude of contrasting compounds in a single cell line. This would dramatically facilitate the applicability of the method for characterization of novel compounds. Therefore, we built OPLS-DA models for every molecule in this study, contrasting its proteome signature (class I) against all the other compounds (class II). OPLS-DA model loadings for methotrexate, paclitaxel, and vincristine show the deconvolution of cognate targets in **Figure 20A-C**. DHFR was found as a straightforward target for methotrexate. Paclitaxel and vincristine affect tubulin depolymerization and polymerization, respectively. Tubulins could be found among top-regulated proteins for both compounds. 30 top specifically regulated proteins close to x axis extremities were submitted to StringDB and the enriched pathways are shown in the right panel in **Figure 20**. These pathways reflect the known MOAs for the corresponding compounds.

The OPLS-DA models also revealed compound effects on protein complexes. For example, for bortezomib, which is a proteasome inhibitor, specific up-regulation of proteasome subunits was noted (**Figure 20D**). The average fold change of proteasome subunits was 1.15 vs. control. Not only this shows the efficiency of OPLS-DA in pulling out subtle effects, but it also shows how a small perturbation in proteasome expression can have deleterious consequences for the cell. Likewise in sorafenib model (**Figure 20E**), NADH dehydrogenases and mitochondrial ribosomal proteins were specifically down-regulated, validating previous findings in human neuroblastoma cells (Bull, Rajalingam et al. 2012). These results demonstrate that ProTargetMiner results can also be carefully generalized to other cell lines.

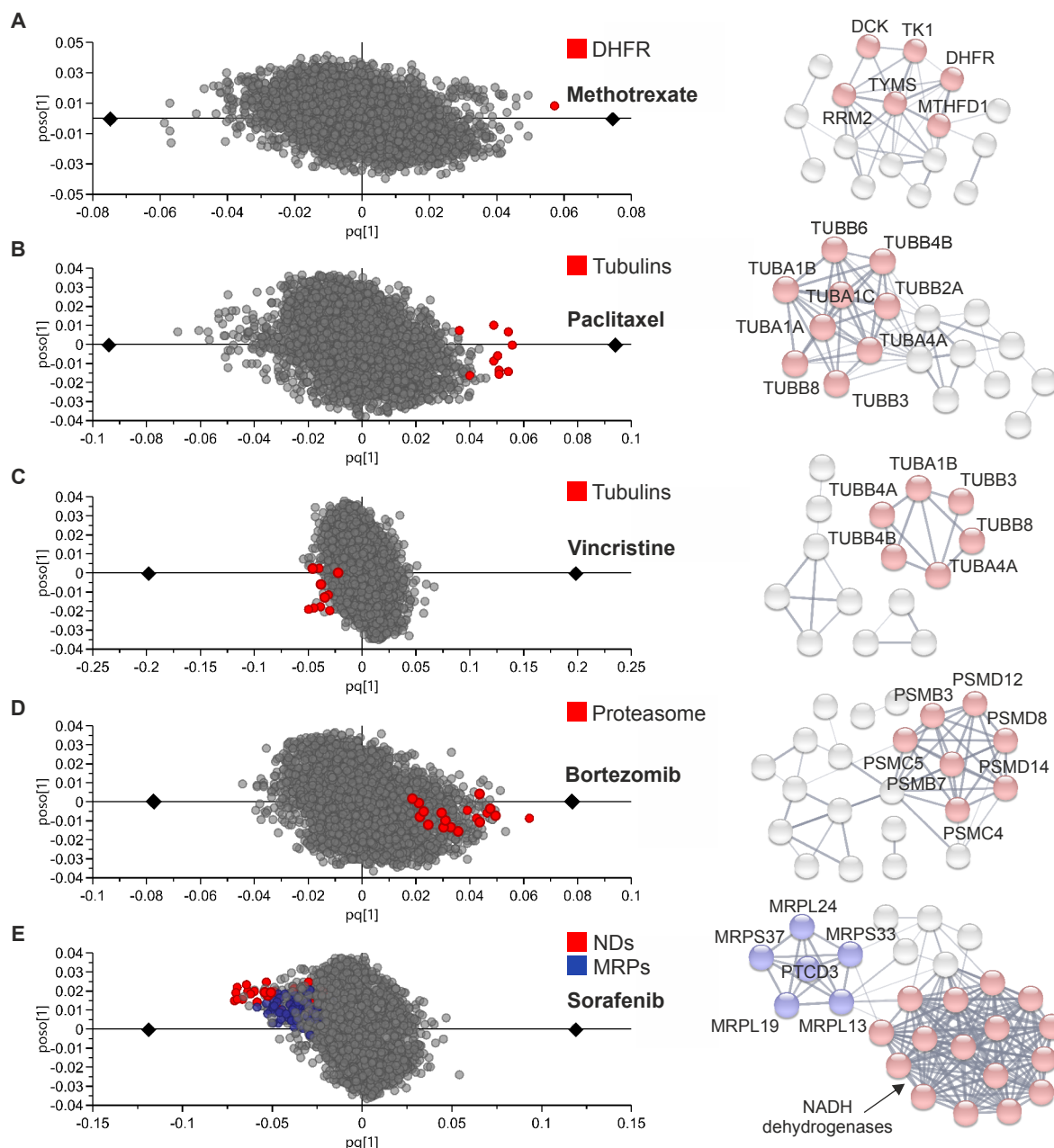


Figure 20. ProTargetMiner reveals the targets, MOA and affected cellular complexes. A-E) Exemplary OPLS-DA model loadings for several compounds show the successful deconvolution of targets and MOA. The results of pathway analysis for the 30 top specifically up- or down-regulated proteins is shown on the right panel: KEGG pathways for methotrexate - “one carbon pool by folate and pyrimidine metabolism”; paclitaxel and vincristine - “tubulin”; bortezomib - “proteasome”. GO terms for sorafenib - “NADH dehydrogenase activity” (in red) and “mitochondrial translation” (in blue) (all p values < 0.001). RPL = ribosomal proteins of the large and RPS = ribosomal proteins of the small subunit. MRP = mitochondrial ribosomal proteins, ND = NADH dehydrogenase. Disconnected proteins not involved in the pathway has been omitted. Figure adapted with permission from (Saei, Chernobrovkin et al. 2018).

4.2.5 The degree of drug-induced proteome changes

The next hypothesis was if the extent of proteome perturbation can be associated with the specificity of the compounds (the number of affected targets and MOAs). Therefore, we ranked all compounds by the global deviation of their proteome signatures from control (**Figure 21**). As expected, proteasome inhibitors were among the compound inducing the largest deviations. The largest proteome deviation was induced by bortezomib, while on the other side of the

spectrum, tubulin inhibitors caused the smallest proteome perturbation. While the large proteome deviation can be the result of protein accumulation upon inhibition of the proteasome, the small perturbation of the proteome with tubulin inhibitors at LC50 concentrations, potentially indicates the high specificity and a lack of off-target effects for the latter. Furthermore, the slightest proteome change induced by dasatinib is unexpected, as this drug has 23 annotated targets in DrugBank as of January 2019. There was no particular correlation between compound LC50 and the extent of proteome changes.

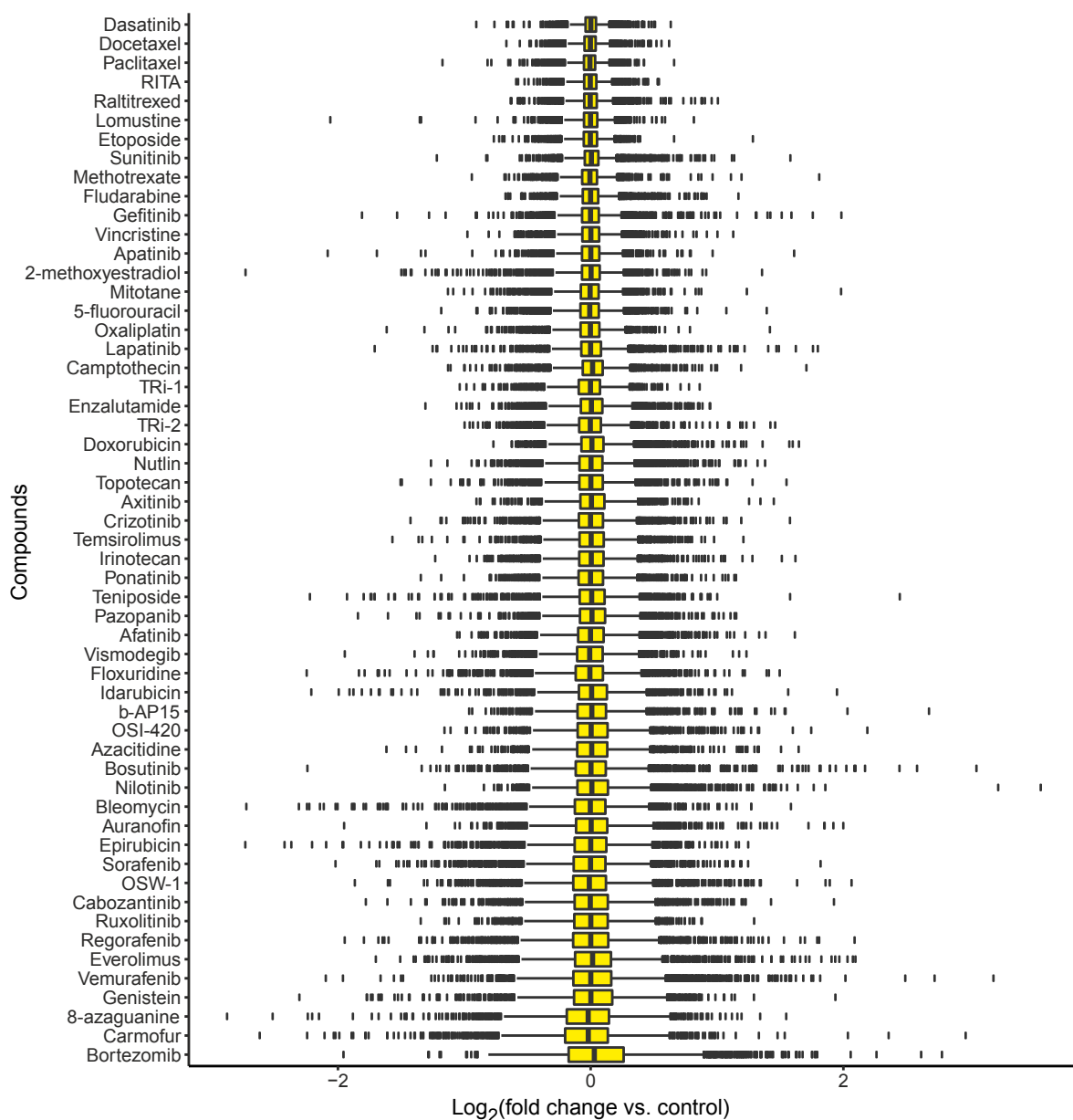


Figure 21. The extent of proteome deviation in response to each molecule, might be an indicator of compound specificity. Center line, median; box limits contain 50%; upper and lower quartiles, 75% and 25%; maximum, greatest value excluding outliers; minimum, least value excluding outliers; outliers, more than 1.5 times of upper and lower quartiles. Figure reused with permission from (Saei, Chernobrovkin et al. 2018).

4.2.6 Assessing the miniaturization possibility of ProTargetMiner concept

Connectivity map efforts aim to saturate the mechanistic space by profiling myriad of compounds (Subramanian, Narayan et al. 2017). There are certain inherent limitations to these approaches. First, this would be an open-end project, as inhibitors of every cellular protein is not available. Even if inhibitors were available, most would be multi-targeted and unspecific. Second, the results of a connectivity map is a ranking of perturbations, which if too many would hamper the interpretation of results.

Furthermore, in drug discovery and development, a detailed characterization of target and mechanism space, especially at the proteome level is desirable. Compound-induced effects are most advantageous if obtained in the most relevant biological models and cell types. However, building databases such as ProTargetMiner can be time-consuming, expensive and therefore not rational, at least not with the current technological platforms. Therefore, we tested if ProTargetMiner methodology could be miniaturized to a minimal compound panel size for effective deconvolution of drug targets and MOA. To obtain the optimal number of contrasting molecule needed for effective target and MOA characterization, we built PLS-DA models in R by using only a subset of contrasting compounds in the model ($n = 1-54$, 50 molecule combinations randomized for each n). In each iteration, the ranking of known targets for representative molecules camptothecin (target: TOP1), paclitaxel (target: tubulins), methotrexate (target: DHFR) and OSW-1 (target: OSBP1) and was calculated and the mean ranking for each n was obtained. As expected, the deconvolution process was more efficient with higher number of contrasting compounds (and not for randomly chosen non-target proteins) (**Figure 22**). Encouragingly, 8-10 contrasting signatures were already efficient for obtaining target rankings below 10.

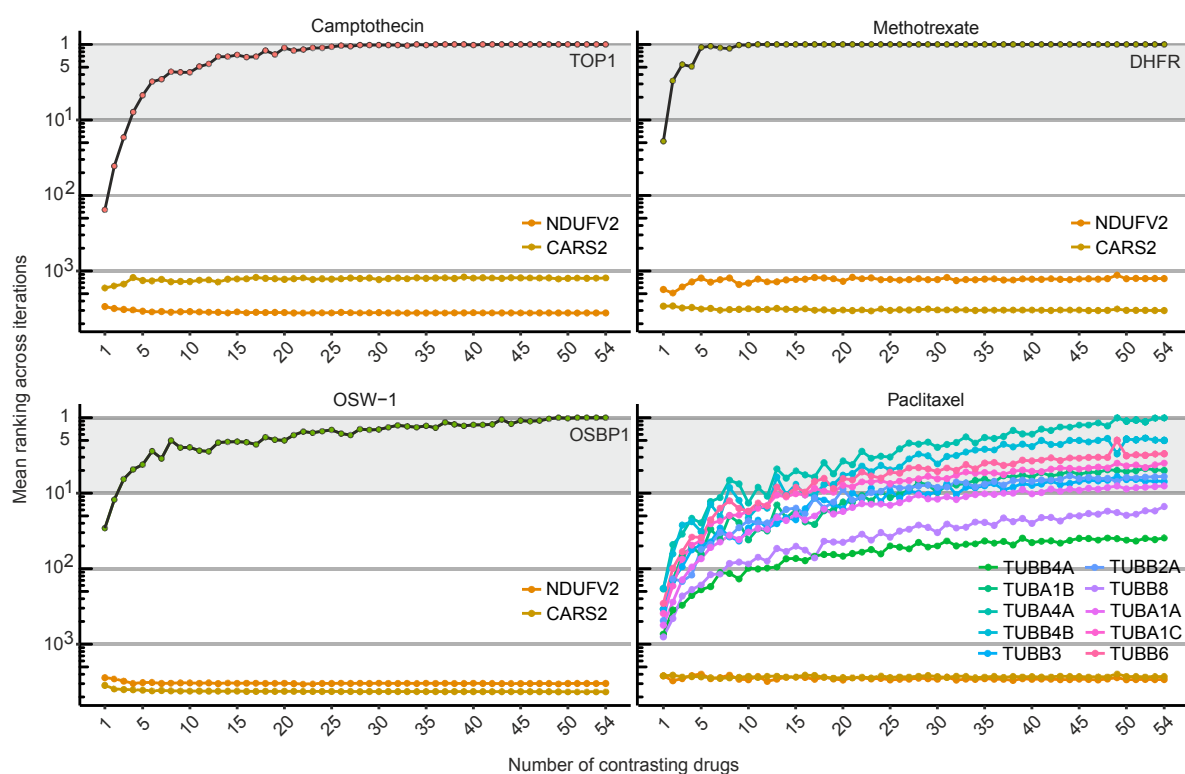


Figure 22. Assessing the miniaturization possibility of ProTargetMiner concept by calculating the optimal number of contrasting molecules. PLS-DA models were built in R for 4 representative compounds, where they were contrasted vs. 50 random combinations of 1-54 compounds and the drug target ranking was averaged for each number. NDUFV2 and CARS2 proteins were randomly selected as non-target proteins. Figure reused with permission from (Saei, Chernobrovkin et al. 2018).

4.2.7 Miniaturization of ProTargetMiner

With miniaturization opportunity at hand, we decided to make ultra-deep proteomic databases in three extensively used cancer cell lines, A549, MCF-7 and RKO, for 9 molecules representing most diverse MOAs for providing optimal contrast. These datasets could then be merged with external data from users interested in target deconvolution for their compounds. These 9 molecules were chosen from different orthogonal dimensions in the factor analysis of the original ProTargetMiner dataset with 56 compounds, and include 8-azaguanine (target: PNP), raltitrexed (target: TYMS), topotecan (target: TOP1), floxuridine (target: TYMS), nutlin (target: MDM2), dasatinib (target: at least 23 kinase targets), gefitinib (target: EGFR), vincristine (target: tubulin) and bortezomib (proteasome).

These deep dataset have a total depth of 11562 proteins, out of which 11293 are quantified with at least two peptides. The number of refined proteins with at least two peptides and no missing values in all the treatments was 7398, 8735 and 8551 in A549, MCF-7 and RKO cell lines, respectively. Therefore in total, 6496 proteins were quantified in all cell lines with no missing values.

Once again, we used the leave-one-out approach for confirming the validity of the databases and their applicability in target/MOA deconvolution. First OPLS-DA models were built for compounds in single cell lines and the representative examples are shown in **Figure 23**. TYMS

was among the top three protein for raltitrexed in MCF-7 cells (**Figure 23A**) and the four top proteins mapped to KEGG pyrimidine metabolism (3 proteins, $p = 0.0003$), which is basically the drug MOA. Similarly, tubulins were among the top proteins for vincristine in A549 cells (**Figure 23B**), and several kinase targets were identified as top proteins for dasatinib in different cell lines (**Figure 23C, F and G**). Interestingly, CYP1A1, which is involved in dasatinib metabolism (Wang, Christopher et al. 2008), was identified as the top specifically up-regulated protein in MCF-7 cells (**Figure 23C**) and its expression level was extraordinary compared to other drugs in this cell line (**Figure 23D**). The top specifically down-regulated protein in response to dasatinib in MCF-7 cells was Poly(ADP-ribose) glycohydrolase (PARG) (**Figure 23C**) which was specifically down-regulated in comparison to the other compounds (**Figure 23E**). Specific down-regulation of PARG was also noted in the other two cell lines (**Figure 23F-G**). While in all the cell lines, the (30) top specifically up-regulated proteins for bortezomib show an enrichment of ubiquitin (RPS27A) and other proteins involved in ubiquitination (**Figure 23H and J**), in MCF-7 cells, the top first and fifth most specifically down-regulated proteins were dipeptidyl peptidase 2 (DPP7) and dipeptidyl peptidase 3 (DPP3) respectively, which are known non-proteasomal targets of the proteasome inhibitor bortezomib (Arastu-Kapur, Anderl et al. 2011). The specific down-regulation of these two proteins in comparison with other compounds is shown in **Figure 23I**. However, these proteins were specific to MCF-7 cells and were not among the top proteins in A549 or RKO cells. These results clearly show that inclusion of 9 compounds in a single cell line provides enough specificity for drug target and MOA deconvolution in majority of cases. Furthermore, due to the special depth of and diversity of cell lines in the dataset, cell-specific targets, low-abundant targets (such as kinases), and drug-metabolizing enzymes might also be found.

