Exploring Cellular Signalling

Perspectives on Phosphoproteomic data use and interpretation in a model of drug resistance

Celine Mulder

The research presented in this thesis was performed in the Biomolecular Mass Spectrometry & Proteomics Group at the Bijvoet Center of Biomolecular Research, Utrecht University, The Netherlands.

ISBN 978-90-393-7151-0
Author: Celine Mulder

Coverdesign and illustrations by Ben Westervoorde

Printed by ProefschriftMaken | Proefschriftmaken.nl

Exploring Cellular Signalling

Perspectives on Phosphoproteomic data use and interpretation in a model of drug resistance

Cellulaire signalering onder de loep
Perspectieven op fosfoproteomic data gebruik en interpretatie in de context van geneesmiddelenresistentie

(met een samenvatting in het Nederlands)

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

woensdag 10 juli 2019 des middags te 12.45 uur

door

Teska Celine Mulder

Geboren op 7 juli 1990
te Hengelo

Promotor: Prof. Dr. A.J.R. Heck
co-promotor: Dr. S.M. Lemeer

-- The truth is never simple and rarely pure --

Oscar Wilde - The importance of being Earnest

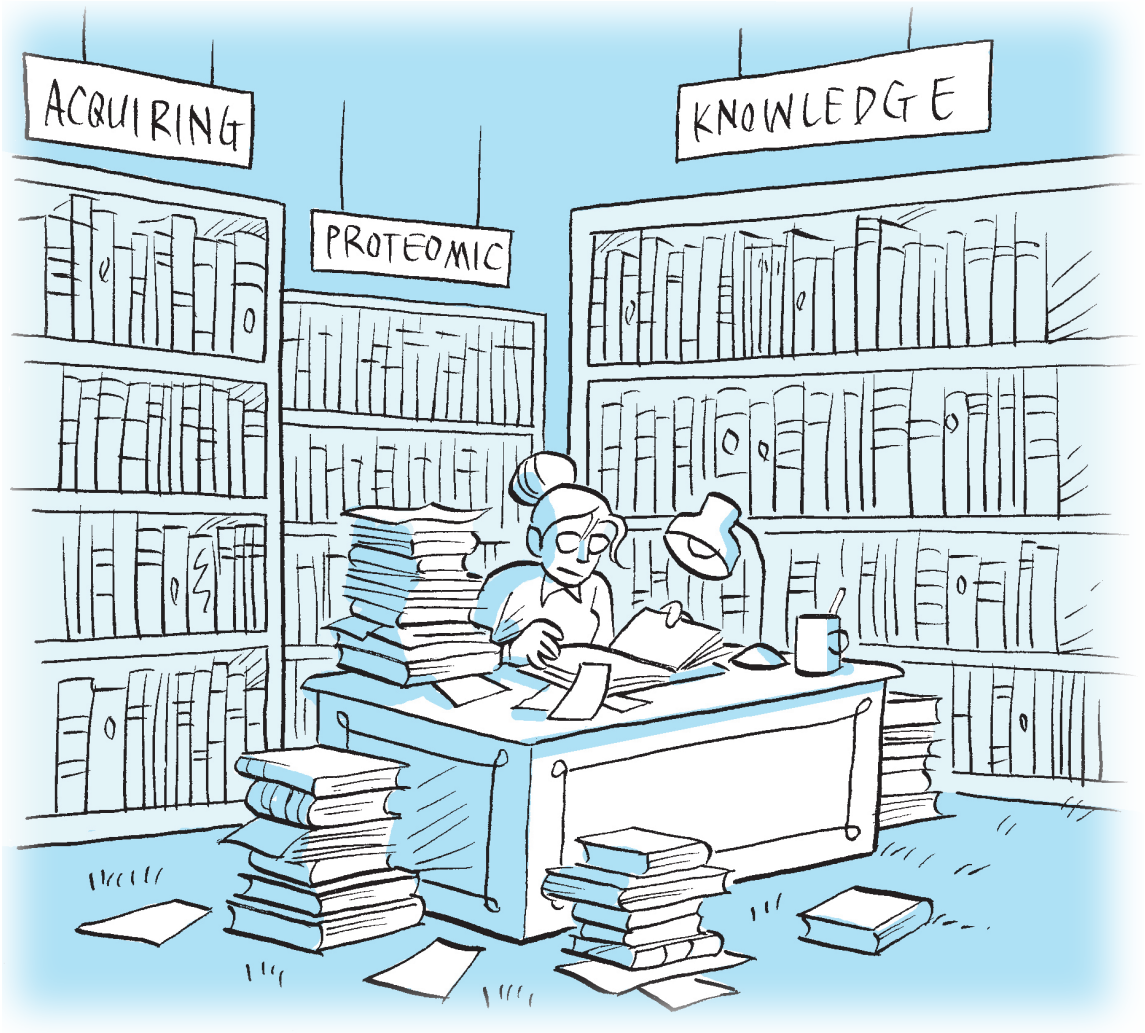
Content

Part 1 - Acquiring Proteomic Knowledge

Part 2 - Increasing Proteomic Depth

Part 3 - Studying Adaptive Resistance

Part 4 - Synopsis


ACQUIRING
PROTEOMIC
KNOWLEDGE

Part 1

Acquiring Proteomic Knowledge


TAMOXIFEN
RETIREMENT HOME

CHAPTER ONE

Proteomics as studied by Mass Spectrometry

Celine Mulder

ONE

Proteomics studied by Mass Spectrometry

After the field of genomics rose during the last decades of the previous century, the term ‘proteome’ was introduced in 1995 [1]. Now, with the information of the genome at hand, research moved towards studying the products of these genes to expand the knowledge about their biological function and to use information of proteins in the attempt to explain phenotypes that could not be explained by genomics alterations.
The word ‘proteome’ refers to the large scale description of an entire set of proteins produced by an organism, and can most robustly be analysed by mass spectrometry (MS). The origin of the mass spectrometer lays in the work of J.J. Thomson, for which he received the Nobel Price in 1906. Despite this success, it took almost 50 years before MS was going to be used to measure organic compounds, including amino acids and peptides. Currently, the entire proteome has been mapped for various organisms, with humans by far being studied the most: following the iconic ‘human genome’ paper, published in 2001 [2], more recently in 2014, the first drafts of the ‘human proteome’ have been published [3,4].