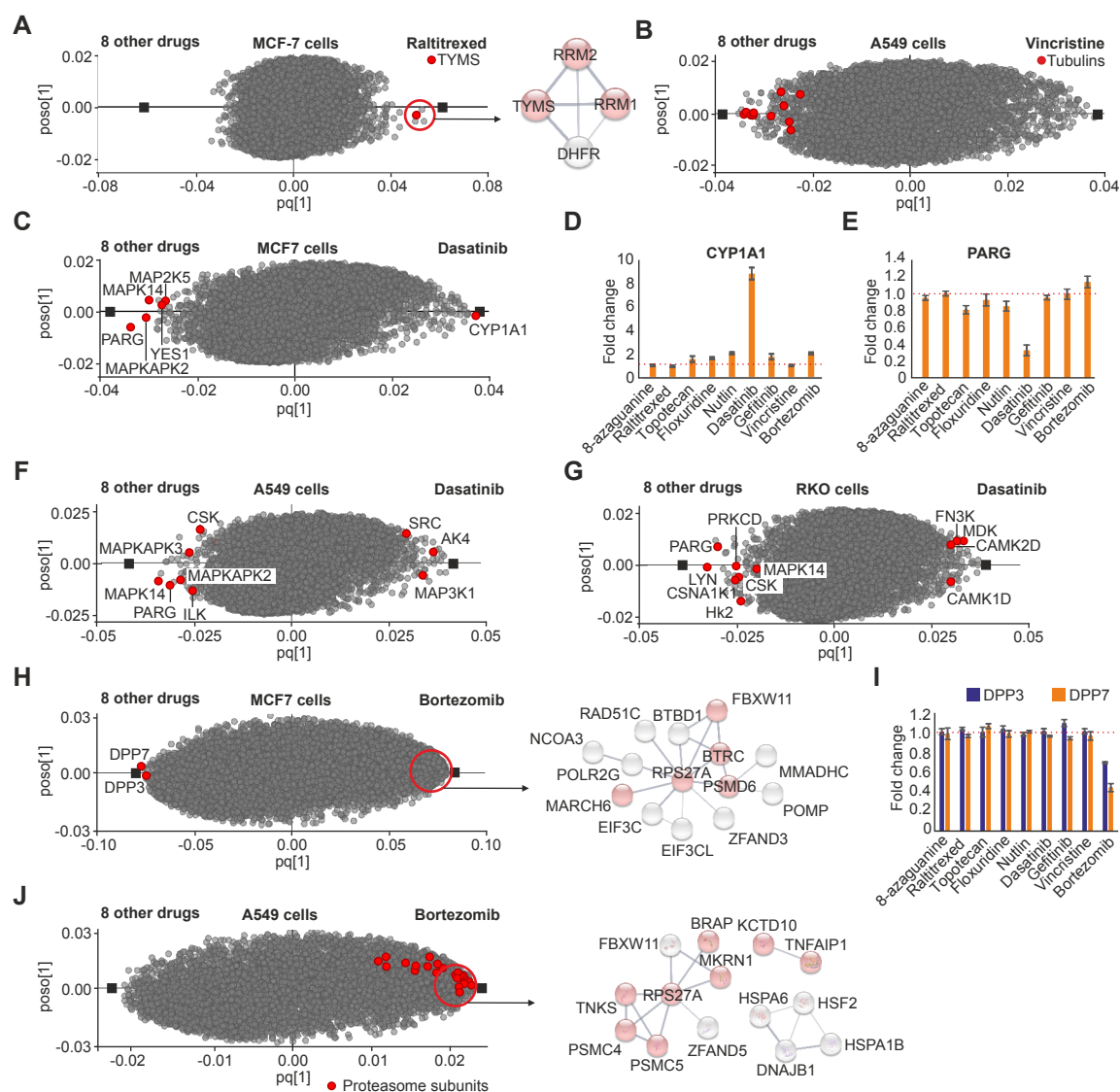


Figure 23. Miniaturized ProTargetMiner captures the targets of compounds in single cell lines with 8 contrasting compounds. OPLS-DA loadings for **A)** raltitrexed in MCF-7 cells and **B)** vincristine in A549 cells. The 4 top up-regulated proteins in OPLS-DA loading map to KEGG pathway pyrimidine metabolism (3 proteins, $p = 0.0003$). The targets are highlighted. **C)** OPLS-DA loading for dasatinib shows multiple kinase targets on top rankings in MCF-7 cells; CYP1A1 and PARG as the top most specifically up- and down-regulated proteins, respectively. The expression pattern of **D)** CYP1A1 and **E)** PARG in the panel of compounds in MCF-7 cells. **F-G)** The identification of multiple targets for dasatinib in A549 and RKO cells. **H)** Bortezomib targets DPP7 and DPP3 in MCF-7 cells, but still induces the expression of proteins involved in ubiquitination as shown in the pathway (30 top proteins, protein ubiquitination pathway proteins in red). **I)** The specific down-regulation of DPP7 and DPP3 in response to bortezomib in comparison with other compounds in MCF-7 cells. **J)** Up-regulation of protein ubiquitination pathway in response to bortezomib in A549 cells and the corresponding pathway analysis of 30 top proteins (protein ubiquitination pathway proteins in red). Disconnected proteins not involved in the pathway has been omitted.

4.2.8 Merging deep datasets to obtain a general picture of drug targets and MOA

Next we merged the deep datasets in three cell lines (6496 proteins) to evaluate the performance of OPLS-DA in drug target and MOA characterization. Of course, merging would offset the

cell-specific targets and MOA, but would provide a general cell-unspecific MOA for compounds. When all cell lines were used in the OPLS-DA model, TYMS could be identified on first position for raltitrexed and 16th for floxuridine (**Figure 24A**), showing that inclusion of two compounds with similar targets/MOA does not drastically hinder the target deconvolution in OPLS-DA. Furthermore, down-regulation of ribosomal proteins in response to floxuridine in **Figure 24A** is a feature of pyrimidine analogues as discussed before (Marin-Vicente, Lyutvinskiy et al. 2013) and observed in **Figure 19**. The OPLS-DA loading showed the general dasatinib targets and topoisomerase 1 (TOP1) as the target of topotecan among the top proteins (**Figure 24A**). Dasatinib also shows specific up-regulation of CYP51A1 and down-regulation of PARG.

OPLS-DA loading for vincristine in the merged dataset gave microtubule cytoskeleton organization (6 proteins, $p=0.0105$) (**Figure 24B**) and showed an enrichment of tubulins as specifically down-regulated proteins. The down-regulated proteins also mapped to a tight pathway representing “poly(A) RNA binding” (13 proteins, $p=7.47e-06$) and “ribonucleoprotein complex biogenesis” (6 proteins, $p=0.0444$). These pathways are in good agreement with previous research showing that vincristine affects RNA synthesis in human cells (Wagner and Roizman 1968). The final OPLS-DA model represents nutlin and could highlight the cognate target MDM2 at 9th position. While the top specifically up-regulated proteins ($n=30$) mapped to KEGG “p53 signaling pathway” (5 proteins, $p < 1.15e-05$), the specifically down-regulated proteins mapped to multiple pathways in DNA replication and repair (GO “DNA repair” (8 proteins, $p=0.0001$) is highlighted in blue in **Figure 24C**). In **Figure 24D**, we show a dip in the expression of multiple proteins involved in DNA repair in response to nutlin compared to other compounds in A549 cells (similar results were obtained in other cell lines). Nutlin activates the p53 pathway by antagonizing MDM2 (Vassilev, Vu et al. 2004) and is known to slow down DNA repair (Verma, Rigatti et al. 2010). Furthermore, DNA repair processes are now known to mediate the p53-dependent tumor suppression (Janic, Valente et al. 2018). Therefore, the subtle down-regulation of these proteins can explain why nutlin slows down DNA repair. The subtlety of this effect might be the reason why this effect has not been observed before. The expression of p53 and MDM2 is shown in **Figure 24E**. While some other compounds also activate p53, the up-regulation is significantly higher in response to nutlin, which directly affects p53.

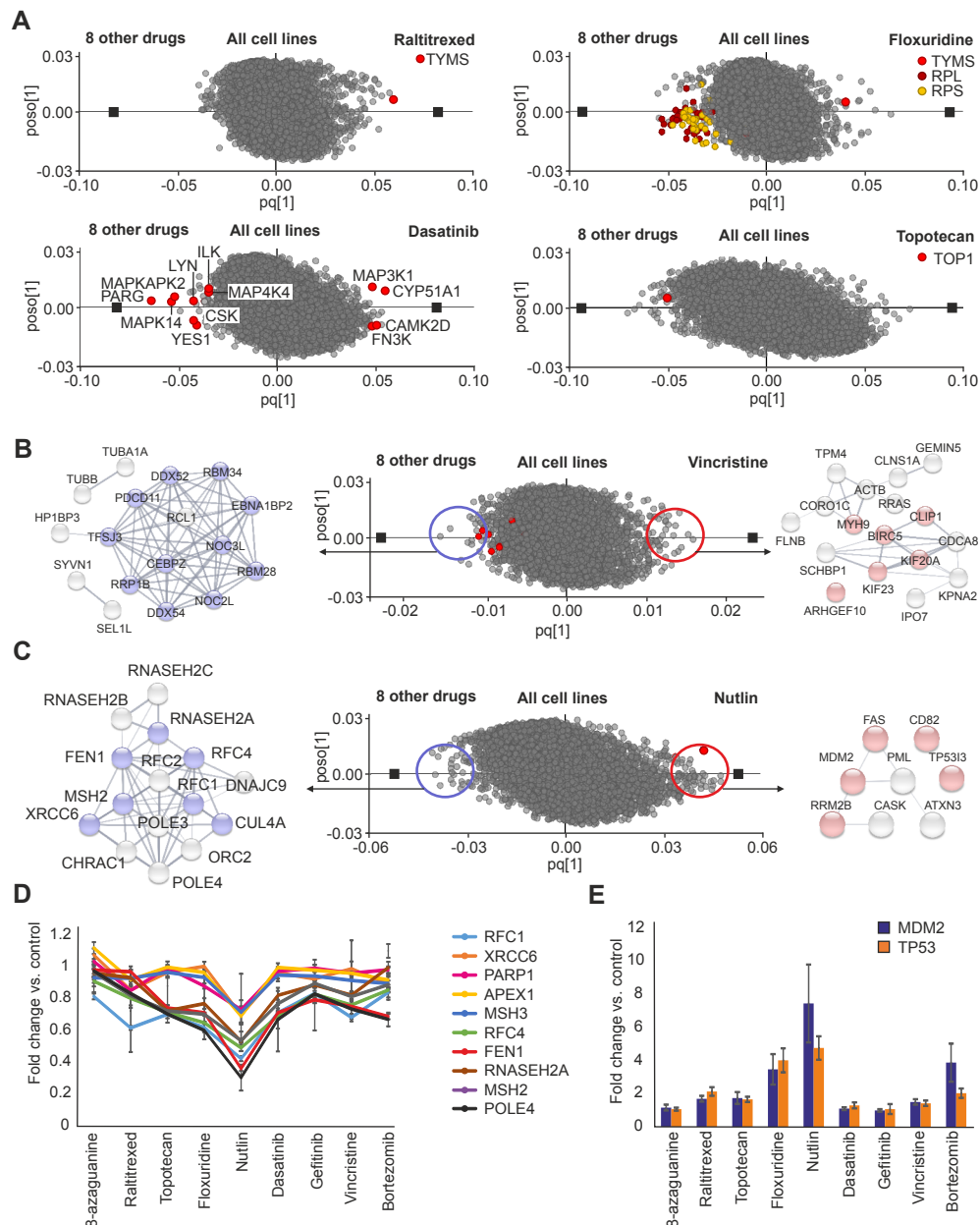


Figure 24. Merging deep datasets from three cell lines improves the target ranking in ProTargetMiner. A) OPLS-DA loadings for raltitrexed, floxuridine, dasatinib and topotecan. The known drug targets are highlighted. For floxuridine, specific down-regulation of ribosomal proteins is shown. **B)** OPLS-DA loading for vincristine showing the up-regulation of microtubule cytoskeleton organization (6 proteins, $p=0.0105$) and down-regulation of poly(A) RNA binding (13 proteins, $p=7.47e-06$). The known tubulin targets are shown with red circles. **C)** OPLS-DA loading for nutlin shows the up-regulation of p53 signaling pathway (5 proteins, $p=1.15e-05$), and specific down-regulation of proteins in DNA replication and repair (7 proteins, $p=0.0001$). **D)** The subtle down-regulation of top DNA repair proteins specifically down-regulated with nutlin in A549 cells. **E)** The expression of p53 and MDM2, the cognate target of nutlin in response to the panel of compounds in A549 cells. Panels D and E reused with permission from (Saei, Chernobrovkin et al. 2018).

4.2.9 ProTargetMiner R Shiny package as a public tool

The original and deep ProTargetMiner datasets can be easily extended with novel compounds in pursuit of their targets and MOA. This possibility was actually shown in the original ProTargetMiner dataset when 9 experiments were combined. Therefore, to make the resource directly available to the community, a Shiny package was written in R, which provides an

interface for integration of user data, matching the proteins with the datasets and OPLS-DA modeling in the desired cell line or in the merged dataset (schematics of Shiny package are shown in **Figure 25**). The OPLS-DA loading ranking of all proteins for the submitted compound can be extracted. In summary, the user will obtain the proteome signature of a compound of interest at LC50 concentration in 48h in any of the above cell lines (or all of them for obtaining more power) and feed the gene names and regulations in three replicates to the Shiny package. The package output will be an OPLS-DA model contrasting the given compound against the other proteome signatures in that cell line, from which a ranking of specificity for proteins can be extracted. The state-of-the-art depth of our dataset would accommodate the majority of proteome signatures.

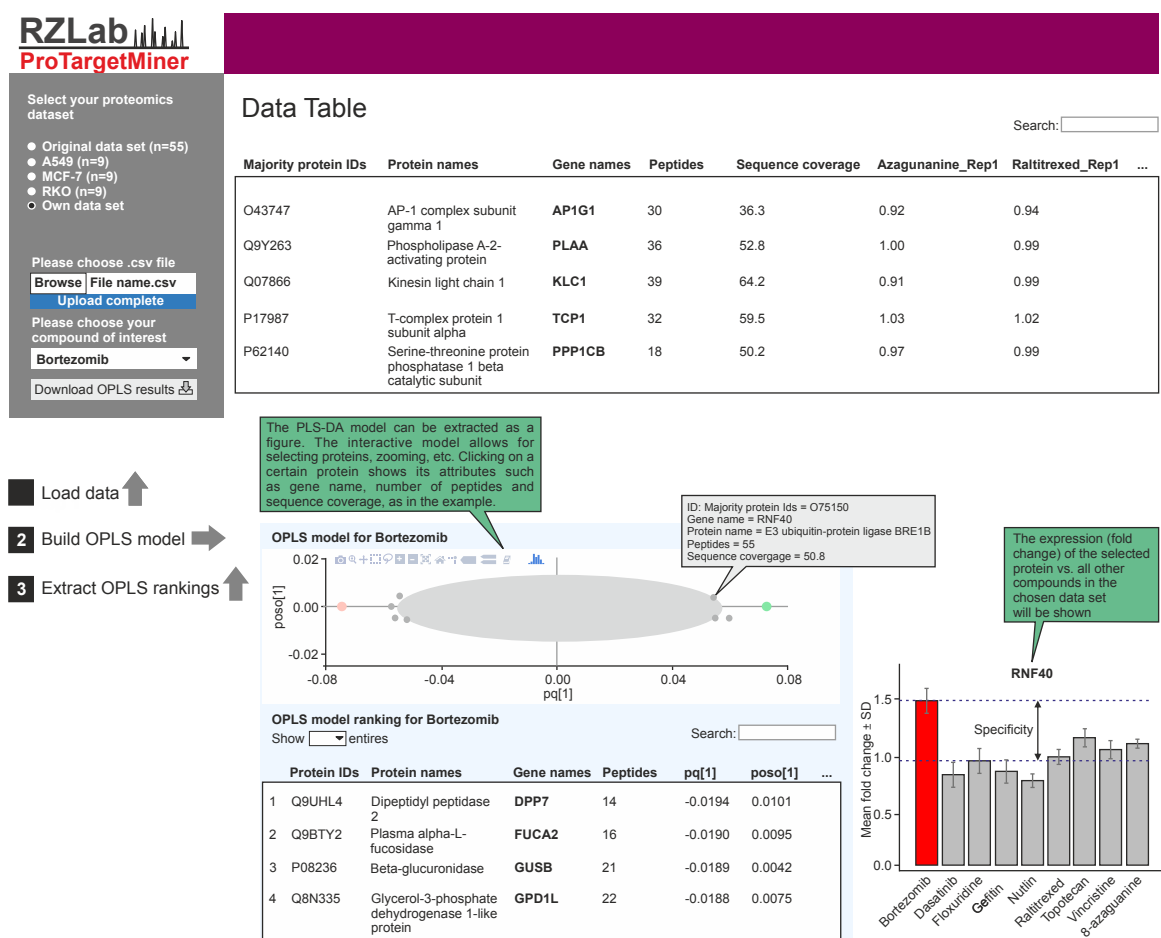


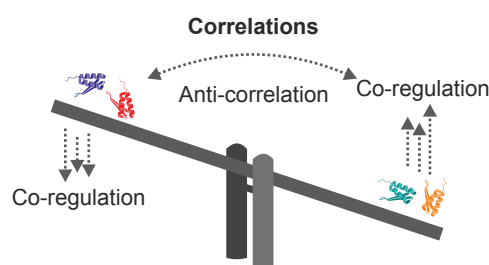
Figure 25. The ProTargetMiner R Shiny package for drug target and MOA deconvolution. In summary, the user could obtain the proteome signature of a compound of interest at LC50 concentration in 48h in any of the above cell lines (or all of them for obtaining more power) and feed the gene names and regulations in three replicates to the Shiny package. The package output will be an OPLS-DA model contrasting the given compound against the 9 diverse proteome signatures in that cell line (or 55 in the original set), from which a ranking of specificity for proteins can be extracted. Clicking on the interactive PLS-DA plot gives the attributes of that protein, e.g. name, number of peptides and sequence coverage, and will show the expression of that protein in the relevant dataset compared to the other perturbations.

4.2.10 The first human protein correlation database uncovers unexpected complexity in protein regulation

Due to the resource demanding nature of protein synthesis, cells spatiotemporally control protein expression (Scott and Pawson 2009). Thus cells reinforce the coordinated expression of genes in the same protein complex or pathway (DeRisi, Iyer et al. 1997). Since such co-expression can indicate functional relationship, “Guilt by association” (Stuart, Segal et al. 2003) approach has been used for characterization of protein function by analyzing co-expression (Hughes, Marton et al. 2000). Such databases have so far been based on transcripts expression (Stuart, Segal et al. 2003; Zuberi, Franz et al. 2013; van Dam, Craig et al. 2014), while it is known that proteomics is more accurate in capturing co-regulation (Wang, Ma et al. 2017). To the best of our knowledge, no large-scale proteomics-based protein correlation database exists.

As the 55 compounds used in this study perturbed the majority of the proteome, ProTargetMiner presented the opportunity to build the first human protein pairwise correlation database solely based on proteomics data (**Figure 26**).

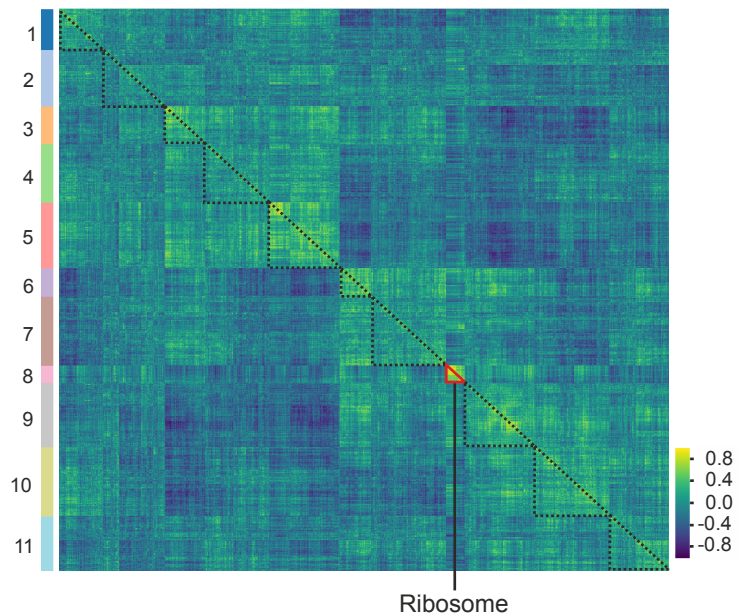
Figure 26. Analyzing pair-wise protein correlation in ProTargetMiner. The proteome perturbation by 55 anticancer compounds will uncover protein co-regulation and anti-correlation networks and facilitate the functional annotation of uncharacterized proteins. Figure adapted with permission from (Saei, Chernobrovkin et al. 2018).



Moreover, such a database can provide information on the complexity of protein regulation. Furthermore, the pair-wise anti-correlation of protein abundances is usually neglected in such databases. This is while there is a lower chance for negative correlations (than positive correlations) to result from technically-induced artifacts (Lee, Hsu et al. 2004). Such protein anti-correlation can potentially represent opposing biological processes, where they are wired for example in active transcriptional repression, transcriptional activation or even canceling of such events (Struhl 1999). Protein pairwise anti-correlation was therefore included in the current study, and physical interpretation of anti-correlation is an open opportunity.

After refining, removal of the batch effects and data filtering, a 4212 x 4212 correlation matrix was made for the original ProTargetMiner dataset (**Figure 27**). At least 11 clusters could be recognized in the matrix. For instance, the 129 proteins present in cluster 8 mapped to ribosomal proteins (n=72) and ribosome biogenesis (n=10). From the 17,740,944 correlation pairs in the matrix, a high-confidence set (FDR<0.001) of 103,928 positively and 51,137 negatively correlating protein pairs were chosen. The lower frequency of negative in comparison with positive correlations was expected and was consistent with a previous study (Lee, Hsu et al. 2004).

Figure 27. Analyzing pairwise protein correlations in ProTargetMiner across 55 perturbations. A correlation matrix of 4,212 proteins. Vertical axis: broken-down to 11 clusters. Note the high abundance of anti-correlations (blue color) and the existence of clear cut corners in the matrix. Figure reused with permission from (Saei, Chernobrovkin et al. 2018).



4.2.11 The database pulls out dense regions of protein interaction networks

The 10 top correlating pairs ($r > 0.98$) mostly mapped to dense interaction networks and protein superfamilies, e.g., protein complexes, such as MCM, condensin, ribosome, chaperonin-containing T and mitochondrial respiratory chain as well as tubulins. In **Figure 28** we mapped the top 2500 co-regulated pairs, which clearly form functionally coherent groups of genes.

For making the database available for network visualization and pathway analysis, we mapped the data to StringDB entities and created a set of external payload data. These payload data can thus be uploaded to StringDB as nodes and edges. Known StringDB interactions and ProTargetMiner correlations can be visualized simultaneously using this approach.

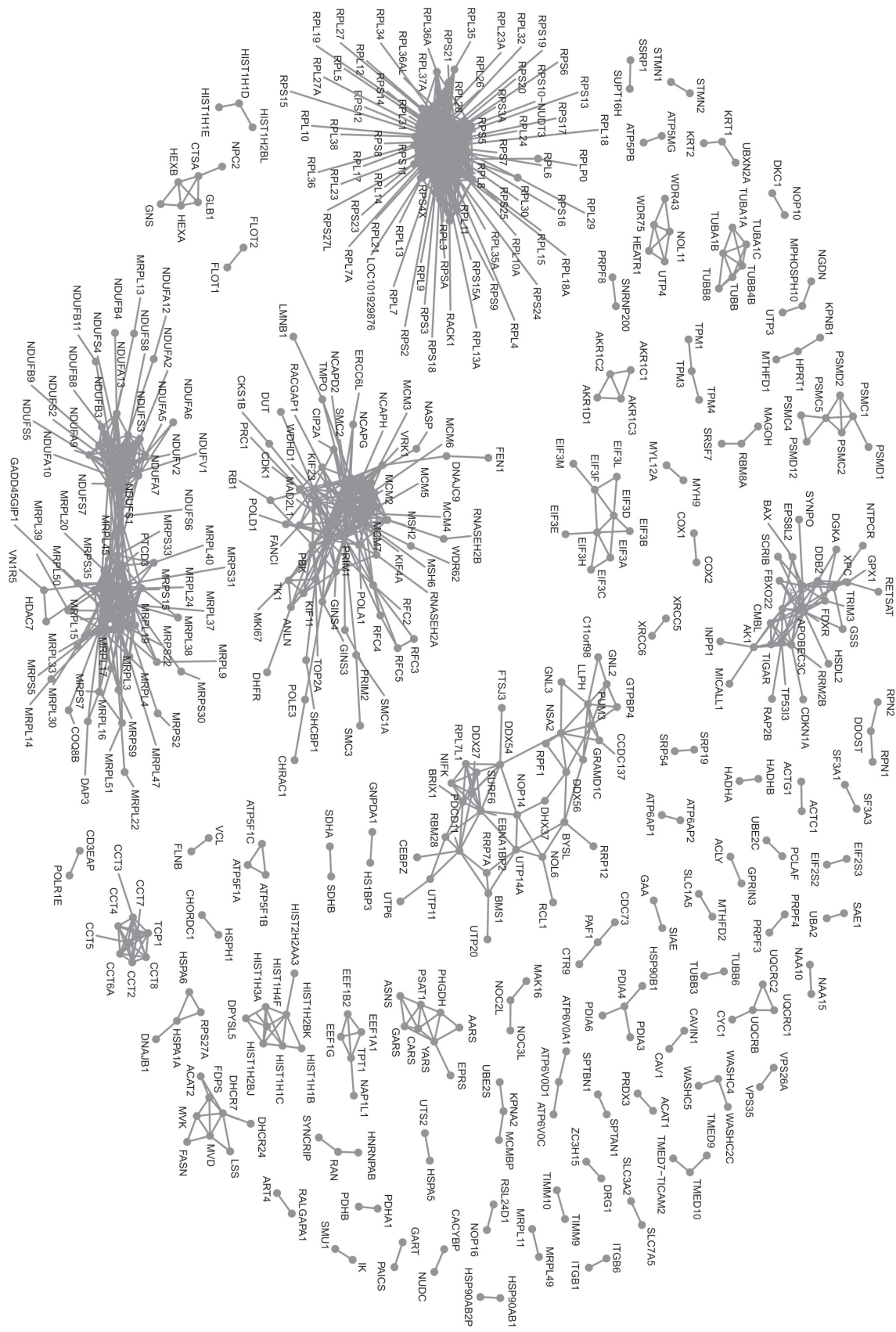


Figure 28. Top co-regulating proteins in ProTargetMiner. The top 2500 correlating pairs map to dense protein interaction networks, indirectly validating the approach. Figure reused with permission from (Saei, Chernobrovkin et al. 2018).

4.2.12 Protein anti-correlation, true or false?

TIGAR (TP53-inducible glycolysis and apoptosis regulator) was one of the top and enriched anti-correlating proteins. Therefore, we set out to examine its expression pattern against its most anti-correlating pair NCAPG (condensin complex subunit 3) ($r=-0.84$) (**Figure 29A**). A pathway analysis was performed for TIGAR and NCAPG, along with 5 most correlating proteins for each (**Figure 29B**). While 3 proteins from the TIGAR group gave p53 signaling pathway ($p < 0.002$), 5 proteins from the NCAPG group mapped to condensin complex and 6 to cell division ($p < 0.001$). This result is in good agreement with the observation that p53 inhibits entry into mitosis upon a blockage of DNA synthesis. Furthermore, TIGAR is known to protect from accumulation of genomic damage in a p53-dependent manner (Bensaad, Tsuruta et al. 2006). Interestingly, one of the proteins strongly co-regulating with TIGAR ($r=0.89$) was CMBL, but this protein was not annotated in the p53 signaling pathway in StringDB. However, CMBL has been identified as a p53-inducible protein in a microarray screen (Jiang, Kon et al. 2015). These findings confirm the predictive power of the database and serve as an *in silico* proof-of-concept.

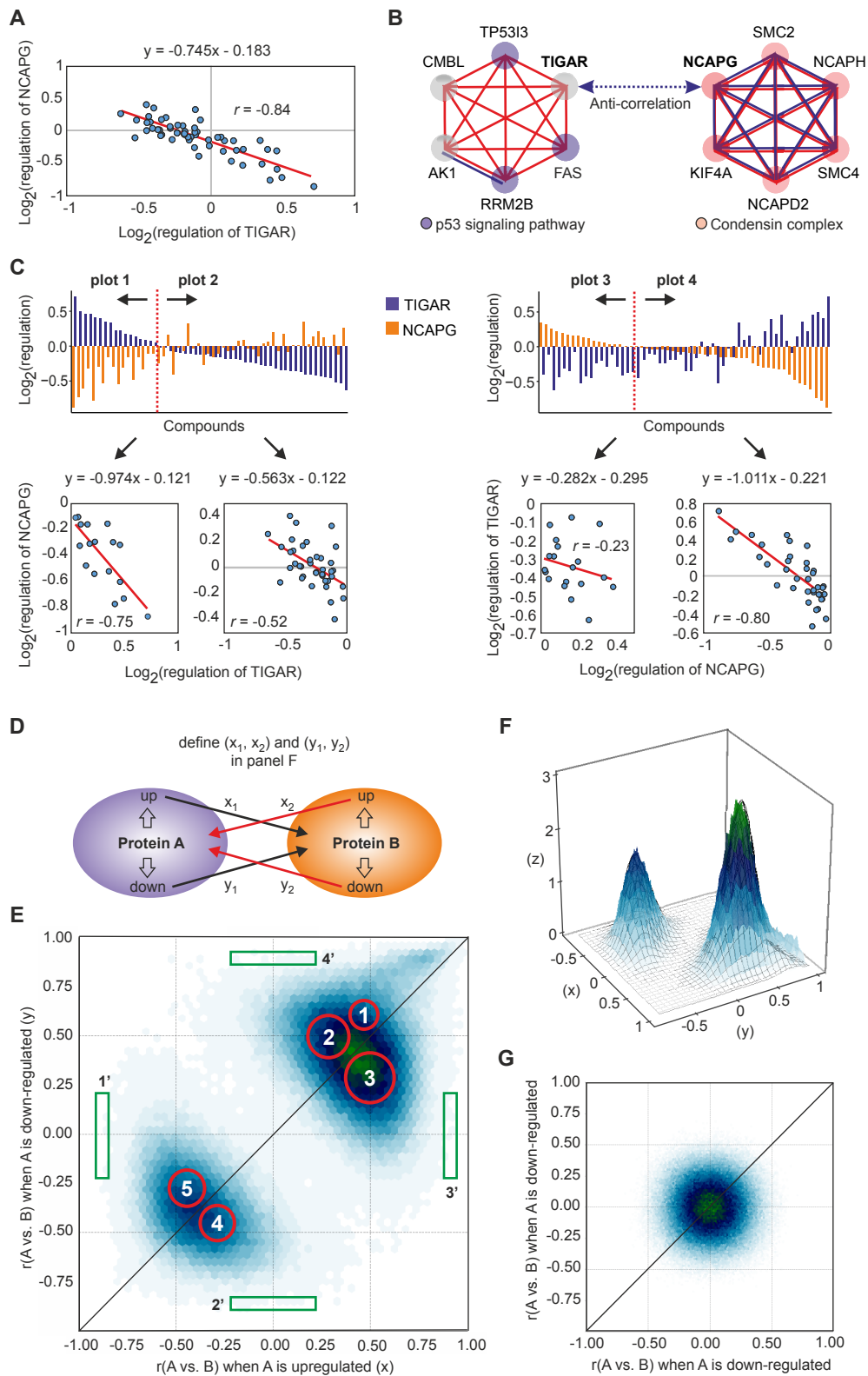


Figure 29. Comprehensive analysis of protein pairwise correlation considering up- and down-regulation. **A)** The anti-correlation of TIGAR with NCAPG. **B)** All the top proteins which are co-regulating with TIGAR anti-correlate with those co-regulating with NCAPG. The blue links reflects information available in StringDB and the red links reflects newly established links in ProTargetMiner. **C)** The anti-correlation of up or down-regulated TIGAR vs. NCAPG (plots 1 and 2) and up- or down-regulated NCAPG vs. TIGAR (plots 3 and 4). **D)** Any protein

pairs give two sets of correlations separately considering their up and down-regulation status. **E)** Map of pairwise correlations between proteins A and B, separately for up- and down-regulation states of protein A. The red circles reflect the positions (center) and volumes of 5 fitted 2D Gaussian distributions. The green boxes reflect protein pairs that only have strong correlations in one direction. **F)** 2D Gaussian fitting performs best with 5 components. **G)** Similar analysis as in panel F using scrambled protein abundances. Caption adapted from **Paper II** (Saei, Chernobrovkin et al. 2018). Figure adapted with permission from (Saei, Chernobrovkin et al. 2018).

4.2.13 Adding complexity to protein regulation

However, this was not the whole story. While the overall TIGAR vs. NCAPG anti-correlation coefficient was significant ($r=-0.84$) (**Figure 29A**), the slope was -0.75 and not -1 , signifying an underlying complexity. When TIGAR's regulation was above-average, it anti-correlated more strongly with NCAPG ($r=-0.75$, plot 1 on **Figure 29C**) than when it had a regulation below average ($r=-0.52$, plot 2). Also, in the first scenario, the slope was close to -0.97 , as if in perfect association; but with TIGAR down-regulated, the slope reduced to -0.56 . On the contrary, when NCAPG had a regulation above average, the anti-correlation with TIGAR was weak ($r=-0.23$, plot 3), whereas when down-regulated, it had a strong anti-correlation with TIGAR ($r=-0.80$, slope -1.01 , plot 4).

These findings revealed an expected but neglected bimodality in anti-correlation. This prompted us to calculate two correlation coefficients for every protein pair (**Figure 29D**). Therefore, we calculated two separate correlations for above and below the median abundance of protein A (in the pair A vs. B). The heatmap of these correlations on a 2D plot with a cutoff of $|r|>0.54$ (**Figure 29E**) unexpectedly revealed five dense areas (circles 1-5 in **Figure 29E**), which were confirmed by fitting symmetric 2D Gaussian distributions (**Figure 29F**). The positions of these Gaussians are shown as the centers of red circles in **Figure 29E**, and their volumes (frequency of pairs) correspond to circle areas. Scrambled protein abundances gave a single central spot (**Figure 29G**) rather than an elliptic symmetry, confirming that spots 2-4 are not artifacts.

Entity 1 or “highly co-regulated” (average $r=0.52$) contains 12% of the total number of (anti-) correlating protein pairs, and the 300 top protein pairs ($r \geq 0.90$) map to ribosome, MCM, condensin, chaperonin-containing T, NADH dehydrogenases and pyruvate dehydrogenase complexes as well as tubulin superfamily, some integrins and spectrins. The center of entity 1 is above the diagonal, indicating a higher co-regulation when the protein pairs are down-regulated than up-regulated. This asymmetry can be explained by the fact that down-regulation can reach zero, whereas up-regulation cannot reach infinity. Another biological explanation is that there is usually no natural stopper for up-regulation, while degradation is under a higher level of control, and slows down or stops with only strongly bound stoichiometric complexes remaining in the system.

In general, positive and negative components account for 55% and 33% of total protein pairs, respectively. The protein pairs in entity 2 or “co-down-regulated” show better co-regulation when down-regulated (average $r=0.5$ vs. average $r=0.3$). Entity 2 harbors 23% of correlating pairs. Entity 3 or “co-up-regulated” proteins is composed of proteins correlating more when

up-regulated (average $r=0.5$) than when down-regulated (average $r=0.28$). Entity 3 possesses 32% of correlating pairs.

The protein pairs in entity 4 or “anti-correlating” exhibit a higher anti-correlation when protein A regulation is below average (average $r=-0.29$ vs. $r=-0.46$), while entity 5 or “negatively regulated” displays higher anti-correlation when protein A is up-regulated (average $r=-0.45$ vs. $r=-0.29$). Entities 4 and 5 have a comparable frequency of protein pairs (16% and 17%, respectively). These findings are discussed in more detail in **Paper II** (Saei, Chernobrovkin et al. 2018).

In **Figure 29E**, a number of protein pairs show strong anti-correlation in one and around zero correlation in the other direction. For example, in region 1', out of the 50 top pairs, 28 proteins (both “A” and “B” type) map to negative regulation of cellular processes ($p < 0.03$). In region 2', 11 “A” proteins from 50 pairs give “cell cycle” ($p < 0.005$), while 5 “B” proteins gave “nucleotide excision repair” ($p < 2.1E-06$) and 7 “B” proteins – mapped to “cellular response to DNA damage stimulus” ($p < 0.022$), exposing the former and the two latter pathways as potentially opposing ones (Hustedt and Durocher 2017). The top 50 pairs in region 4' give “ribosome” (16 proteins, $p < 2.3E-15$) and “ribosome biogenesis” pathways (14 proteins, $p < 1.0E-12$).

4.2.14 Anti-correlation provides further information on protein regulation

The anti-correlation data can potentially furnish further information on protein regulation. For instance, TRIM28 when regulated above average, strongly ($r < -0.61$) anti-correlates with 34 proteins mapping to focal adhesion (7 proteins, $p < 8.1E-06$). Interestingly, a recent study has directly linked TRIM28 to cell adhesion (Klimczak, Czerwińska et al. 2017). Furthermore, the other enriched pathway was that of ErbB signaling (4 proteins, $p < 0.001$). TRIM28 is known to directly interact with ErbB4 and inhibit its transcriptional activity (Gilmore-Hebert, Ramabhadran et al. 2010).

4.2.15 Untouchable proteome reflects essential cell functions

We further analyzed the data to investigate if there are proteins with steady expression in response to the compound panel, designating the untouchable or core proteome (**Figure 30**).