Mass spectrometry is becoming more and more popular in biological research, creating enormous quantities of data to be analysed. As to give an idea of the magnitude, a single run of MS analysis can result in several gigabytes of data containing several hundreds of thousands raw spectra. Within a single biological experiment this can add up to several hundred GB of data, resulting in the identification of tens of thousands proteins and (modified) peptides. Meanwhile, as mass spectrometers keep becoming increasingly sensitive these numbers will only go up. Consequently, one of the biggest challenges in this, relatively young, omics field is the downstream analysis and processing of this large body of information. Extracting true and, most of all, biologically relevant information from these large amounts of data remains a challenge, and it is a vital, but highly underestimated, component of the MS workflow.

Choose your flavour: bottom-up, middle-down or top-down

Proteins can be analysed by MS using various approaches. A popular method of choice is the bottom-up approach, also called shotgun proteomics, where proteins of interest are digested into peptides which are subsequently analysed in the MS individually. In the last step, proteins are identified by computational modelling based on the measured peptides [5]. This method is highly suitable for high-throughput analysis of up to thousands of proteins simultaneously. By digesting proteins into peptides, dynamic range issues due to differences in protein sizes no longer exist and complex 3D structures are dismantled, therefore no longer interfering with detection. For bottom-up protein analysis, generated peptides can be analysed used using data-dependent acquisition (DDA), data-independent acquisition (DIA) [6] or targeted MS methods [7].
Unfortunately, by using a bottom-up approach valuable information about secondary and tertiary characteristics of proteins will be lost. In order to maintain this information, middle-down proteomics can be used [8]. In this case, a protein is only partially digested, making it possible to preserve parts of its shape and details, including co-existence of post-translational modifications (PTMs) and the identifications of dif-

ferent proteoforms.
If complete analysis of a protein is desired, top-down or native proteomics is a possibility, where undigested proteins are analysed in very low complex samples [9]. Using top down MS, denatured but undigested proteins are studied, making it is possible to get near to 100% sequence coverage and detailed information about PTMs. Native MS is used to study proteins in their 3D form, thereby enabling study of tertiary protein folding and dynamic protein-protein complex formations.
The trait-off in choosing which approach to take is a matter of complexity: the more proteins you would like to analyse in a single experiment, the less information you will get per protein, and visa versa.
Since all of the analyses in this thesis are done with a bottom-up approach using DDA, I will not discuss further middle-down, native, DIA or targeted MS. For this, I refer to some excellent reviews that have been written about these topics [5,7–9].