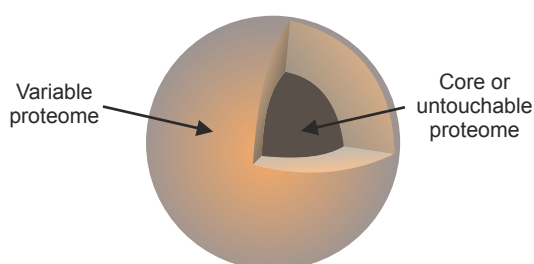


Figure 30. The core and variable proteomes. The untouchable or core proteome is defined as a set of proteins with steady expression across the treatments. Figure reused with permission from (Saei, Chernobrovkin et al. 2018).

To identify core proteins, the standard deviation of expression for all the proteins was calculated in the original ProTargetMiner dataset. The 100 most stably expressed proteins mapped to proteasome (9 proteins, $p < 2E-10$), spliceosome (8 proteins, $p < 3E-05$), and mRNA surveillance pathway (6 proteins, $p < 0.001$). These pathways might represent the core cellular functionalities. The fact that proteins from the proteasome and spliceosome were among the enriched pathways is remarkable, as we had a number of proteasome inhibitors and proteins targeting components of the spliceosome in the library. This shows the great level of control and buffering of expression in these complexes and networks.

4.2.16 Untouchable proteome unravels house-keeping proteins

In comparative proteomics and other molecular biology techniques, and in particular Western blots, specific proteins are used as internal references, assuming that they have a steady expression. Examples of such housekeeping proteins (HKPs) include GAPDH, β -tubulin and β -actin. Such proteins have been selected because of their high mRNA expression levels and consistent expression in different cells and tissues. However, stable expression under treatment conditions has not been studied for most of these proteins. Indeed, in practice, a number of these HKPs show inconsistent expression levels under different treatments (Vigelsø, Dybboe et al. 2014; Janes 2015; Lee, Jo et al. 2016). For example, GAPDH, one of the most extensively used HKPs showed variable expression in our compound panel especially for bortezomib (**Figure 31A**). These inconsistencies can lead to inaccurate estimation of protein abundance, as reported before (Dittmer and Dittmer 2006; Li and Shen 2013; Collins, An et al. 2015; Gough 2015). Therefore, characterization of ideal HKPs is an active area of research (Eisenberg and Levanon 2013).

ProTargetMiner presented an opportunity to identify reliable HKPs. We used standard deviation across all the treatments for rank ordering of most stably expressed proteins (**Figure 31B**). The exemplary expression profiles of two proteins with most steady and variable expression are depicted in **Figure 31C**. More analysis will be done to assess the reliability of potential HKPs (with regards to abundance, number of peptides, presence in various cell lines and tissues, availability of reliable antibodies, etc.) and presenting a shortlist of top candidates.

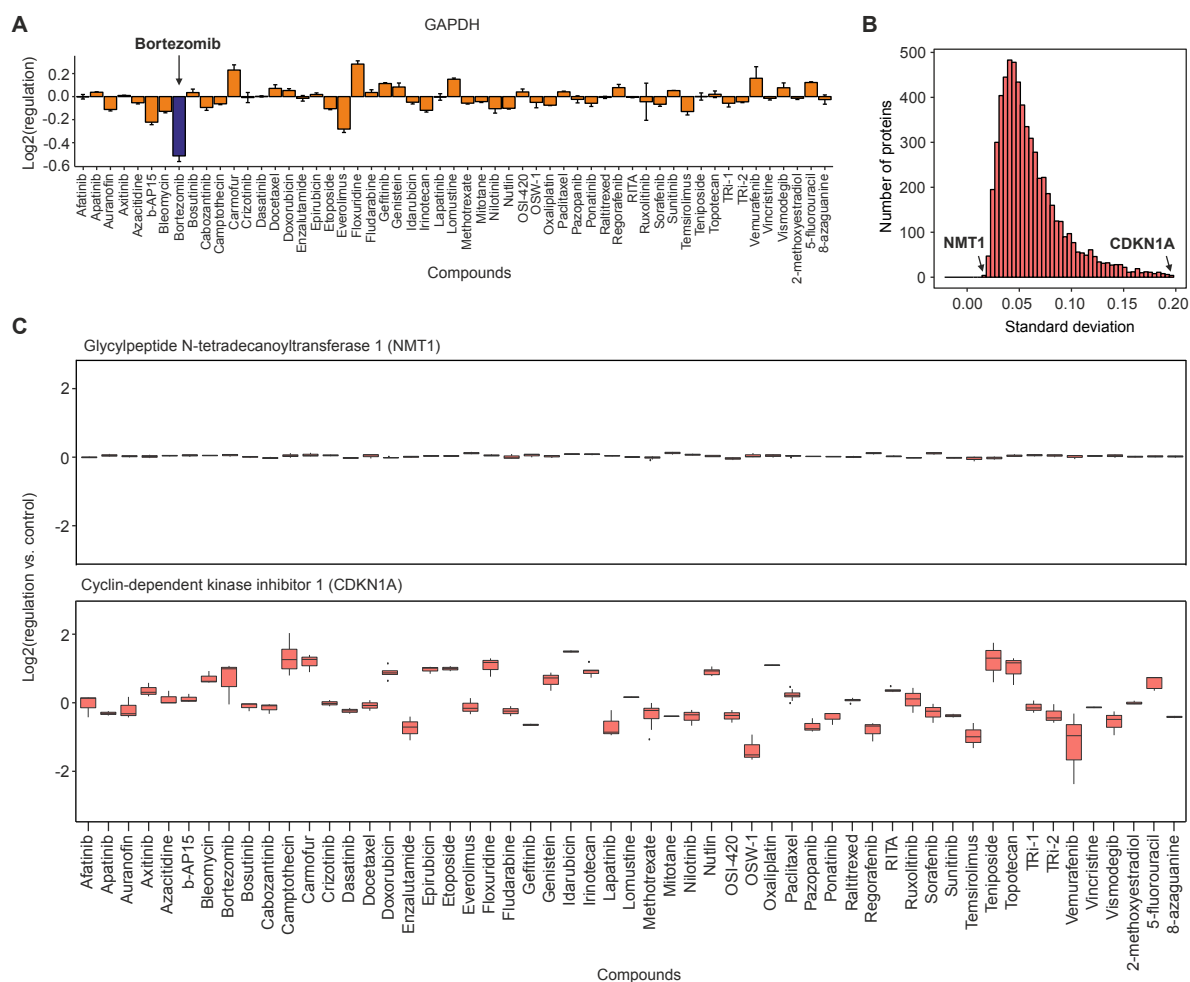


Figure 31. In search for more reliable housekeeping proteins. A) GAPDH expression across the treatments. **B)** Distribution of standard deviations of 6,032 proteins across the compounds panel. **C)** Typical proteins exhibiting highest and lowest stability across the compound panel. Data are represented as mean±SD. Figure adapted with permission from (Saei, Chernobrovkin et al. 2018).

Conclusions

In summary, ProTargetMiner can be considered as a resource of proteome signatures for a panel of FDA-approved compounds. ProTargetMiner R Shiny package serves as a platform for deconvolution of targets, MOA, resistance factors and other specific effects for novel compounds. The specificity concept enabled by OPLS-DA modeling can be a viable approach in characterization of subtle but biologically meaningful phenomena in target deconvolution in proteomics and other big data, as well as nuances of differences in protein expression potentially involved in tissue diversification (Silva and Vogel 2016). Furthermore, the analysis of ProTargetMiner data provided insight on the complexity of protein regulation. The correlation database can be used for characterization of proteins with unknown functions. Finally, the most stably expressed proteins in the ProTargetMiner data can be used as HKPs in molecular biology experiments.

4.3 PAPER III. COMPREHENSIVE CHEMICAL PROTEOMICS PREDICTS DRUG TARGETS: AURANOFIN AS EXAMPLE

In **Paper III**, we used a series of chemical proteomics tools –namely TPP, FITeXP and multiplexed redox proteomics for prediction of targets and characterization of mechanism space for auranofin. Auranofin was approved by FDA in 1985 as an antirheumatic agent, but recent studies have shown its potent antitumor activity (Rios Perez, Roife et al. 2016) and is therefore being evaluated in clinical trials against ovarian cancer (NCT03456700), chronic lymphocytic leukemia (NCT03456700) and lung cancer (NCT03456700) (Roder and Thomson 2015). Auranofin is known to bind to the selenocysteine- (Sec-) active site in thioredoxin reductase 1 (TXNRD1). However, surprisingly several different mechanisms have been proposed for auranofin and controversies exist regarding its cellular targets and potentially important off-targets remain unexplored. Full characterization of auranofin mechanistic space can help adjust and justify its repurposing context.

4.3.1 Unbiased prediction of auranofin targets using a combination of chemical proteomics tools

Since each chemical proteomics methods yields both false positives and false negatives, multiple tools must be explored and combined to get a reliable list of target proteins, each supported by one or more technique. Therefore, a series of experiments were performed for auranofin: Two TR-TPP experiments in cells (2 and 3 μ M in HCT116 cells for 2h), a TR-TPP experiment in lysate (500 nM in HCT116 cells), a FITeXP experiment (HCT116, RKO and A375 cell lines treated for 48h with LC50 concentrations; methotrexate, OSW-1 and paclitaxel were used as contrasting compounds) and a redox experiment (3 μ M in HCT116 cells for 2h). An overview of the results is shown in **Figure 32A-C**.

For *de novo* and unbiased identification and validation of auranofin targets, we ranked the proteins in each chemical proteomics experiment. In TR-TPP experiment in cells (3 μ M auranofin) and lysate, the proteins were ranked based on the ΔT_m between auranofin and control treatments, and in FITeXP, absolute magnitude of consistent regulation in three cell lines was used for ranking of proteins. In redox proteomics, the proteins were ranked based on the difference in oxidation level and p values of auranofin vs. vehicle-treated samples. The rankings from all mentioned experiments were combined to shortlist candidate drug targets (**Figure 32D**). Proteins with the lowest cumulative ranking are therefore the best drug target candidates.

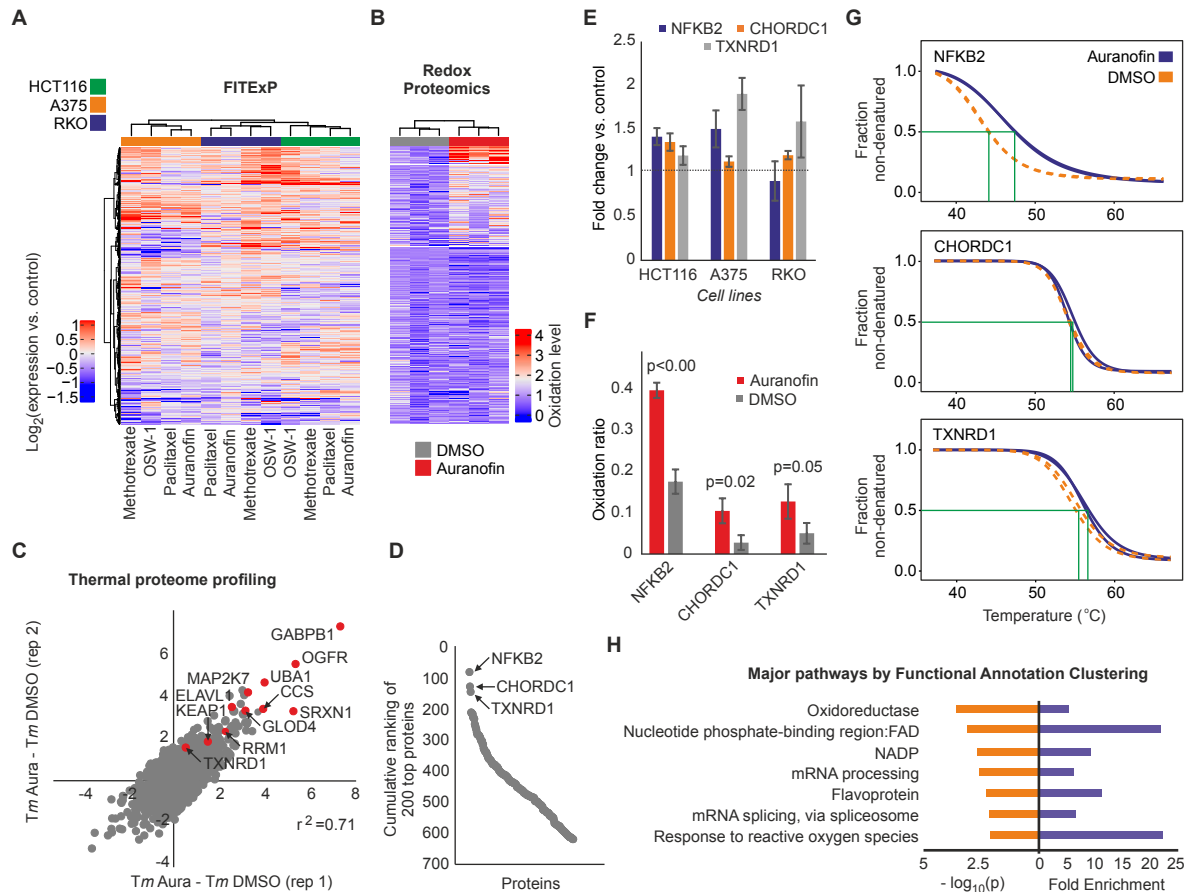


Figure 32. Unbiased prediction of auranofin targets by chemical proteomics tools. **A)** Overview of FITExP data for auranofin and three contrasting compounds from three cell lines. **B)** Redox proteomics revealed an increase in the oxidation level of many peptides and a decrease in some upon treatment with auranofin. **C)** TPP data for 1887 proteins with ≤ 1 °C difference between the two replicates in 3 μ M auranofin treatment of HCT116 cells shows stabilization of several proteins. Some of these proteins, including GABPB1, RRM1 and SRXN1 are known to be regulated by thioredoxin system. **D)** The cumulative ranking of target candidates in four different types of analysis (FITExP, TR-TPP in cells and lysate, as well as redox proteomics). Proteins with the lowest overall rankings are top candidate targets (TXNRD1 is the cognate target). **E-G)** Changes in the expression, oxidation of top constituent peptides and stability of three top target proteins in cells (none of these proteins changed their stability in cell lysate). **H)** Top 15 proteins were analyzed with Functional Annotation Clustering tool in DAVID. Top pathways enriched with minimal redundancy (fold enrichment >5 and $p < 0.01$), representing the dominant mechanisms for auranofin, are shown.

The expression vs. control in three cell lines, the oxidation level of the most oxidized peptide as well as the melting of three top proteins in cells in response to auranofin are shown in **Figure 32E-G**. Using this unbiased approach, TXNRD1 was on found on the 3rd position. Since TXNRD1 is the cognate target for auranofin, the predictive power of chemical proteomics is confirmed. This was an impressive result, as the lowest ranking for auranofin in any method was in FITExP, where TXNRD1 was ranked 15th. TXNRD1 had a cell line dependent up-regulation. CDYENVPTTVFTPLEYGACGLSEEK was the top oxidized peptide for TXNRD1 with a ratio of 2.53 ($p = 0.051$) auranofin:DMSO. TXNRD1 was slightly but reproducibly stabilized in cells (1.14 °C with 3 μ M and 1.02 °C with 2 μ M auranofin) and not in the lysate, perhaps indicating a chemical transformation prior to attachment to TXNRD1. Indeed, it is known that auranofin binds to TXNRD1 selenocysteine and subsequently gold is transferred to the

redox active cysteine couples inside the protein (Angelucci, Sayed et al. 2009). The slight change in TXNRD1 stability though might be linked to auranofin binding to the penultimate amino acid, which may not induce a significant stability change in the 71 kDa macromolecule.

The other two top shortlisted proteins were not known as auranofin targets. The top protein was Nuclear factor NF-kappa-B p100 subunit (NF-kB2), the inhibition of which is responsible for the anti-inflammatory effects of auranofin (Jeon, Jeong et al. 2000; Youn, Lee et al. 2006). Although the direct binding of auranofin to NF-kB2 has not been reported in literature, the appearance of this protein in the list cannot be by chance. Therefore, we will examine if NF-kB2 is a direct target of auranofin by follow-up experiments. CHORDC1 or cysteine and histidine-rich domain-containing protein ranked second before TXNRD1 and is known to be involved in HSP90 chaperone complex and stress response (Ferretti, Palumbo et al. 2010). CHORDC1 is rich in cysteines, but it is an unknown protein in the context of thiol-reactive metal compounds.

4.3.2 Pathway analysis reveals the major auranofin mechanism

To lower the risk of false negatives stemming from taking the interception of top results from different analyses, we selected fifteen top proteins from each of the four methods and subjected the combined list to pathway analysis by Functional Annotation Clustering tool in DAVID. This new tool helps reduce the number of redundant biological pathways by grouping similar annotations together (Huang, Sherman et al. 2008). The top enriched pathways by p value were “oxidoreductase”, “nucleotide phosphate binding region (NAD)” and “NADP”, in line with the MOA of auranofin (**Figure 32H**).

We looked further in the expression data to decipher which proteins are involved in oxidoreductase pathways. The expression data for auranofin were clustered (**Figure 33A**). The clustering grouped the data first based on the three cell lines. Interestingly, the cluster 6 consisted of 15 proteins up-regulated in all three cell lines, demonstrating that FITeXP is an effective approach for discovery of proteins with consistent behavior in a panel of cell lines (Chernobrovkin, Marin-Vicente et al. 2015). These 15 proteins formed a tight group, irrespective of the chosen number of clusters (n=6-10) and mapped to “glutathione (GSH) metabolic process” (4 proteins, $p < 0.0004$) and “response to oxidative stress” (4 proteins, $p < 0.0005$). Although auranofin induced different proteome signatures in different cell lines, the up-regulation of GSH metabolism related proteins was noted in all cell lines. The induction of enzymes playing a role in GSH metabolism is consistent with the backup role of GSH in the absence of TXNRD1 activity (Du, Zhang et al. 2012). Furthermore, interestingly, most of the proteins in cluster 6 are Nrf2 target proteins (highlighted in red in **Figure 33A**). TXNRD1 inhibition is known to activate Nrf2 (Cebula, Schmidt et al. 2015), which in turn induces the expression of genes involved in the response to oxidative stress such as GCLM and GCLC (GSH synthesis enzymes) (Malhotra, Portales-Casamar et al. 2010). The biological pathways enriched for 30 top

up-regulated proteins from FITExP are shown in **Figure 33B**.

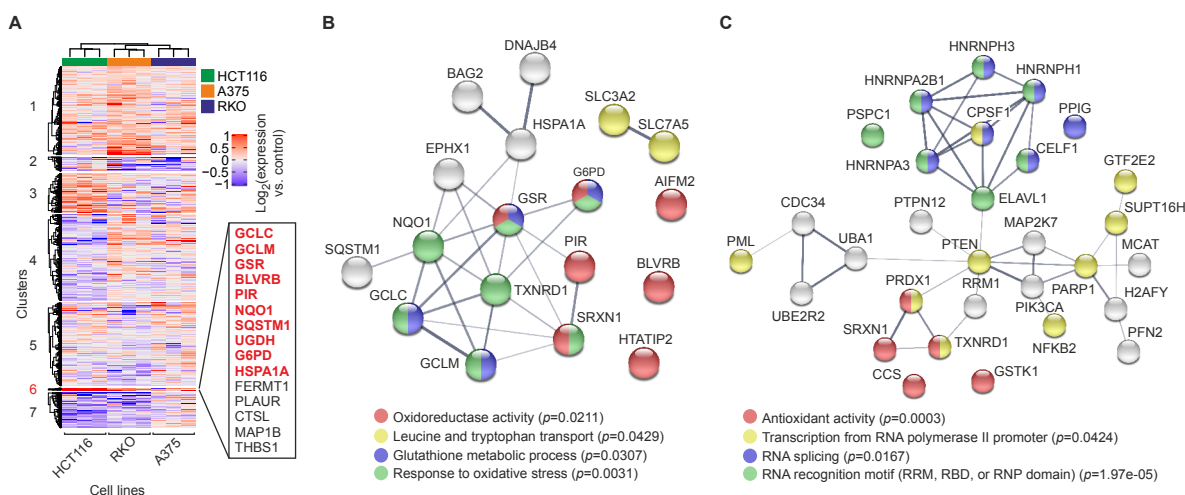


Figure 33. The over-representation of Nrf-2 targets and oxidoreductase pathways in response to auranofin. A) The clustering of auranofin expression in three cell line gave a tight cluster (6), consisting of 15 proteins up-regulated in all cell lines. The pathways enriched for top 30 proteins in FITExP (B) and proteins changing stability in TPP (C).

Next we analyzed the stability data in cells for finding proteins responding to auranofin treatment by a change in stability. We subjected the proteins (with at least three peptides) with ≥ 1.5 °C ΔT_m between auranofin and DMSO treatments in the TR-TPP with 3 μM auranofin to pathway analysis. These proteins mapped to “antioxidant activity” (**Figure 33C**). Therefore, both FITExP and TPP methods showed the perturbation of oxidoreductase pathways in cells treated with auranofin indicating that perturbation of oxidoreductases is the main MOA for auranofin.

4.3.3 Redox proteomics links cysteine oxidation and protein stability

Another hypothesis that we pursued in this study was if protein oxidation in the peptide level (measured with redox proteomics) can be linked to change in stability of the respective proteins (measured in TPP). Cellular thioredoxin system is tasked with reduction of specific substrate proteins (Arnér and Holmgren 2000), which in turn reduce other cellular proteins. Auranofin-mediated inhibition of this system could consequently lead to stabilization of these substrates and perhaps down-stream proteins. Therefore, the redox proteomics experiment was performed under similar conditions to TPP, treating HCT116 cells with 3 μM of auranofin for 2h. The higher oxidation of peptides in auranofin vs. DMSO treatment was obvious in the asymmetric volcano plot (**Figure 34A**) in favor of more oxidation. We could link the oxidation/reduction of 11 to increased or decreased thermal stability in TPP (shown in red in **Figure 34A**). Interestingly, the active sites of SRXN1 (Cys99) –which is the only cysteine residue in this protein, (Sunico, Sultan et al. 2016) and PRDX5 (Cys100) (Seo, Kang et al. 2000; Hall, Parsonage et al. 2010) were among these sites (**Figure 34B**). The melting curves of SRXN1

and PRDX5 along with two other representative proteins with simultaneous change in oxidation and stability is shown in **Figure 34C**. Therefore, a combination of TPP and redox proteomics can be used for discovery of potential redox regulatory switches in proteins, and extend the applications of both techniques.

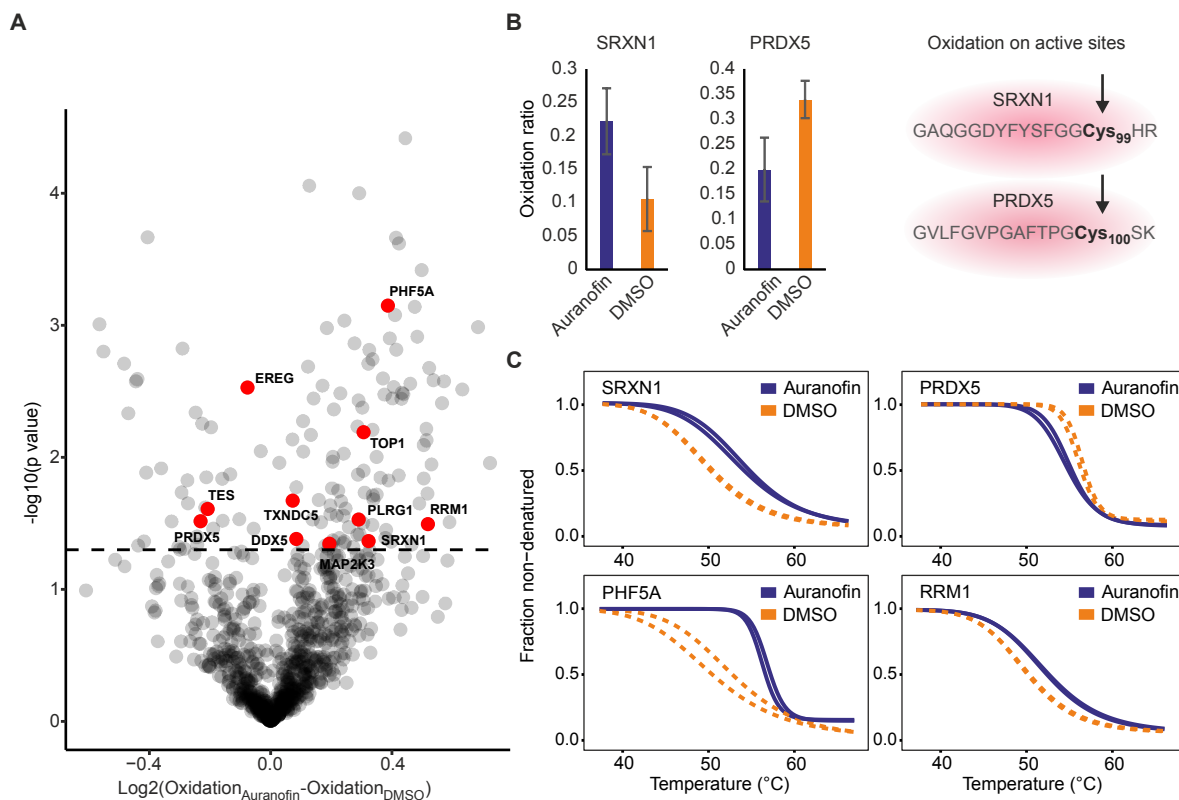


Figure 34. Peptide oxidation state can be potentially linked to protein stability. **A)** The oxidation rate of cellular peptides after treatment with 3 μ M auranofin for 2h. The highlighted peptides with significant change in oxidation state are those for which the respective proteins change stability in TPP. **B)** Significant oxidation and reduction of SRXN1 and PRDX5 active sites in redox proteomics. **C)** Change in the stability of SRXN1 and PRDX5 in TPP. PHF5A and RRM1 as other representative proteins which had significantly oxidized peptides and higher stability in response to auranofin.

Conclusions

Taken together, we showcased the efficiency of chemical proteomics for prediction of the target and MOA of auranofin. As expected, different techniques have various strengths and weaknesses, and must be used in parallel for obtaining the most reliable targets. Another important finding of this study was that redox proteomics can be combined with TPP to discover downstream redox events with effect on protein stability, and potentially map them to redox-active or generally active sites.

4.4 PAPER IV. THE DEUBIQUITINASE INHIBITOR B-AP15 INDUCES STRONG PROTEOTOXIC STRESS AND MITOCHONDRIAL DAMAGE

In **Paper IV**, we applied quantitative multiplexed proteomics to delineate the phenotypic response of colon cancer cells to b-AP15 in comparison with another proteasome inhibitor bortezomib. b-AP15 belongs to the bis-benzylidene piperidone compound family, which have been shown to effectively kill apoptosis-resistant cells, as shown in several tumor models. b-AP15 is known to inhibit deubiquitinases USP14 and UCHL5. Although the mRNA and protein expression profiles for b-AP15 and bortezomib were quite similar, b-AP15 induced a more significant expression of chaperones. Furthermore, polyubiquitinated proteins accumulated to a higher degree in response to b-AP15 than bortezomib, which were shown to co-localize with organelle membranes such as mitochondria. Such a phenomenon reduced the mitochondrial oxidative phosphorylation. This reduction was exacerbated by severe proteotoxic stress, down-regulation of VCP/p97 and inhibition of endoplasmic reticulum translocation. The overall results indicated that the co-localization of misfolded proteins with mitochondrial membrane might underlie the atypical cell death mode. The effect of b-AP15 on mitochondrial cluster of proteins was also obvious in **Paper II**. Therefore, compounds targeting mitochondria might be promising for eradication of apoptosis resistant tumors.

4.5 PAPER V. SYSTEM-WIDE IDENTIFICATION OF ENZYME SUBSTRATES BY THERMAL ANALYSIS (SIESTA)

Many proteins modify their specific substrates by post-translation modification. In addition to expression level, protein regulation and numerous cellular processes are modulated with PTMs (Mann and Jensen 2003). PTMs are capable of changing several protein attributes such as function, stability, hemostasis and localization (Houde, Peng et al. 2010), and can ultimately dictate cellular diversification as an independent regulator along with protein expression. Therefore, the identification and characterization of PTMs and their biological consequences has emerged as a growing and lively research area (Weinert, Narita et al. 2018). A central paradigm in PTM research is characterizing enzyme-substrate associations. Such knowledge is crucial for the fundamental understanding of cell biology in general and disease mechanisms in particular. In addition, modified substrates provide an output and are an integral component of many high-throughput screening assays in drug discovery (Von Ahsen and Bömer 2005). The development of therapeutics for several conditions such as Parkinson's disease (Steger, Tonelli et al. 2016) and cancer (Byrd and Blagg 2018) is in certain cases hampered by the lack of knowledge on the physiological substrates of enzymes for developing such screens.

In spite of the importance of enzyme-substrate associations, generic proteome-wide methods for characterizing the molecular entities partaking in these reactions are not available. Specific substrates for enzymes can be experimentally identified by several techniques such as creating substrate-trapping mutants (Flint, Tiganis et al. 1997), peptide immunoprecipitation (Matsuoka, Ballif et al. 2007), affinity purification-mass spectrometry (Low, Peng et al. 2014), employing peptide (Köhn, Gutierrez-Rodriguez et al. 2007) or protein arrays (Feilner, Hultschig et al. 2005), using genetic and pharmacologic perturbations (Yen and Elledge 2008) and client proteins tagging by substrate analogues using engineered enzymes (Ubersax, Woodbury et al. 2003). However these methods are developed for a certain enzyme (class), take long to optimize and are not straightforward. Furthermore, engineered enzymes might change the biology of the system under study, and are prone to false positive discoveries. Therefore, developing an unbiased system-wide tool for discovery of specific enzyme substrates can be a significant advance.

It was previously known that protein stability can shift upon interaction with metabolites, proteins and nucleic acids (Park and Marqusee 2005; Niesen, Berglund et al. 2007; Molina, Jafari et al. 2013; Savitski, Reinhard et al. 2014). Proteome-wide techniques such as TPP can be used to measure protein stability changes arising from such interactions (Savitski, Reinhard et al. 2014). PTMs have also been postulated to alter thermal stability of substrate proteins (Becher, Andres-Pons et al. 2018; Dai, Zhao et al. 2018; Drake, Hou et al. 2018). Therefore, one could theoretically add a recombinant enzyme and its co-factor to the cell lysate and monitor the thermal stability of proteins in a proteome-wide manner, to reveal enzyme substrate proteins with shift in stability (Savitski, Reinhard et al. 2014). A PTM-induced shift in stability can hypothetically arise from conformational transitions, promoting or disruption of protein-protein interactions, etc.

However, the enzyme-protein and co-factor-protein interactions can also happen and mask the substrate proteins. In our method called System-wide Identification of Enzyme Substrates by Thermal Analysis (SIESTA), we solved this problem by tracking specific thermal stability changes of substrate proteins (Saei, Astorga Wells et al. 2018). Meaning that stability changes in the candidate substrate proteins induced by a combination of enzyme and co-factor are contrasted to stability changes induced by enzyme or co-factor alone in OPLS-DA. The method workflow is shown in **Figure 35**.

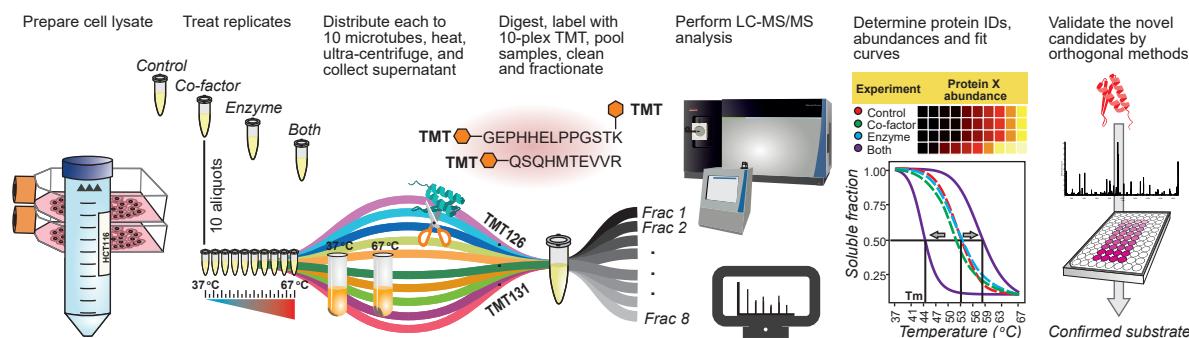


Figure 35. SIESTA workflow. Cell lysate is prepared in a non-denaturing buffer by several freeze-thaw cycles to preserve the native conformation of proteins. Cell lysate aliquots are then treated with vehicle, co-factor, enzyme or both. After the treatment period, each condition is split into 10 microtubes, and each microtube is heated to a temperature point in the 37-67 °C range. After ultracentrifugation to remove unfolded proteins, identical volumes of supernatants are taken for digestion with trypsin. Thereafter, the samples are labeled with 10-plex TMT, pooled, cleaned by SepPak and fractionated by reversed-phase chromatography. The fractions are then analyzed by LC-MS/MS to quantify the proteins. Finally, after sigmoid curve fitting, melting temperature T_m is determined for each protein. The candidate substrate proteins for an enzyme are thus those that show a larger stability change upon addition of enzyme-co-factor combination and should be validated by orthogonal methods. Figure reused with permission from (Saei, Astorga Wells et al. 2018).

4.5.1 SIESTA identifies known and putative TXNRD1 substrates

We used the thioredoxin reductase 1 (TXNRD1) as a proof of principle system. Using NADPH as a co-factor, TXNRD1 reduces the disulfide bonds in a limited number of specific substrate proteins (Arnér and Holmgren 2000). Such a reduction should destabilize the structure of substrate proteins and may lead to negative ΔT_m . We therefore treated the HCT116 cell lysate with vehicle, 1 mM NADPH, 1 μ M TXNRD1, or both. Samples were prepared according to the workflow in **Figure 35**.

As expected, NADPH treatment stabilized several proteins known to interact with this co-factor (**Figure 36A-B**). TXNRD1+NADPH treatment specifically destabilized known substrate proteins (**Figure 36C**). Furthermore, as observed in **Figure 36C**, an expected asymmetry in T_m shifts was noted in favor of destabilization, in line with TXNRD1 function. In an OPLS-DA model, TXNRD1+NADPH treatment was contrasted with the three other treatments to reveal the specifically destabilized proteins (**Figure 36D**). The melting curves for a number of known and putative TXNRD1 substrates are shown in **Figure 36E**. In total, 28 candidate substrates were found and mapped to the following INTERPRO Protein Domains and Features

pathways: “thioredoxin-like fold” (8 proteins) and “peroxiredoxin, C-terminal” (4 proteins) ($p < 0.001$). These pathways are in line with the key physiological functions of TXNRD1 (Amér and Holmgren 2000).

The protein showing the largest shift in stability was GPX1 (**Figure 36C-E**), which is a selenoprotein usually supposed to be GSH-dependent (Brigelius-Flohé and Maiorino 2013). However, TXNRD1 can directly reduce some GPX isoenzymes (Björnstedt, Xue et al. 1994). Several identified candidates such as peroxiredoxins (PRDXs) (Chae, Kim et al. 1999), TXNL1 (or TRP32) (Jiménez, Pelto-Huikko et al. 2006), NXN (Funato, Michiue et al. 2006), as well as GSTO1 and GSTO2 (Board, Coggan et al. 2000) are well-known substrates of TXNRD1 or TXNRD1-dependent enzymes. The other proteins identified in this screen are thus putative substrates of TXNRD1.

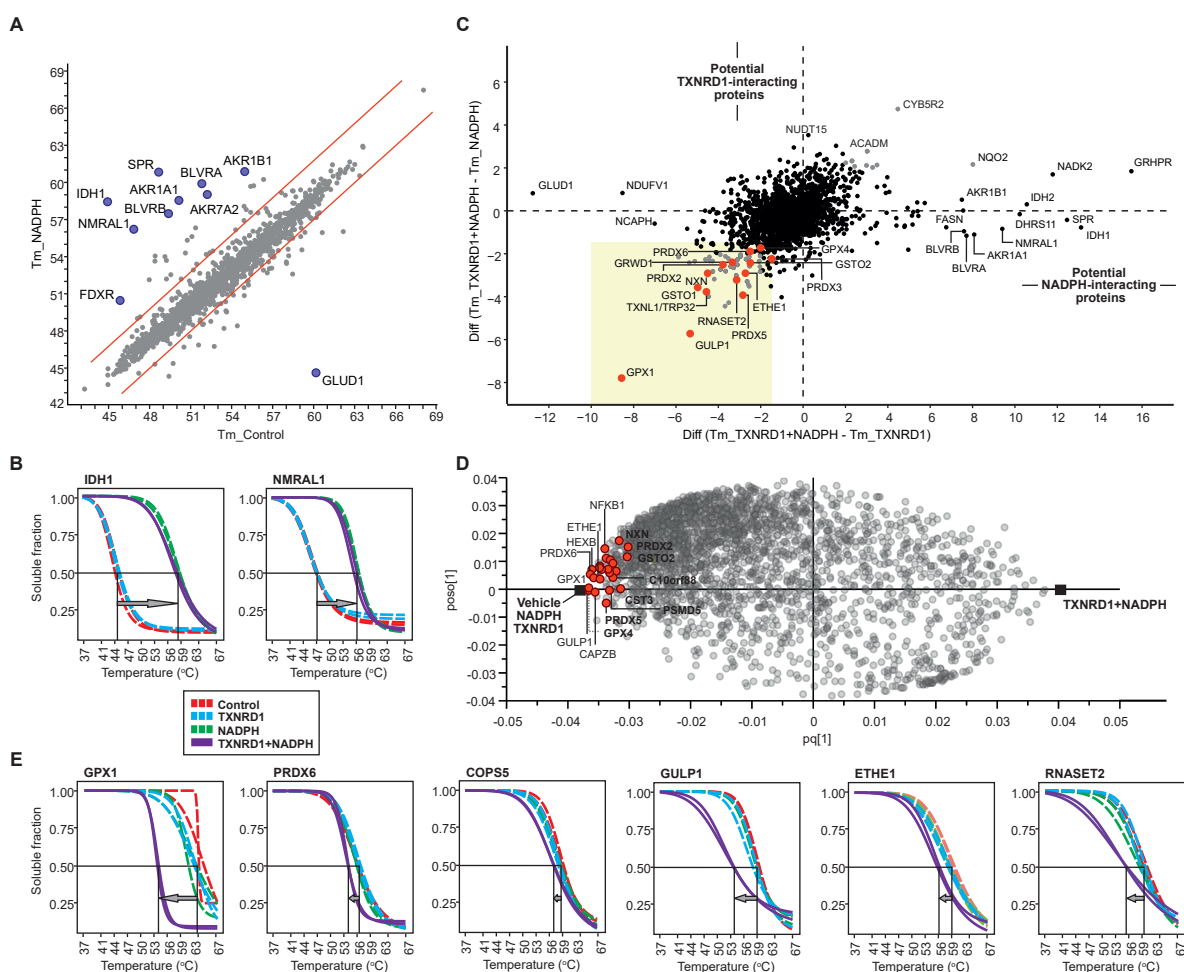


Figure 36. SIESTA experiment on the proof-of-principle TXNRD1 system revealed a number of known and putative substrates. **A)** A linear regression was performed on SIESTA results for control and NADPH (95%-prediction interval as a cutoff). The outlier proteins are known (purple circles) or putative NADPH binders. **B)** Thermal stability shifts for known NADPH binding proteins IDH1 and NMRAL1. **C)** A T_m differences scatterplot shows the shifts occurring only after adding a combination of TXNRD1 and NADPH; these shifts arising from enzymatic modifications (known and putative substrates as red circles). **D)** An OPLS-DA model contrasting the “TXNRD1+NADPH” T_m vs. all other treatments reveals potential substrates (red circles) located on the negative extremity of x axis (black square). **E)** Exemplary melting curves of known (GPX1 and PRDX6) and putative substrates (COPS5, GULP1, ETHE1 and RNASET2) for TXNRD1. Figure reused with permission from (Saei, Astorga Wells et al. 2018).