A Shotgun-Proteomics Workflow: From cell to spectrum

Harvest and digestion

The first step in processing a sample for MS analysis is the extraction of proteins from cells by destroying the cell membrane and solubilising the proteins in lysis buffer (Figure 1). Within a cell, proteins are well organised into their respective cellular compartments, according to their function. These different compartments require different shapes and characteristics of the proteins, for example differences in hydrophobicity. Consequently, proteins behave differently in terms of solubilisation if all lysed in the same buffer. Therefore, with every buffer, there will be a trait-off to which proteins will solubilise best. An example for this are, the often very hydrophobic, transmembrane proteins. Their hydrophobic nature makes them reside inside cellular lipid bilayers, making them difficult to solubilise and digest in standard lysis buffer. Several studies focussed specifically on the analysis of this difficult to detect class of proteins, mainly by optimising the composition of the lysis buffer [10–12]. Besides the buffer, several other factors during lysis can be varied, ranging from the method of lysis (e.g. bead-beating, scraping or sonication) to the removal of contaminants (e.g. lipid, salt or nucleotides like DNA and/or RNA). Adequate optimisation of the lysis protocol can, on its own, greatly influence the depth of a proteomics experiment [13].
The next step in the process is protein digestion (Figure 1). Using an endoprotease, the now denatured proteins are being cleaved into easy to detect, short peptides. The most used protease in shotgun proteomics is trypsin. This protease cleaves on the carboxyl terminus of lysine and arginine, creating peptides with a desirable length and charge, favouring MS detection [14]. Because of these characteristics of the produced peptides, and because the enzyme is relatively cheap, most bottom-up studies in literature have been performed using trypsin. This however, creates a bias within the available data. Some proteins have large parts with either too many or too little lysine and arginine residues, and will therefore be forever missed by the use of trypsin only. As a workaround, several studies have shown that the use of alternative proteases, like AspN, GluC, LysC etc., results in the appearance of complementary parts of the proteome [15,16].