4.5.2 SIESTA identifies many novel putative substrates for PARP10

Poly-(ADP-ribose) polymerase-10 (PARP10) was the second system where we applied SIESTA. PARP10 belongs to the PARP family of proteins and is only capable of performing mono-ADP ribosylation (Kleine, Poreba et al. 2008). This reaction takes place in the presence of NAD as a co-factor. ADP-ribosylation is involved in signaling, DNA repair, regulation of gene expression and cell death (Gupte, Liu et al. 2017). The glycosidic and thus the labile nature of this modification makes it difficult to detect this PTM by MS/MS, as the ADP-ribose can be lost during sample processing. Different strategies have focused on enrichment of the modified peptides, and gentle MS/MS methods have been optimized (Carter-O'Connell, Jin et al. 2014; Martello, Leutert et al. 2016). Mono-ADP-ribosyltransferases such as PARP10, unlike the poly-ADP-ribosyltransferases family members, have not been studied extensively and require further characterization.

On the contrary to the TXNRD1 system, we expected the change in stability of substrates in both direction in the PARP10 system. The stability shift of several proteins known to interact with NAD verified the experiment quality. 28 potential protein substrates were found to change stability only when a combination of PARP10 and NAD was added to the lysate (16 stabilized and 12 destabilized proteins) (**Figure 37A**). Exemplary melting curves for these proteins are shown in **Figure 37B**. The OPLS-DA model was also capable of identifying these substrates by contrasting the T_m of "PARP10+NAD" T_m vs. those from all other treatments. Out of the 28 identified putative substrates, 7 proteins mapped to "ribonucleoprotein complex" ($p < 0.02$), and 19 were localized to the "nucleus" ($p < 0.05$), which is in good agreement with the cellular roles and localization of PARP10 (Bock, Todorova et al. 2015). Several of these proteins such as ILF2 and 3 are already known substrates of PARP10 (Carter-O'Connell, Jin et al. 2016).

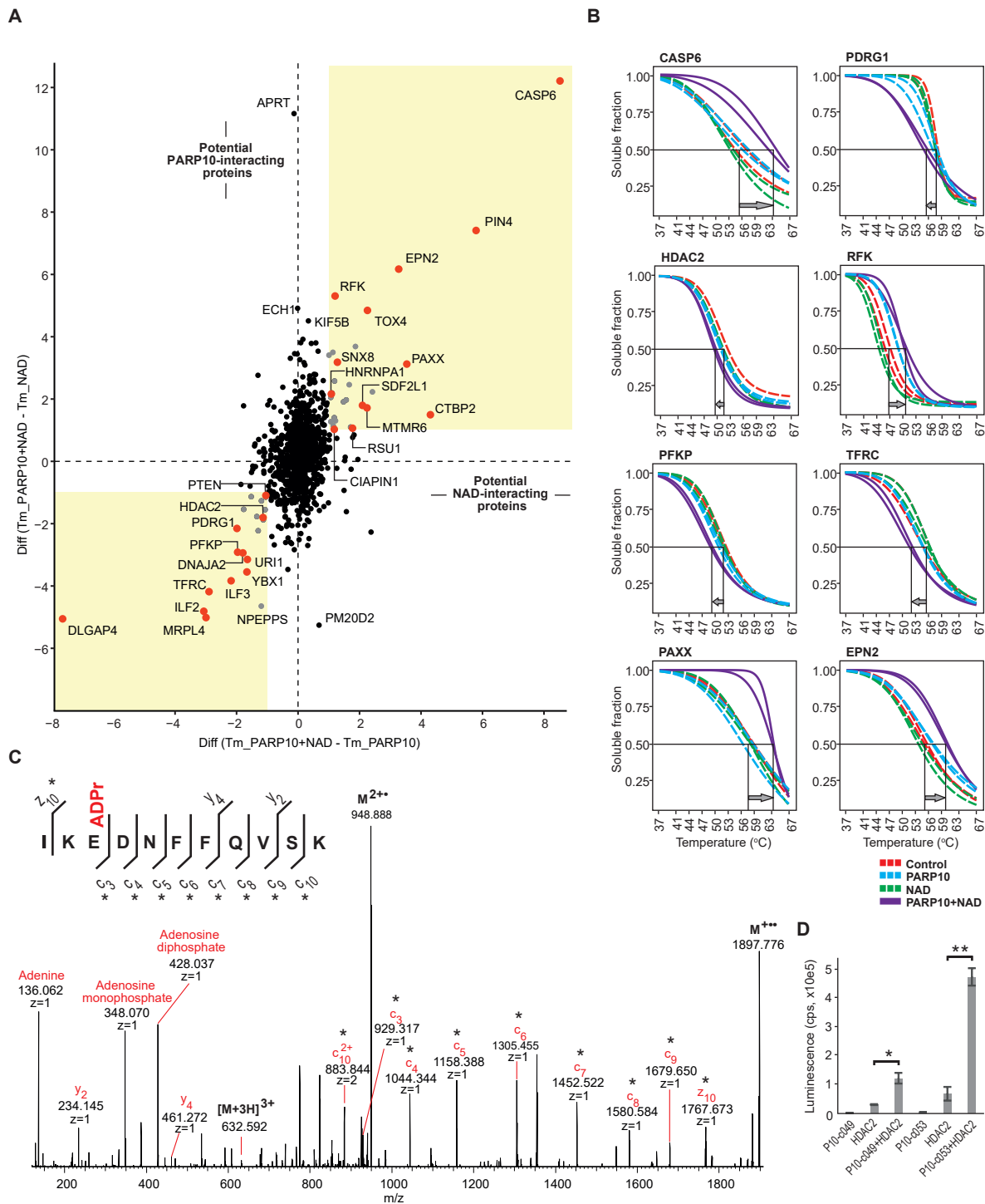


Figure 37. SIESTA identified known and novel protein substrates for PARP10. **A)** A T_m differences scatterplot reveals shifts occurring only upon simultaneous addition of NAD and PARP10 to the cell lysate. Red circles represent potential substrates melting reproducibly. **B)** Exemplary melting curves of putative PARP10 substrates. **C)** Targeted MS/MS analysis of a RFK peptide showed the mono-ADP-ribosylation of a glutamic acid residue (the site with the highest sequence-fitting score). The fragments carrying the ADP-ribose are shown with an asterisk. **D)** HDAC2 mono-ADP-ribosylation was validated in two independent experiments using recombinant HDAC2 and two PARP10 catalytic subunit constructs (n=2 replicates and mean±SD; *p<0.02 and **p<0.005 two-sided student t-test). Figure reused with permission from (Saei, Astorga Wells et al. 2018).

To validate a number of putative substrates, we manually checked their melting curves and extracted their rankings from OPLS-DA model loadings. Another criterion was the availability of high purity recombinant proteins. Finally, caspase-6 and RFK (stabilized) as well as PDRG1

and HDAC2 (destabilized) were chosen to verify the presence of PARP10-catalyzed mono-ADP-ribosylation. These proteins were incubated with PARP10 and NAD, and then digested and analyzed by LC-MS/MS.

The MS/MS events were monitored in real time for signature ions of adenine, adenosine-18 and adenosine monophosphate (m/z 136.0623, 250.094 and 348.0709, respectively). The presence of these signatures would initiate a second MS/MS event using electron-transfer dissociation (ETD) with a supplementary HCD activation. Using this method, we could prove the mono-ADP-ribosylation of RFK on four sites Glu140, Glu131, Glu113 and Arg14 (sequence coverage of 94%), ordered by the peptide score and PDRG1 on Glu110, Glu75 and Asp32 (sequence coverage of 74%). The ETD MS/MS spectrum of a RFK peptide (Glu140 ADP-ribosylation) is shown in **Figure 37C**. Due to the incomplete sequence coverage of HDAC2 with trypsin (with and without LysC) digestion, an *in vitro* chemiluminescence assay was used for validation of mono-ADP-ribosylation on HDAC2. HDAC2 was significantly modified with both PARP10 catalytic domain constructs (**Figure 37D**).

Although caspase-6 exhibited the largest specific stabilization (10.4 °C, **Figure 37A-B**) as a putative substrate for PARP10, we could not verify its modification in the two *in vitro* assays. A review of literature showed that PARP10 is actually a substrate for caspase-6 during apoptosis (Herzog, Hartkamp et al. 2013), having a major cleavage site for this protease at D406 (Herzog, Hartkamp et al. 2013). Therefore, the huge thermal shift specifically observed for caspase 6 might be related to induction of a conformational change in caspase-6 upon PARP10 binding. The conformational change in caspase-6 when binding to its substrates has been reported before (Vaidya, Velázquez-Delgado et al. 2011). However, the auto-modification of PARP10 is required for effective binding, as a thermal stability shift for caspase-6 was not noted in the absence of NAD.

Conclusions

Based on the above results, we believe that SIESTA can be used as a universal approach for unbiased identification of protein substrates for specific enzymes in a proteome-wide manner. SIESTA can also be applied to discover cell- or tissue-specific substrates by comparative stability monitoring of lysates from different sources. Therefore, SIESTA is likely to enhance our understanding of enzyme systems in hemostasis and disease and facilitate the incorporation of substrates in high throughput screening in drug discovery and development.

5 CONCLUSIONS AND FUTURE PERSPECTIVES

In the past decade, proteomics-based target deconvolution technologies have evolved to system-wide methods in which all human proteins are scrutinized for binding or correlation with efficacy of unmodified compounds in living cells. The availability and evolution of several techniques for drug target and MOA characterization indicates that a single method is not capable of—or should not be trusted to, exclusively pinpointing the efficacy targets responsible for the compound effects. Although the readout from each technique is a list of potential targets, a long list of candidate protein can also confound the identification of central efficacy targets. Combining different methods allows for shortlisting the highly potential candidates. The inclusion of matrix-detached cells in the proteomics experiments such as FITeXP was against the mainstream, but showed that specific information about drug targets, survival pathways and processes differentiating life from death can be obtained. We believe that this research will have an impact in later works and contribute to our understanding of cell death.

ProTargetMiner is a resource for catering to the research community and for assisting the phenotypic drug discovery paradigm. This database can also open up an opportunity for reverse chemical proteomics for even linking known targets to novel biological phenomena by interpretation of compound-induced phenotypes, which can prove valuable in drug repurposing. ProTargetMiner can be expanded in future by profiling more cell lines, more drugs (also non-anticancer agents) and different biological concentrations.

The protein pairwise correlation database based on ProTargetMiner dataset uncovered unexpected complexity in protein regulation that might not be an artefact. It is yet to be seen if this database can be used for characterization of proteins and their functions. Furthermore, the pairwise protein anti-correlation might reveal novel regulatory pathways, especially in the context of cancer. The set of proteins with most stable expression in ProTargetMiner is also a resource which must be analyzed in more detail to shortlist a number of top candidates with lowest standard deviation, high abundance and consistent expression in tissues. This goal can be partly achieved with the deep data available in MCF-7 and RKO cell lines in ProTargetMiner.

Finally, the lack of knowledge on specific protein substrates for enzymes hampers the design of high-throughput screening assays in drug discovery. Furthermore, such knowledge is fundamental in the understanding of biology in homeostasis and disease. So far, we have demonstrated the application of SIESTA in the reactions involving electron transfer (TXNRD1 system) and ADP-ribosylation. We also have positive results that SIESTA can be applied to identify phosphorylation events (in the AKT1 system) driving protein stability. Furthermore, SIESTA can be applied to different cell lines and tissues to discover cell- or tissue-specific PTMs. Another focus could be on delineating the isoform specificity in enzyme families (Gonzalez and McGraw 2009). Elucidation of substrate specificity is the key to targeted therapy (Toker 2012).

6 ACKNOWLEDGEMENTS

First and foremost, I would like to thank my principal supervisor Prof. **Roman Zubarev**. I am grateful to Roman for entrusting me, providing me with the opportunity to join his team in the first place, embracing me as a young scientist, and subsequently teaching me most of the things I know, and imparting his theories, hypotheses, knowledge, writing skills, and especially his critical thinking on me. Besides, his logical, realistic, unbiased and statistical mindset were exceptional and crucial in my studies. I wish one day, I will become as good a teacher and mentor as he is. In special cases, Roman has his own creative ways of shortlisting candidates; he told me that I was apparently an outlier for the number of publications as a function of age. In other words, I owe my Ph.D. to mathematics and in particular regression analysis.

I am also very thankful to my co-supervisor Prof. **Thomas Helleday**, who is a super-motivational engine. Thanks for all your scientific support, insightful feedbacks and inspirations. I am certainly in debt to a number of professors in Karolinska Institutet including Profs. **Elias Arnér**, **Herwig Schöler**, **Per Arvidsson**, **Stig Linder** and **Galina Selivanova** with whom we started interesting collaborations. I would like to express my appreciation to my mentor Prof. **Arne Holmgren** for all the discussions we have had and for kindly accepting to be my defense chairperson. I need to extend another gratitude to **Elias** who has been of much guidance, as if to be my second mentor, and to **Per** as he accepted my internship in SciLifeLab Drug Discovery and Development Platform, where we started an interesting project. In that platform, I was assigned to the Biochemical Screening facility and I am grateful to **Bo Lundgren** and **Hjalmar Gullberg** for everything I learned from them during that period. I would like to also acknowledge Prof. **Bernhard Kuster** for accepting me as a visiting researcher in Technische Universität München to master Thermal Proteome Profiling and his valuable advices (also great Oktoberfest and minibar in the lab).

I am thankful to my half-time committee Profs. **Stig Linder**, **Rolf Lewensohn** and **Ola Larsson**, for constructive discussion about and criticism of my research project. I am also grateful to my defense opponent Dr. **Karl Mechtler** for his constructive reading of my thesis and pleasantly opposing me in the dissertation. Special thanks to my defense examination board Profs. **Johan Malmström**, **Rolf Lewensohn**, and **Jonas Bergquist** for their input in my dissertation and their comments on my work.

I am indebted to Dr. **Roozbeh Eskandari** from Cornell University as an invisible mentor and peer who I have never met personally. At the moment of desperation when I was on the verge of giving up, he was the one who shouted at me (actually cursed at), pushed me and motivated me to pursue my dream. Our scientific discussions have been beyond motivating and helpful. I also need to acknowledge Dr. **Zohreh Amoozgar** from Harvard Medical School for all the time she devoted for my guidance and discussions when I was applying to Ph.D. programs. Her constructive criticism greatly improved my writing skills.

Profs. **Morteza Mahmoudi**, **Yadollah Omid**, **Jaleh Barar** and **A. Ghasem Jouyban** it was a pleasure working with you as an undergrad. I owe my scientific foundation to you.

On a more personal note, **Pierre Sabatier**, thank you for all the camaraderie and for the memory of the refreshments, Sunday promenades and the papers we have published together and the many that are to come. We have really done some brainstorming on the hindsight and hopefully this will continue for a lifetime. **Christian Beusch**, thanks a lot for being an awesome funny friend and a valuable colleague, I would have a very hard time without your help. Thanks for all the good and fun company gents. **Bo Zhang**, same goes for you and thanks for being a true friend. Sorry for all the data analysis headaches. We miss your intelligence in the lab. **Hassan Foroughi Asl** thanks for all the scientific and non-scientific dialogs, and all the strategic guidance you have provided, and also showing me ways when I came to Stockholm. You were my first friend in Sweden. **Mohammad Pirmoradian**, thanks for the friendship, getting me going in the lab and showing me the way by being a senior Ph.D. student. **Nataliya Tarasova**, thank you for all the discussions when we were office mates and the numerous times you have answered my questions.

Alexey Chernobrovkin, I am grateful to you for things you taught me and for everything you have done for the Zubarev group in general. Your absence can be felt like a black hole in the lab. Your FITExp method laid the groundwork for the current thesis and without your experience and statistical help, it would be impossible to develop the project this far. I am also acknowledging all the comments, logical criticism and help I have received from my great friend and colleague **Juan Astorga Wells**. His advice has always been from a special perspective and different from the mainstream. Alexey and Juan, the champions of our foosball tournament, we would have beaten you in the foosball tournament if we were not lightheaded; but I am sure we won in the parallel universe. I should also take the opportunity and thank **Peter Teiwick** from Waters Company for funding this wonderful event, through which I got to know many nice people.

Ákos Végvári, you have been of great help and I have learned a lot from you. I love your stories and wisdom. I would like to thank **Sergey Rodin**, for all the career and science guidance, the fun and friendly after work nights and especially for strategizing the writing and submission of our papers. **Hassan Gharibi**, I like your presence in the lab and you have been helpful in ways that will be appreciated by time. It is also nice to have somebody speak in Persian to me from time to time. **Mohsen Ghasemi**, you had a short stay in our lab, but we will be friends of a lifetime. Thanks for all the fun time. I thank **Massimiliano Gaetani** for all the scientific discussions, the collaborative projects, and last but not least, sometimes letting us run the instrument.

Marie Ståhlberg, **Carina Palmberg**, you are remarkable and indispensable people, thank you so much for keeping us up with amazing Swedish traditions in the lab and helping us with the instruments like there is no tomorrow.

Eirini Alexandra Alexandridou, thanks for being a cheer in the lab and waking us up with colorful sweaters in the morning, though sometimes they hurt my eyes. **Ülkü Guler Tokat**, it was fun and fruitful working with you on the ProTargetMiner project, and it is nice that we are in touch. **Eleni Stergiou**, you have been a good student and I wish you success in future, on

the condition that you vortex and/or sonicate more samples. I would like to also take the chance and thank my past and present friends and colleagues in the Zubarev lab in no particular order, including **Susanna Lundström, Jimmy Ytterberg, Rahul Pandey, Zhe Yang, Jijing Janet Wang, Hongyu Xie, Luciano Di Stefano, Hongqian Yang, Dorothea Rutishauser, Xueshu Xie, Alexander Manoylov, Harleen Dhot, Xuepei Zhang** and **Sasha Zdanovich** for all the nice time we have had together. I would like to also acknowledge **Tobias Karlberg, Katja Näreoja, Ann-Gerd Thorsell** from the Schüler lab and **Qing Cheng, Katarina Johansson, Markus Dagnell** and **Belen Espinosa** from the Arnér lab.

Victoria Balabanova and **Alessandra Nanni**, I am very thankful to you for all your help, support and everyday assistance. Our great IT support **Chad Tunell**, you have sometimes been helpful beyond expectations.

My fantastic friends. All of you! Thanks for being there and making my Ph.D. fun and enjoyable: **Andrea Fossati, Niels Leijten, Vincent Gemard**, for all the foosball and FIFA games, grills, hang outs, Valborgs, GOT nights, laughter, sarcasm and of course sometimes science. I vividly remember Andrea's foosball shots like he's in war zone, Vincent's wall shots, magic shots and introduction of pure randomness in foosball, and Niels taking it very seriously. Some fun people have made my Ph.D. enjoyable: **Michael Bonner** and **Jeremy Forsberg**, thanks for all the discussions and consults, I particularly enjoy you guys taking time and thinking sincerely before any conversation. **Shane Wright**, you have been gathering a group of nice people and caused a lot of interactions. Thanks for being up for it and keep it up back in Canada. **Mike Aoun**, you are a very cool friend and collaborator, thanks for the positive energy and all the quality time. **Neil Rumachik** thanks for all the discussions about science and life and being a hilarious Instagram person. **Amit Saxena**, thanks for all the collaborations and talks that we have had, especially the guidance about organizing my defense seminar. Other friends that I have had the pleasure to meet and get influenced by: **Antonio Checa, Mauricio Barrientos, Marcus Lundin, Annina Steinbach, Soudabeh Rad Pour, Alireza Azimi, Alis Salis, Giuseppe Santopolo, Luca D. Retattino** and **Alberto Joven Araus**.

A number of people in my life have been a source of motivation: **Ammar Shamsedini** for we grew the desire together to move and be adventurous, **Bahar Matin** for always encouraging me, and **Arash Manafirad**, for all out motivating discussions on starting Ph.D. During the years I was planning to start my Ph.D., I had the pleasure of working with **Abolfazl Barzegari** and learned a lot of things from him, we had a number of great collaborations and became close friends. I also memorized all the journal impact factors in the process. My special thanks goes to **Nazli Saeedi**, who was meticulous with CorelDraw and from whom I have borrowed some organelles on my thesis cover.

I would like to thank all my Stockholm friends outside Karolinska circle in no particular order; **Behnood Lotfalizadeh, Hamed Kalhor, Marjan Abbasi, Kristers Pavelsons, Taraneh Shirazi, Mahsa Ebrahimi, Mehdi Bakhshian, Mehrnaz Mirzaei, Hengameh Khodadadi, Sepehr Saeidibonab** and **Anna Nazari**. You have been my family in Sweden, have supported

me emotionally and I have had very nice moments with you. I will keep these memories close to my heart.

I would like to thank **Roman Zubarev**, **Christian Beusch**, **Pierre Sabatier** and my sister **Sara Saei** for proofreading this thesis work.

Finally, my most special thanks goes to my family, and in particular my beloved **mom** and adored **dad**, and then my dearest sisters **Sara** and **Sanaz** who always believed in me and wished me the best. Thanks for all the relentless sacrifices you have made to get me here. I owe everything to my parents who have been supportive of my decisions and actions throughout my life.

7 REFERENCES

- Aebersold, R. and M. Mann (2003). "Mass spectrometry-based proteomics." *Nature* 422(6928): 198-207.
- Akbulut, Y., H. J. Gaunt, et al. (2015). "(-)-Englerin A is a Potent and Selective Activator of TRPC4 and TRPC5 Calcium Channels." *Angewandte Chemie* 127(12): 3858-3862.
- Almqvist, H., H. Axelsson, et al. (2016). "CETSA screening identifies known and novel thymidylate synthase inhibitors and slow intracellular activation of 5-fluorouracil." *Nature Communications* 24(7): 11040.
- Angelucci, F., A. A. Sayed, et al. (2009). "Inhibition of *Schistosoma mansoni* thioredoxin-glutathione reductase by auranofin structural and kinetic aspects." *Journal of Biological Chemistry* 284(42): 28977-28985.
- Arastu-Kapur, S., J. L. Anderl, et al. (2011). "Non-proteasomal targets of the proteasome inhibitors bortezomib and carfilzomib: a link to clinical adverse events." *Clinical Cancer Research*. 17(9):2734-2743
- Arnér, E. S. and A. Holmgren (2000). "Physiological functions of thioredoxin and thioredoxin reductase." *European Journal of Biochemistry* 267(20): 6102-6109.
- Bantscheff, M., D. Eberhard, et al. (2007). "Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors." *Nature Biotechnology* 25(9): 1035-1044.
- Bantscheff, M., C. Hopf, et al. (2011). "Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes." *Nature Biotechnology* 29(3): 255-265.
- Bassilana, F., A. Carlson, et al. (2014). "Target identification for a Hedgehog pathway inhibitor reveals the receptor GPR39." *Nature Chemical Biology* 10(5): 343-349.
- Bath, T. S., C. Francavilla, et al. (2014). "Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics." *Journal of Proteome Research* 13(12): 6176-6186.
- Becher, I., A. Andres-Pons, et al. (2018). "Pervasive Protein Thermal Stability Variation during the Cell Cycle." *Cell* 173(6): 1495-1507 e1418.
- Bennett, E. J., J. Rush, et al. (2010). "Dynamics of cullin-RING ubiquitin ligase network revealed by systematic quantitative proteomics." *Cell* 143(6): 951-965.
- Bensaad, K., A. Tsuruta, et al. (2006). "TIGAR, a p53-inducible regulator of glycolysis and apoptosis." *Cell* 126(1): 107-120.
- Björnstedt, M., J. Xue, et al. (1994). "The thioredoxin and glutaredoxin systems are efficient electron donors to human plasma glutathione peroxidase." *Journal of Biological Chemistry* 269(47): 29382-29384.
- Board, P. G., M. Coggan, et al. (2000). "Identification, characterization, and crystal structure of the Omega class glutathione transferases." *Journal of Biological Chemistry* 275(32): 24798-24806.
- Bock, F. J., T. T. Todorova, et al. (2015). "RNA regulation by poly (ADP-ribose) polymerases." *Molecular Cell* 58(6): 959-969.
- Boisvert, F.-M., Y. Ahmad, et al. (2012). "A quantitative spatial proteomics analysis of proteome turnover in human cells." *Molecular & Cellular Proteomics* 11(3): M111. 011429.
- Branca, R. M., L. M. Orre, et al. (2014). "HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics." *Nature Methods* 11(1): 59-62.

Bray, F., J. Ferlay, et al. (2018). "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." *CA: a Cancer Journal for Clinicians* 68(6): 394-424.

Brigelius-Flohé, R. and M. Maiorino (2013). "Glutathione peroxidases." *Biochimica et Biophysica Acta (BBA)-General Subjects* 1830(5): 3289-3303.

Broach, J. R. and J. Thorner (1996). "High-throughput screening for drug discovery." *Nature* 384(6604): 14-16.

Browne, C. M., B. Jiang, et al. (2018). "A Chemoproteomic Strategy for Direct and Proteome-wide Covalent Inhibitor Target-site Identification." *Journal of the American Chemical Society*. doi: 10.1021/jacs.8b07911.

Bruno, P. M., Y. Liu, et al. (2017). "A subset of platinum-containing chemotherapeutic agents kills cells by inducing ribosome biogenesis stress." *Nature Medicine* 23(4): 461.

Budnik, B., E. Levy, et al. (2018). "SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation." *Genome Biology* 19(1): 161.

Bull, V. H., K. Rajalingam, et al. (2012). "Sorafenib-induced mitochondrial complex I inactivation and cell death in human neuroblastoma cells." *Journal of Proteome Research* 11(3): 1609-1620.

Bunnage, M. E., A. M. Gilbert, et al. (2015). "Know your target, know your molecule." *Nature Chemical Biology* 11(6): 368-372.

Burgett, A. W., T. B. Poulsen, et al. (2011). "Natural products reveal cancer cell dependence on oxysterol-binding proteins." *Nature Chemical Biology* 7(9): 639-647.

Bylesjö, M., M. Rantalainen, et al. (2006). "OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification." *Journal of Chemometrics* 20(8-10): 341-351.

Byrd, K. M. and B. S. Blagg (2018). "Chaperone substrate provides missing link for cancer drug discovery." *Journal of Biological Chemistry* 293(7): 2381-2382.

Campbell, D. H., E. Luescher, et al. (1951). "Immunologic adsorbents I. Isolation of antibody by means of a cellulose-protein antigen." *Proceedings of the National Academy of Sciences* 37(9): 575-578.

Campuzano, I. D., T. San Miguel, et al. (2016). "High-Throughput Mass Spectrometric Analysis of Covalent Protein-Inhibitor Adducts for the Discovery of Irreversible Inhibitors: A Complete Workflow." *Journal of Biomolecular Screening* 21(2): 136-144.

Carter-O'Connell, I., H. Jin, et al. (2014). "Engineering the substrate specificity of ADP-ribosyltransferases for identifying direct protein targets." *Journal of the American Chemical Society* 136(14): 5201-5204.

Carter-O'Connell, I., H. Jin, et al. (2016). "Identifying family-member-specific targets of mono-ARTDs by using a chemical genetics approach." *Cell Reports* 14(3): 621-631.

Cebula, M., E. E. Schmidt, et al. (2015). "TrxR1 as a potent regulator of the Nrf2-Keap1 response system." *Antioxidants & Redox Signaling* 23(10): 823-853.

Chae, H. Z., H. J. Kim, et al. (1999). "Characterization of three isoforms of mammalian peroxiredoxin that reduce peroxides in the presence of thioredoxin." *Diabetes Research and Clinical Practice* 45(2): 101-112.

Chamrád, I., U. Rix, et al. (2013). "A miniaturized chemical proteomic approach for target profiling of clinical kinase inhibitors in tumor biopsies." *Journal of Proteome Research* 12(9): 4005-4017.

- Checconi, P., S. Salzano, et al. (2015). "Redox proteomics of the inflammatory secretome identifies a common set of redoxins and other glutathionylated proteins released in inflammation, influenza virus infection and oxidative stress." *PloS one* 10(5): e0127086.
- Chernobrovkin, A., C. Marin-Vicente, et al. (2015). "Functional Identification of Target by Expression Proteomics (FITExP) reveals protein targets and highlights mechanisms of action of small molecule drugs." *Scientific Reports* 5. doi: 10.1038/srep11176.
- Chernobrovkin, A. L. and R. A. Zubarev (2016). "How well can morphology assess cell death modality? A proteomics study." *Cell Death Discovery* 2. <https://doi.org/10.1038/cddiscovery.2016.68>.
- Chick, J. M., D. Kolippakkam, et al. (2015). "A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides." *Nature Biotechnology* 33(7): 743-749.
- Christophorou, M. A., G. Castelo-Branco, et al. (2014). "Citruination regulates pluripotency and histone H1 binding to chromatin." *Nature* 507(7490): 104-108.
- Cohen, A. A., N. Geva-Zatorsky, et al. (2008). "Dynamic proteomics of individual cancer cells in response to a drug." *Science* 322(5907): 1511-1516.
- Collins, M. A., J. An, et al. (2015). "Total protein is an effective loading control for cerebrospinal fluid western blots." *Journal of Neuroscience Methods* 251: 72-82.
- Cox, J. and M. Mann (2008). "MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification." *Nature Biotechnology* 26(12): 1367-1372.
- Cox, J., N. Neuhauser, et al. (2011). "Andromeda: a peptide search engine integrated into the MaxQuant environment." *Journal of Proteome Research* 10(4): 1794-1805.
- D'arcy, P., S. Brnjic, et al. (2011). "Inhibition of proteasome deubiquitinating activity as a new cancer therapy." *Nature Medicine* 17(12): 1636-1640.
- da Silva, D. C., P. B. Andrade, et al. (2017). "Neurotoxicity of the steroidal alkaloids tomatine and tomatidine is RIP1 kinase-and caspase-independent and involves the eIF2 α branch of the endoplasmic reticulum." *The Journal of Steroid Biochemistry and Molecular Biology* 171: 178-186.
- Dai, L., T. Zhao, et al. (2018). "Modulation of Protein-Interaction States through the Cell Cycle." *Cell* 173(6): 1481-1494 e1413.
- Daum, G., N. D. Lees, et al. (1998). "Biochemistry, cell biology and molecular biology of lipids of *Saccharomyces cerevisiae*." *Yeast* 14(16): 1471-1510.
- DeRisi, J. L., V. R. Iyer, et al. (1997). "Exploring the metabolic and genetic control of gene expression on a genomic scale." *Science* 278(5338): 680-686.
- Dittmer, A. and J. Dittmer (2006). " β -Actin is not a reliable loading control in Western blot analysis." *Electrophoresis* 27(14): 2844-2845.
- Drake, W. R., C.-W. Hou, et al. (2018). "New use for CETSA: monitoring innate immune receptor stability via post-translational modification by OGT." *Journal of Bioenergetics and Biomembranes*: 50(3): 231-240.
- Du, Y., H. Zhang, et al. (2012). "Glutathione and glutaredoxin act as a backup of human thioredoxin reductase 1 to reduce thioredoxin 1 preventing cell death by aurothioglucose." *Journal of Biological Chemistry* 287(45): 38210-38219.
- Eder, J., R. Sedrani, et al. (2014). "The discovery of first-in-class drugs: origins and evolution." *Nature reviews Drug discovery* 13(8): 577-587.

- Edfors, F., F. Danielsson, et al. (2016). "Gene-specific correlation of RNA and protein levels in human cells and tissues." *Molecular Systems Biology* 12(10): 883.
- Eisenberg, E. and E. Y. Levanon (2013). "Human housekeeping genes, revisited." *Trends in Genetics* 29(10): 569-574.
- Emanuele, M. J., A. Ciccia, et al. (2011). "Proliferating cell nuclear antigen (PCNA)-associated KIAA0101/PAF15 protein is a cell cycle-regulated anaphase-promoting complex/cyclosome substrate." *Proceedings of the National Academy of Sciences* 108(24): 9845-9850.
- Feilner, T., C. Hultschig, et al. (2005). "High throughput identification of potential Arabidopsis mitogen-activated protein kinases substrates." *Molecular & Cellular Proteomics* 4(10): 1558-1568.
- Ferretti, R., V. Palumbo, et al. (2010). "Morgana/chp-1, a ROCK inhibitor involved in centrosome duplication and tumorigenesis." *Developmental cell* 18(3): 486-495.
- Flint, A. J., T. Tiganis, et al. (1997). "Development of "substrate-trapping" mutants to identify physiological substrates of protein tyrosine phosphatases." *Proceedings of the National Academy of Sciences* 94(5): 1680-1685.
- Franken, H., T. Mathieson, et al. (2015). "Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry." *Nature Protocols* 10(10): 1567-1593.
- Funato, Y., T. Michiue, et al. (2006). "The thioredoxin-related redox-regulating protein nucleoredoxin inhibits Wnt- β -catenin signalling through Dishevelled." *Nature Cell Biology* 8(5): 501-508.
- Gaetani, M., P. Sabatier, et al. (2018). "Proteome Integral Stability Alteration assay dramatically increases throughput and sensitivity in profiling factor-induced proteome changes." *bioRxiv*. doi: <https://doi.org/10.1101/496398>.
- Galluzzi, L., J. Bravo-San Pedro, et al. (2015). "Essential versus accessory aspects of cell death: recommendations of the NCCD 2015." *Cell Death & Differentiation* 22(1): 58-73.
- Ganter, B., S. Tugendreich, et al. (2005). "Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action." *Journal of Biotechnology* 119(3): 219-244.
- Gholami, A. M., H. Hahne, et al. (2013). "Global proteome analysis of the NCI-60 cell line panel." *Cell reports* 4(3): 609-620.
- Gilmore-Hebert, M., R. Ramabhadran, et al. (2010). "Interactions of ErbB4 and Kap1 connect the growth factor and DNA damage response pathways." *Molecular Cancer Research*: 1541-7786. MCR-1510-0042.
- Gonzalez, E. and T. E. McGraw (2009). "The Akt kinases: isoform specificity in metabolism and cancer." *Cell Cycle* 8(16): 2502-2508.
- Gough, N. R. (2015). Focus Issue: Tackling reproducibility and accuracy in cell signaling experiments." *Science Signaling*. 8(371): eg4.
- Gregori-Puigjané, E., V. Setola, et al. (2012). "Identifying mechanism-of-action targets for drugs and probes." *Proceedings of the National Academy of Sciences* 109(28): 11178-11183.
- Gupte, R., Z. Liu, et al. (2017). "PARPs and ADP-ribosylation: recent advances linking molecular functions to biological outcomes." *Genes & development* 31(2): 101-126.
- Hall, A., D. Parsonage, et al. (2010). "Structural evidence that peroxiredoxin catalytic power is based on transition-state stabilization." *Journal of Molecular Biology* 402(1): 194-209.

- Harper, J. W., G. R. Adami, et al. (1993). "The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases." *Cell* 75(4): 805-816.
- Hebert, A. S., A. L. Richards, et al. (2014). "The one hour yeast proteome." *Molecular & Cellular Proteomics* 13(1): 339-347.
- Herzog, N., J. D. Hartkamp, et al. (2013). "Caspase-dependent cleavage of the mono-ADP-ribosyltransferase ARTD10 interferes with its pro-apoptotic function." *The FEBS Journal* 280(5): 1330-1343.
- Hieronimus, H., J. Lamb, et al. (2006). "Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators." *Cancer Cell* 10(4): 321-330.
- Houde, D., Y. Peng, et al. (2010). "Post-translational modifications differentially affect IgG1 conformation and receptor binding." *Molecular & Cellular Proteomics* 9(8): 1716-1728.
- Huang, D. W., B. T. Sherman, et al. (2008). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." *Nature Protocols* 4(1): 44-57.
- Hughes, T. R., M. J. Marton, et al. (2000). "Functional discovery via a compendium of expression profiles." *Cell* 102(1): 109-126.
- Hustedt, N. and D. Durocher (2017). "The control of DNA repair by the cell cycle." *Nature Cell Biology* 19(1): 1-9.
- Janes, K. A. (2015). "An analysis of critical factors for quantitative immunoblotting." *Science Signaling* 8(371): rs2-rs2.
- Janic, A., L. J. Valente, et al. (2018). "DNA repair processes are critical mediators of p53-dependent tumor suppression." *Nature Medicine* (24): 947-953.
- Jeon, K.-I., J.-Y. Jeong, et al. (2000). "Thiol-reactive metal compounds inhibit NF- κ B activation by blocking I κ B kinase." *The Journal of Immunology* 164(11): 5981-5989.
- Jiang, L., N. Kon, et al. (2015). "Ferroptosis as a p53-mediated activity during tumour suppression." *Nature* 520(7545): 57-62.
- Jiménez, A., M. Pelto-Huikko, et al. (2006). "Characterization of human thioredoxin-like-1: Potential involvement in the cellular response against glucose deprivation." *FEBS Letters* 580(3): 960-967.
- Kanoh, N., K. Honda, et al. (2005). "Photo-Cross-Linked Small-Molecule Affinity Matrix for Facilitating Forward and Reverse Chemical Genetics." *Angewandte Chemie International Edition* 44(23): 3559-3562.
- Klaeger, S., B. Gohlke, et al. (2016). "Chemical proteomics reveals ferrochelatase as a common off-target of kinase inhibitors." *ACS Chemical Biology* 11(5): 1245-1254.
- Klaeger, S., S. Heinzlmeir, et al. (2017). "The target landscape of clinical kinase drugs." *Science* 358(6367): eaan4368.
- Kleine, H., E. Poreba, et al. (2008). "Substrate-assisted catalysis by PARP10 limits its activity to mono-ADP-ribosylation." *Molecular Cell* 32(1): 57-69.
- Klimczak, M., P. Czerwińska, et al. (2017). "TRIM28 epigenetic corepressor is indispensable for stable induced pluripotent stem cell formation." *Stem Cell Research* 23: 163-172.
- Köhn, M., M. Gutierrez-Rodriguez, et al. (2007). "A microarray strategy for mapping the substrate specificity of protein tyrosine phosphatase." *Angewandte Chemie International Edition* 46(40): 7700-7703.
- Kotz, J. (2012). "Phenotypic screening, take two." *SciBX* 5(15): 380-380.