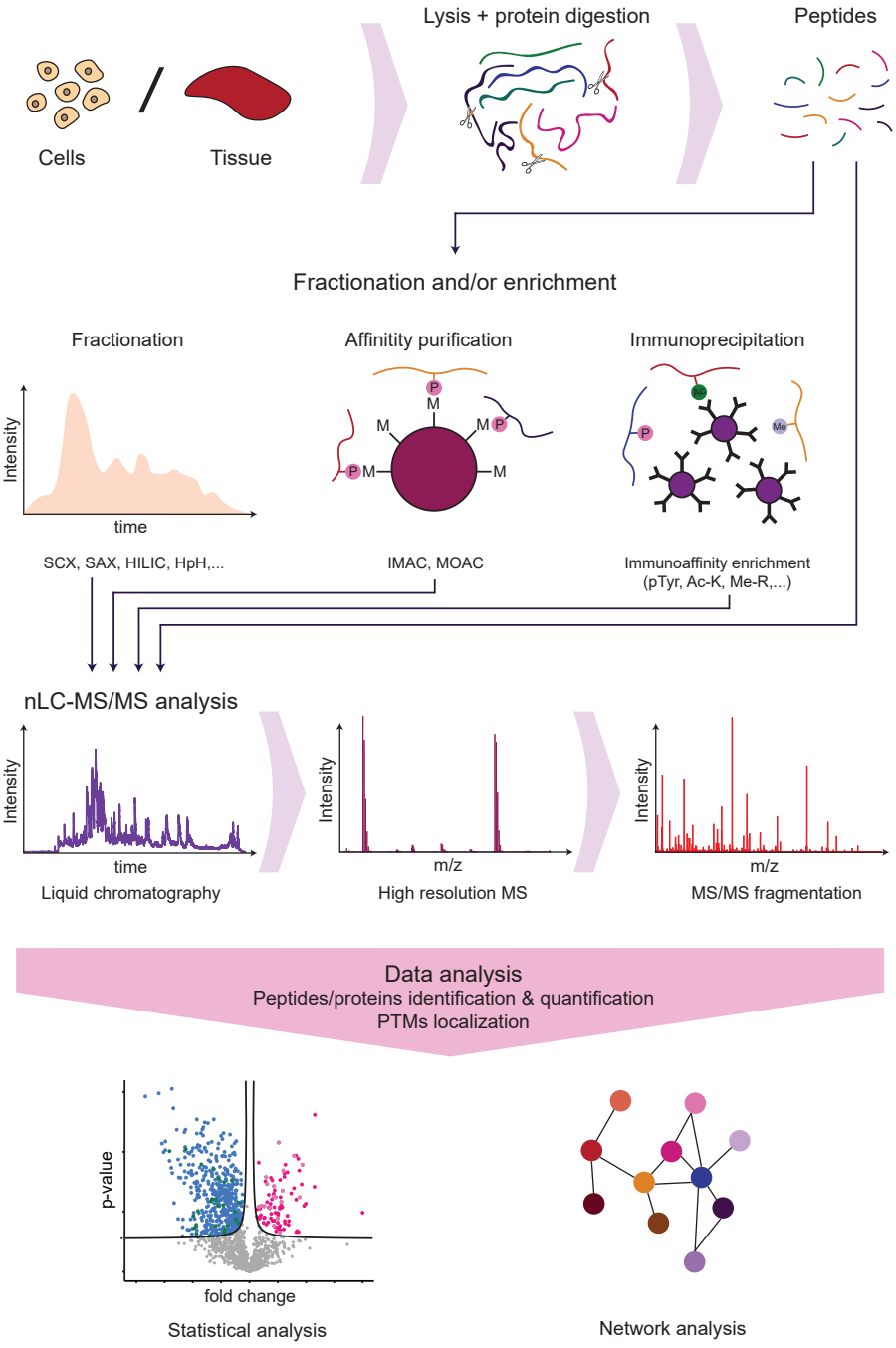


Figure 1 – Standard Proteomics Workflow. Based on Alterlaar *et al* [5], adjusted by Clément Potel.

Quantification – approaches and considerations

The mere presence of proteins within samples is often lacking the desired information. Quantitative data on the other hand is more revealing, allowing the researcher to asses whether proteins change their abundance under certain conditions compared to a control state. Mass spectrometric measurements are not quantitative; due to distinct chemical properties, different peptides behave uniquely within the mass spectrometer. Therefore, it is not possible to compare different peptides with each other. However, by using either label free quantification or a labelling strategy, it is possible to compare the same peptide between different samples (Figure 2)[17]. It is possible to label either proteins (e.g. with SILAC [18,19]) or peptides (e.g. with tandem mass tags (TMT) [20] or dimethyl labelling [21]) with stable isotopes. The main advantage of labelling approaches is the control over a large part of variation induced by (additional) sample preparations (e.g. due to lysis, enrichment or fractionation) or during MS analysis. Secondly, with labelling approaches it is possible to multiplex, and win hours of MS time by analysing several conditions within a single MS run. In addition, multiplexing decreases the amount of missing values, thereby increasing the amount of quantified peptides.

With the use of stable isotope labelling with amino acids in cell culture (SILAC) variations that are introduced as early as lysis or digestion can be taken into account for, since labelling was performed during cell culture. In 2008, this labelling technique was taken a step further by the development of labelled *in vivo* models, like the SILAC mouse [22] and SILAC fly [23]. SILAC is highly accurate, but limited to a small number of amino acids that can be used for labelling. Furthermore it increases MS1 complexity, thereby decreasing the number of identifications per run.

Labelling before protein extraction is not always a possibility, for example when samples are originating from patients or if cells are difficult to grow, for example in the case of neurons or other primary cells. In these cases, peptides can be labelled after extraction and digestion, for example by TMT or dimethyl labelling. In contrast to SILAC labelling, TMT labelling is measured at the MS2 level, therefore MS1 level complexity does not increase. Although this will increase the number of identifications, ratio compression causes a lower quantification accuracy. The latter can be solved by using MS3 measurements, however this is only possible on the latest generation of instruments and has therefore thus far been less accessible by the majority of the scientific community. Further disadvantages of labelling methods are the general high costs and the addition of several steps of processing in the protocol. The latter can possibly induce unwanted modifications, loss of material and extra variation between samples.

If labelling of samples is not desirable, label free quantification (LFQ) can be used [24]. In a LFQ approach, samples to be compared are processed and measured separately from each other, after which the resulting spectra are combined during the database search. The advantage of LFQ methods is that no additional sample handling is necessary, making it fast and cheap, and decreasing possibilities of sample loss and modifications. The disadvantage is that the sample preparations and the MS runs need to be very consistent over different samples and conditions, which is very difficult or even impossible when samples are not harvested simultaneously

or when a fractionation step is included. In addition, LFQ is highly affected by the presence of missing values due to the stochastic nature of peak picking in data dependent acquisition.

Enrichment and fractionation

Although mass analysers have become faster and increasingly sensitive, detecting all generated peptides is still impossible. Additional sample preparation can be performed, either if a specific part of the proteome is of special interest, or if the maximum depth of the proteome is required.

Because of a large dynamic range within proteome samples, low abundant peptides or peptides with low stoichiometry are easily missed in MS analysis. Therefore, it can be beneficial to use adapted sample preparation workflows. By focussing on a

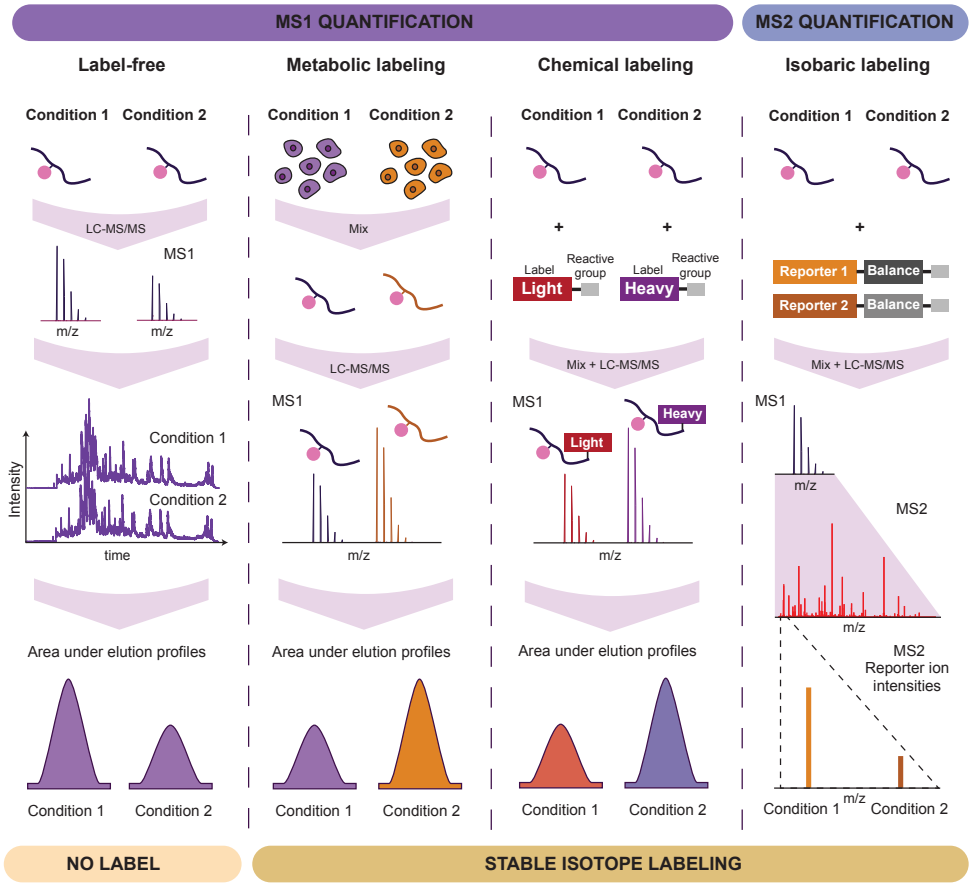


Figure 2 – Quantification Strategies. Based on Riley and Coon [17], adjusted by Clément Potel.

specific part of the proteome the mix of peptides becomes less complex and easier to analyse. Examples include immunoprecipitation [25], subcellular fractionation [26], extraction of a specific protein class, like kinases [27] or the enrichment of a post-translational modification (PTM) [28].
In contrast, it can be desired to expand the view of the proteome as much as possible and reach the maximum depth of the analysed biological model. For this, fractionation can be used, after which different fractions are analysed separately by MS. Different fractionation methods can be combined with MS, for example size exclusion chromatography (SEC) [29], strong cation-exchange (SCX) [30] and high-pH fractionation [31], separating the peptides based on different characteristics (size, charge or hydrophobicity, respectively). After collecting separate fractions, the use of a concatenation strategy decreases the amount of fractions, thereby decreasing MS running time, without losing the power of the fractionation [32].