Kwon, H. J. and P. Karuso (2018). Chemical proteomics, an integrated research engine for exploring drug-target-phenotype interactions, *BioMed Central*. 16: 1.

Lamb, J., E. D. Crawford, et al. (2006). "The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease." *Science* 313(5795): 1929-1935.

Lee, H. G., J. Jo, et al. (2016). "State-of-the-art housekeeping proteins for quantitative western blotting: Revisiting the first draft of the human proteome." *Proteomics* 16(13): 1863-1867.

Lee, H. K., A. K. Hsu, et al. (2004). "Coexpression analysis of human genes across many microarray data sets." *Genome Research* 14(6): 1085-1094.

Lee, J. A., M. T. Uhlik, et al. (2012). "Modern phenotypic drug discovery is a viable, neoclassic pharma strategy." *Journal of Medicinal Chemistry* 55(10): 4527-4538.

Lee, R. F., A. Chernobrovkin, et al. (2017). "Expression proteomics study to determine metallodrug targets and optimal drug combinations." *Scientific Reports* 7(1): 1590.

Legewie, S., H. Herzelt, et al. (2008). "Recurrent design patterns in the feedback regulation of the mammalian signalling network." *Molecular Systems Biology* 4(1): 190.

Leuenberger, P., S. Ganscha, et al. (2017). "Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability." *Science* 355(6327): eaai7825.

Li, J. J. and M. D. Biggin (2015). "Statistics requantitates the central dogma." *Science* 347(6226): 1066-1067.

Li, R. and Y. Shen (2013). "An old method facing a new challenge: re-visiting housekeeping proteins as internal reference control for neuroscience research." *Life Sciences* 92(13): 747-751.

Litichevskiy, L., R. Peckner, et al. (2018). "A Library of Phosphoproteomic and chromatin signatures for characterizing cellular responses to drug perturbations." *Cell Systems* 6(4): 424-443.

Liu, Y., A. Beyer, et al. (2016). "On the dependency of cellular protein levels on mRNA abundance." *Cell* 165(3): 535-550.

Lo, Y.-C., S. Senese, et al. (2015). "Large-scale chemical similarity networks for target profiling of compounds identified in cell-based chemical screens." *PLoS Computational Biology* 11(3): e1004153.

Lohmann, W., H. Hayen, et al. (2008). "Covalent protein modification by reactive drug metabolites using online electrochemistry/liquid chromatography/mass spectrometry." *Analytical Chemistry* 80(24): 9714-9719.

Lomenick, B., R. Hao, et al. (2009). "Target identification using drug affinity responsive target stability (DARTS)." *Proceedings of the National Academy of Sciences* 106(51): 21984-21989.

Low, T. Y., M. Peng, et al. (2014). "A systems-wide screen identifies substrates of the SCF β TrCP ubiquitin ligase." *Science Signaling* 7(356): rs8-rs8.

Maaten, L. v. d. and G. Hinton (2008). "Visualizing data using t-SNE." *Journal of Machine Learning Research* 9: 2579-2605.

MacKinnon, A. L., J. L. Garrison, et al. (2007). "Photo-leucine incorporation reveals the target of a cyclodepsipeptide inhibitor of cotranslational translocation." *Journal of the American Chemical Society* 129(47): 14560-14561.

Malhotra, D., E. Portales-Casamar, et al. (2010). "Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through ChIP-Seq profiling and network analysis." *Nucleic Acids Research* 38(17): 5718-5734.

- Maniatis, T. and M. Ptashne (1973). "Multiple repressor binding at the operators in bacteriophage λ ." *Proceedings of the National Academy of Sciences* 70(5): 1531-1535.
- Mann, M. and O. N. Jensen (2003). "Proteomic analysis of post-translational modifications." *Nature Biotechnology* 21(3): 255-261.
- Marin-Vicente, C., Y. Lyutvinskiy, et al. (2013). "The effects of 5-fluorouracil on the proteome of colon cancer cells." *Journal of Proteome Research* 12(4): 1969-1979.
- Martello, R., M. Leutert, et al. (2016). "Proteome-wide identification of the endogenous ADP-ribosylome of mammalian cells and tissue." *Nature Communications* 7: 12917.
- Matsuoka, S., B. A. Ballif, et al. (2007). "ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage." *Science* 316(5828): 1160-1166.
- Medina-Franco, J. L., M. A. Giulianotti, et al. (2013). "Shifting from the single to the multitarget paradigm in drug discovery." *Drug Discovery Today* 18(9-10): 495-501.
- Meier, F., P. E. Geyer, et al. (2018). "BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes." *Nature Methods*: 15, 440-448.
- Moffat, J. G., J. Rudolph, et al. (2014). "Phenotypic screening in cancer drug discovery—past, present and future." *Nature Reviews Drug Discovery* 13(8): 588-602.
- Molina, D. M., R. Jafari, et al. (2013). "Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay." *Science* 341(6141): 84-87.
- Niesen, F. H., H. Berglund, et al. (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." *Nature Protocols* 2(9): 2212-2221.
- Ntai, I., L. Fornelli, et al. (2018). "Precise characterization of KRAS4b proteoforms in human colorectal cells and tumors reveals mutation/modification cross-talk." *Proceedings of the National Academy of Sciences* 115(16): 4140-4145.
- Nuñez, G., M. A. Benedict, et al. (1998). "Caspases: the proteases of the apoptotic pathway." *Oncogene* 17(25): 3237-3245.
- Oda, Y., T. Owa, et al. (2003). "Quantitative chemical proteomics for identifying candidate drug targets." *Analytical Chemistry* 75(9): 2159-2165.
- Palmer, A. C. and R. Kishony (2014). "Opposing effects of target overexpression reveal drug mechanisms." *Nature Communications* 5: 4296.
- Pantoliano, M. W., E. C. Petrella, et al. (2001). "High-density miniaturized thermal shift assays as a general strategy for drug discovery." *Journal of Biomolecular Screening* 6(6): 429-440.
- Park, C. and S. Marqusee (2005). "Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding." *Nature Methods* 2(3): 207-212.
- Park, H., J. Ha, et al. (2017). "Label-free target identification using in-gel fluorescence difference via thermal stability shift." *Chemical Science*. 8(2): 1127-1133.
- Peck, D., E. D. Crawford, et al. (2006). "A method for high-throughput gene expression signature analysis." *Genome Biology* 7(7): R61.
- Piazza, I., K. Kochanowski, et al. (2018). "A map of protein-metabolite interactions reveals principles of chemical communication." *Cell* 172(1-2): 358-372. e323.
- Pirmoradian, M., J. Astorga-Wells, et al. (2015). "Multijunction capillary isoelectric focusing device combined with online membrane-assisted buffer exchanger enables isoelectric point fractionation of intact human plasma proteins for biomarker discovery." *Analytical Chemistry* 87(23): 11840-11846.

Pirmoradian, M., H. Budamgunta, et al. (2013). "Rapid and deep human proteome analysis by single-dimension shotgun proteomics." *Molecular & Cellular Proteomics* 12(11): 3330-3338.

Readhead, B., B. J. Hartley, et al. (2018). "Expression-based drug screening of neural progenitor cells from individuals with schizophrenia." *Nature Communications* 9(1): 4412.

Rees, M. G., B. Seashore-Ludlow, et al. (2016). "Correlating chemical sensitivity and basal gene expression reveals mechanism of action." *Nature Chemical Biology* 12(2): 109-116.

Reinhard, F. B., D. Eberhard, et al. (2015). "Thermal proteome profiling monitors ligand interactions with cellular membrane proteins." *Nature Methods* 12(12): 1129-1131.

Rios Perez, M. V., D. Roife, et al. (2016). Auranofin to prevent progression of pancreatic ductal adenocarcinoma, *American Society of Clinical Oncology*. suppl 34(4): 236-236.

Rix, U. and G. Superti-Furga (2009). "Target profiling of small molecules by chemical proteomics." *Nature Chemical Biology* 5(9): 616-624.

Roddick, J. G. and R. Drysdale (1984). "Destabilization of liposome membranes by the steroidal glycoalkaloid α -tomatine." *Phytochemistry* 23(3): 543-547.

Roder, C. and M. J. Thomson (2015). "Auranofin: repurposing an old drug for a golden new age." *Drugs in R&D* 15(1): 13-20.

Sabatier, P., A. A. Saei, et al. "Dynamic proteomics reveals high plasticity of cellular proteome: growth-related and drug-induced changes in cancer cells are comparable." *Proteomics* 18(24):e1800118.

Saei, A. A., A. Chernobrovkin, et al. (2018). "ProTargetMiner: A proteome signature library of anticancer molecules for functional discovery." *bioRxiv*. <https://doi.org/10.1101/421115>.

Saei, A. A., J. Astorga Wells, et al. (2018). "System-wide Identification of Enzyme Substrates by Thermal Analysis (SIESTA)." *bioRxiv*. <https://doi.org/10.1101/423418>.

Saei, A. A., P. Sabatier, et al. (2018). "Comparative proteomics of dying and surviving cancer cells improves the identification of drug targets and sheds light on cell life/death decisions." *Molecular & Cellular Proteomics* 17(6): 1144-1155.

Salisbury, C. M. and B. F. Cravatt (2007). "Activity-based probes for proteomic profiling of histone deacetylase complexes." *Proceedings of the National Academy of Sciences* 104(4): 1171-1176.

Savitski, M. M., M. L. Nielsen, et al. (2006). "ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures." *Molecular & Cellular Proteomics* 5(5): 935-948.

Savitski, M. M., F. B. Reinhard, et al. (2014). "Tracking cancer drugs in living cells by thermal profiling of the proteome." *Science* 346(6205): 1255784.

Savitski, M. M., N. Zinn, et al. (2018). "Multiplexed proteome dynamics profiling reveals mechanisms controlling protein homeostasis." *Cell* 173(1): 260-274. e225.

Scheltema, R. A., J.-P. Hauschild, et al. (2014). "The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer." *Molecular & Cellular Proteomics* 13(12): 3698-3708.

Schenone, M., V. Dančik, et al. (2013). "Target identification and mechanism of action in chemical biology and drug discovery." *Nature Chemical Biology* 9(4): 232.

Schirle, M., M. Bantscheff, et al. (2012). "Mass spectrometry-based proteomics in preclinical drug discovery." *Chemistry & Biology* 19(1): 72-84.

Schirle, M. and J. L. Jenkins (2016). "Identifying compound efficacy targets in phenotypic drug discovery." *Drug Discovery Today* 21(1): 82-89.

Scott, J. D. and T. Pawson (2009). "Cell signaling in space and time: where proteins come together and when they're apart." *Science* 326(5957): 1220-1224.

Seo, M. S., S. W. Kang, et al. (2000). "Identification of a new type of mammalian peroxiredoxin that forms an intramolecular disulfide as a reaction intermediate." *Journal of Biological Chemistry* 275(27): 20346-20354.

Siegel, R. L., K. D. Miller, et al. (2018). "Cancer statistics, 2018." *CA: a Cancer Journal for Clinicians* 68(1): 7-30.

Silva, G. M. and C. Vogel (2016). "Quantifying gene expression: the importance of being subtle." *Molecular Systems Biology* 12(10): 885.

Sirota, M., J. T. Dudley, et al. (2011). "Discovery and preclinical validation of drug indications using compendia of public gene expression data." *Science Translational Medicine* 3(96): 96ra77-96ra77.

Smits, W. K., O. P. Kuipers, et al. (2006). "Phenotypic variation in bacteria: the role of feedback regulation." *Nature Reviews Microbiology* 4(4): 259-271.

Somlyai, G., T. Q. Collins, et al. (2017). "Structural homologies between phenformin, lipitor and gleevec aim the same metabolic oncotarget in leukemia and melanoma." *Oncotarget*. 8(30):50187-50192.

Stafford, W. C., X. Peng, et al. (2018). "Irreversible inhibition of cytosolic thioredoxin reductase 1 as a mechanistic basis for anticancer therapy." *Science Translational Medicine* 10(428): eaaf7444.

Steger, M., F. Tonelli, et al. (2016). "Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases." *eLife* 5: e12813.

Struhl, K. (1999). "Fundamentally different logic of gene regulation in eukaryotes and prokaryotes." *Cell* 98(1): 1-4.

Stuart, J. M., E. Segal, et al. (2003). "A gene-coexpression network for global discovery of conserved genetic modules." *Science* 302(5643): 249-255.

Subramanian, A., R. Narayan, et al. (2017). "A next generation connectivity map: L1000 platform and the first 1,000,000 profiles." *Cell* 171(6): 1437-1452. e1417.

Sunico, C. R., A. Sultan, et al. (2016). "Role of sulfiredoxin as a peroxiredoxin-2 denitrosylase in human iPSC-derived dopaminergic neurons." *Proceedings of the National Academy of Sciences* 113(47): E7564-E7571.

Swinney, D. C. and J. Anthony (2011). "How were new medicines discovered?" *Nature Reviews Drug Discovery* 10(7): 507-519.

Tan, C. S. H., K. D. Go, et al. (2018). "Thermal proximity coaggregation for system-wide profiling of protein complex dynamics in cells." *Science* 359(6380): 1170-1177.

Tarasova, N. K., A. Gallud, et al. (2016). "Cytotoxic and Proinflammatory Effects of Metal-Based Nanoparticles on THP-1 Monocytes Characterized by Combined Proteomics Approaches." *Journal of Proteome Research* 16(2): 689-697.

Thakur, S. S., T. Geiger, et al. (2011). "Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation." *Molecular & Cellular Proteomics* 10(8): M110. 003699.

Toker, A. (2012). "Achieving specificity in Akt signaling in cancer." *Advances in Biological Regulation* 52(1): 78.

- Ubersax, J. A., E. L. Woodbury, et al. (2003). "Targets of the cyclin-dependent kinase Cdk1." *Nature* 425(6960): 859-864.
- Vaidya, S., E. M. Velázquez-Delgado, et al. (2011). "Substrate-induced conformational changes occur in all cleaved forms of caspase-6." *Journal of Molecular Biology* 406(1): 75-91.
- van Dam, S., T. Craig, et al. (2014). "GeneFriends: a human RNA-seq-based gene and transcript co-expression database." *Nucleic Acids Research* 43(D1): D1124-D1132.
- Van Esbroeck, A. C., A. P. Janssen, et al. (2017). "Activity-based protein profiling reveals off-target proteins of the FAAH inhibitor BIA 10-2474." *Science* 356(6342): 1084-1087.
- Vassilev, L. T., B. T. Vu, et al. (2004). "In vivo activation of the p53 pathway by small-molecule antagonists of MDM2." *Science* 303(5659): 844-848.
- Verma, R., M. J. Rigatti, et al. (2010). "DNA damage response to the Mdm2 inhibitor nutlin-3." *Biochemical Pharmacology* 79(4): 565-574.
- Vigelsø, A., R. Dybboe, et al. (2014). "GAPDH and β -actin protein decreases with aging, making Stain-Free technology a superior loading control in Western blotting of human skeletal muscle." *Journal of Applied Physiology* 118(3): 386-394.
- Visnes, T., A. Cázares-Körner, et al. (2018). "Small-molecule inhibitor of OGG1 suppresses proinflammatory gene expression and inflammation." *Science* 362(6416): 834-839.
- Von Ahsen, O. and U. Bömer (2005). "High-throughput screening for kinase inhibitors." *ChemBioChem* 6(3): 481-490.
- Vora, A., N. Goulden, et al. (2013). "Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): a randomised controlled trial." *The Lancet Oncology* 14(3): 199-209.
- Wagner, E. K. and B. Roizman (1968). "Effect of the vinca alkaloids on RNA synthesis in human cells in vitro." *Science* 162(3853): 569-570.
- Wang, J., Z. Ma, et al. (2017). "Proteome profiling outperforms transcriptome profiling for coexpression based gene function prediction." *Molecular & Cellular Proteomics* 16(1): 121-134.
- Wang, L., L. J. Christopher, et al. (2008). "Identification of the human enzymes involved in the oxidative metabolism of dasatinib: an effective approach for determining metabolite formation kinetics." *Drug Metabolism and Disposition* 36(9): 1828-1839.
- Weigmann, K. (2016). "Releasing the brakes to fight cancer." *EMBO reports*: e201643038.
- Weinert, B. T., T. Narita, et al. (2018). "Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome." *Cell*. 174(1): 231-244.
- Ye, C., D. J. Ho, et al. (2018). "DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery." *Nature Communications* 9(1): 4307.
- Yen, H.-C. S. and S. J. Elledge (2008). "Identification of SCF ubiquitin ligase substrates by global protein stability profiling." *Science* 322(5903): 923-929.
- Youn, H. S., J. Y. Lee, et al. (2006). "Auranofin, as an anti-rheumatic gold compound, suppresses LPS-induced homodimerization of TLR4." *Biochemical and Biophysical Research Communications* 350(4): 866-871.
- Zhan, S., J. Li, et al. (2018). "Quantitative Proteomics Analysis of Sporadic Medullary Thyroid Cancer Reveals FN1 as a Potential Novel Candidate Prognostic Biomarker." *The Oncologist* 23(12): 1415-1425.

Zhu, Y., L. M. Orre, et al. (2018). "Discovery of coding regions in the human genome by integrated proteogenomics analysis workflow." *Nature Communications* 9(1): 903.

Zubarev, R. A., M. L. Nielsen, et al. (2008). "Identification of dominant signaling pathways from proteomics expression data." *Journal of Proteomics* 71(1): 89-96.

Zuberi, K., M. Franz, et al. (2013). "GeneMANIA prediction server 2013 update." *Nucleic Acids Research* 41(W1): W115-W122.