LC-MS/MS analysis
After all desired sample preparations, in the final step peptides are separated using liquid chromatography (LC) prior to analysis in the mass spectrometer. Reversed phase chromatography is the most widely used separation method prior to MS analysis. In this method the peptides are separated based on their hydrophobicity by using a C18 column and a gradient of increasing organic modifier. Different gradient lengths can be used: the longer the gradient, the more peptides are separated and therefore a better resolution per peptide. The length of the gradient together with the flow rate determines the peak shape and the performance of the separation, thereby influencing the MS1 and MS2 intensity of the measured peptide.
After peptides elute from the LC system, they are brought into the MS and given a charge by electrospray ionisation [33]. Subsequently, the ions are tunnelled into the instrument through a series of lenses before they are measured by a mass analyser, e.g. by time-of-flight (TOF), an Orbitrap [34] or a linear ion trap [35]. In order to gain sequence information, peptides can be fragmented by using different strategies, e.g. CID [36], HCD [37], ETD [38] or EtHCD [39]. Within the latest type of hybrid instruments, a combination of fragmentation and mass analysers can be used in parallel [40].

Database search and statistics
After the peptide spectra are acquired in the mass analyser, the measured peptides are identified by an automated and high-throughput database search. In order to identify the peptides, the experimental acquired fragmentation spectra are compared to theoretical mass spectra within a database of *in silico* digested proteins, using preset parameters like protease, taxonomy and fixed and variable peptide modifications [41]. After identification, the search algorithm will calculate a peptide score to control for false discoveries, based on aspects including number of identified fragments, the probability of a match occurring at chance and the size of the database used. Next, a score cut-off to control for false and true matches can be determined by the False Discovery Rate (FDR), which is calculated by a second search against a decoy database containing random sequences [42]. Different search algorithms exist, using

different approaches for identification and scoring of the acquired mass spectra. Two often used examples of search algorithms are Mascot [43] and Sequest [44].
When the database search is complete, the data is ready for further processing. Before the actual data processing can be done (see following section), some pre-processing steps are necessary, including normalisation of the data, filtering the data for known contaminants and considering missing values. The latter are caused by stochastic variations between the different samples within one experiment. Although a complete dataset is necessary for multivariate analyses, such as a principle component analysis (PCA), as described below, it is unnecessary to discard all incomplete data entries. Peptides that are measured in only some of the samples will still provide valuable information. However, it is considered good practice to either filter the data for a minimum amount of detected peptides or to impute the missing values.

Data Processing – Let the data speak

After the acquisition of the data, the data analysis and visualisation can be a laborious and complex process. Every researcher, lab and even project has its own ways of dealing with this. Some pick their favourite protein, and discard the rest of the data, others write their own code and create maps of their identified proteins completely suited to their project [26]. Many pieces of software, both in terms of statistics as well as visualisation, exist to assist researches in the process of clustering and organising their data. However still, manual evaluation, reading literature and using your own experience on the subject and common sense will eventually be the most beneficial strategy.
What matters most in answering the question how to analyse and display the data is the research question that led to the existence of the current piece of data. Shotgun proteomic experiments can be designed as either a hypothesis-generating or hypothesis-driven experiment. In the first scenario, little is known beforehand on which direction the data will take. Ideally, not the researcher, but the data will lead the data processing in a specific direction. Often, different types of analyses need to be tested to discover where the most valuable information lies within the data and what approach will provide the most informative visualisation. In an oppositely, hypothesis-driven approach, a clear hypothesis is stated before the experiment starts, and the MS analysis is used to either confirm or reject this hypothesis. In this case data processing is simpler, since it is more clear where to look for. Still, it can be of great value to treat a dataset originating from a hypothesis-driven experiment as it was a hypothesis-generating experiment. Because of the depth of mass spectrometric data, possibly new, unknown and unexpected findings can be gained by looking at any dataset with an open mind.
In the last section of this part I will discuss some common and less common alternatives for data processing that I am familiar with, for this list is far from complete [45] (Figure 3). I will discuss several ways of moving forward from this point on, as it is in my personal opinion, one of the most challenging and underestimated parts of the entire mass spec workflow. However keep in mind while reading that the preferred data processing for every dataset is unique, and that finding the right approach requires a recurring process of fine-tuning and trial and error.

Filtering

An important choice to be made before any type of analysis is which proteins or peptides to include. Often, a large group of data entries lack an immediate interest, for example proteins that do not change in abundance between samples, with low statistical relevance or with too many missing values. Therefore, filtering for the most significant and biological relevant events often helps subsequent data analysis (Figure 3). For this purpose, most often used are either a Student T-test or an analysis of variance (ANOVA) to select for the proteins or peptides of statistical interest.

Although filtering is necessary to some extent, it should be done with great caution and awareness of what is being in- or excluded. Important in this process is the decision of cut-offs for *p*-values and ratios, which can be set in an arbitrary or logical manner. An example of the latter is the use of an FDR based cut-off for both *p*-value and ratio, called Significance Analysis of Microarray (SAM), or S0 curve, in which for proteins bearing a higher ratio, a lower *p*-value is allowed [46]. Although useful, it might be a too stringent cut-off filter, a reason why many researches still use the arbitrary cut-offs of an absolute $\log_2$ ratio of >1 and a significant *p*-value of <0.05.

An example of where unobservant filtering can go wrong is the lack of consideration of *which* proteins are assessed. Some proteins are in need of an enormous fold change in order to make a biological difference inside a cell, where others can have an amplifying effect with only a very minor fold change. These different type of proteins should be treated differently while filtering before and during data analysis. The same is true for the case if many proteins within a single pathway all change with a small ratio, but with high confidence. Because a large group of proteins attribute to the same hypothesis (i.e. this pathway is regulated), even though with a small change, it adds to its biological relevance (further discussed in “Enrichment and Networks” section). If the data would have been filtered blindly for individual proteins with large fold changes, such an observation goes completely unnoticed. It is therefore considered good practice to filter multiple times during analysis, for criteria of in- or exclusion of a protein might change after certain observations are made along the way.

Clustering

The favourable first step in data-processing is to get a global overview of trends within the data. This can best be done by clustering the data as a whole, instead of the examination of individual proteins or peptides. By using unsupervised machine learning, different algorithms will find patterns and trends in the data, without the use of prior classification [45] (Figure 3). As an additional quality control, it is then possible to examine if the beforehand known classifications are indeed found by the unsupervised clustering (e.g.: do replicates cluster together?).

An example of an unsupervised clustering is a Principle Component Analysis (PCA), in which all data is compressed in a set of components that explain the variation between the different samples [47]. A PCA plot is an easy way of detecting batch effects between, for example, moment of harvesting or biological and technical replicates. Using this analysis, the most global trends of the entire dataset can be seen within a single graph, immediately providing valuable information about where the largest

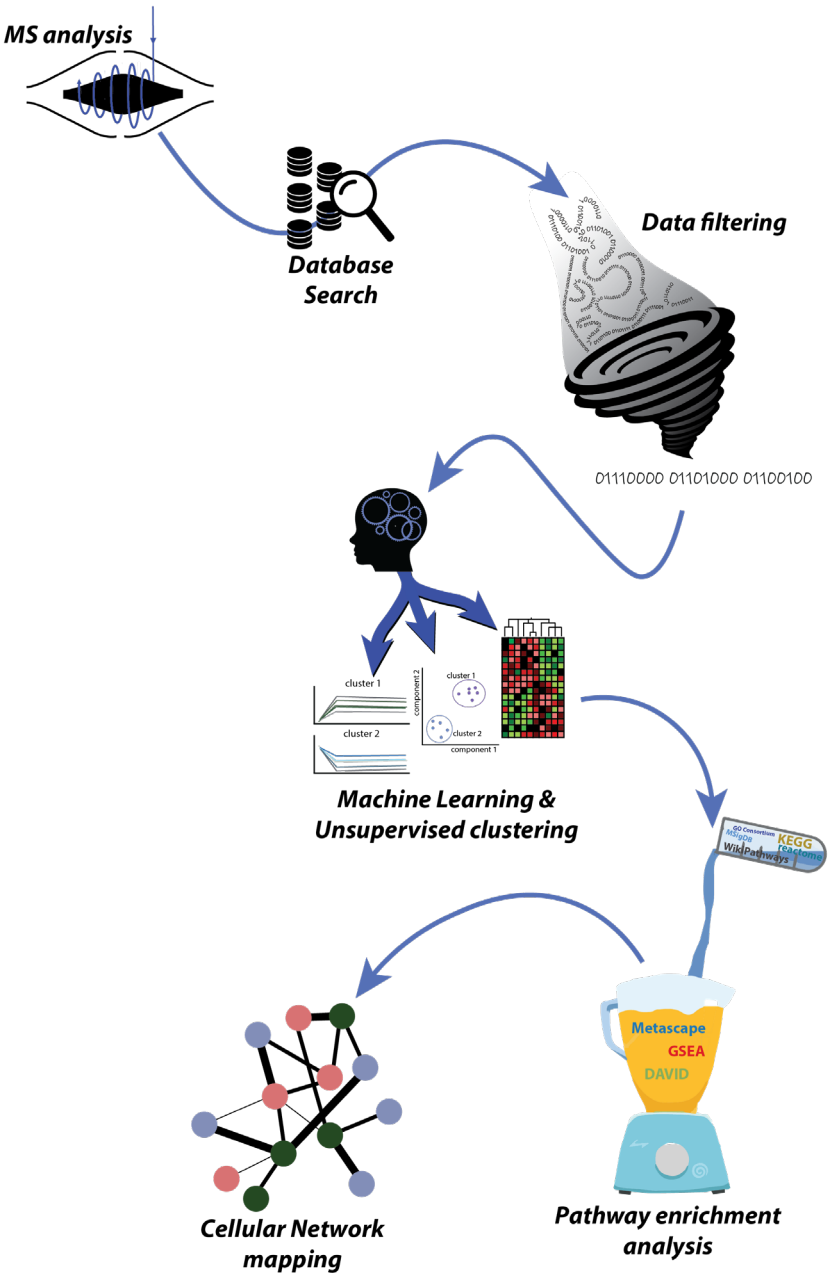


Figure 3 – Dataprocessing of proteomics data

variation exists within the data.
Next, a more detailed analysis can be done in the form of hierarchical clustering [48,49]. In this type of analysis, the data is organised by building a hierarchy of clusters and therefore grouping either samples, proteins or both based on similarities. Depending on the data, different choices can be made considering the parameters the clustering is based on, e.g. distance and linkage type. A similar type of clustering is by k-means clustering [50]. In contrast to hierarchical clustering, in k-means clustering the data is clustered in a beforehand set number of clusters. To determine the right number of clusters, a visual inspection of a hierarchical clustering can be used or the number can be determined mathematically, for example by the inspection of percentage of variance explained as a function of number of clusters (i.e. the elbow method).

Enrichments and Networks
The regulation of cellular pathways is based on the cooperation of multiple proteins and processes. Therefore, in the study of biological processes, minor regulation of a group of proteins that work together in the same pathway might be more informative than a large regulation of a single protein. In addition, since a single protein can function in multiple pathways, it is difficult and maybe even impossible, to assign the regulation of an entire pathway to the regulation of a single protein. Therefore, analysing a large dataset, such as the result of a shotgun proteomics experiment, is mostly about finding regulated pathways and trends, rather than showing the regulation of individual events. Moreover, in order to combine or compare multiple datasets, data analysis has to be robust. Data originating from different experiments is well able to display regulation of the same pathway, resulting in the same conclusions, despite that each dataset consist of a completely different set of proteins. Therefore, after filtering and clustering the data, the discovery of an over-representation of pathways amongst regulated proteins can be used to find their connecting link, resulting in a conclusion about the biology contained within the experimental data (Figure 3).
In 2005 the Gene Set Enrichment Analysis (GSEA) was developed, providing a tool to identify alternatively regulated biological processes (called gene sets) between different samples, based on the regulations of clusters rather than individual proteins [51]. Other tools, such as Metascape [52] or DAVID [53], calculate an enrichment score, usually depicted as a *p*-value. The latter is done by the comparison of the selected protein cluster to a background of proteins (usually all proteins measured in the experiment at hand, or the entire genome/proteome) and with the subsequent identification of over- or underrepresentation of proteins belonging to the same pathway. A great variety of tools exist providing algorithms to identify enriched biological processes [54]. Nonetheless, the success of finding the desired biological meaning within the data is highly dependent on the availability of literature and knowledge concerning the considered proteins. Therefore, it is not always possible to explain the clustering of certain proteins, and what this might mean for the biological implications of the experiment. However, with the increasing availability of knowledge about the (human) genome around the year 2000, multiple consortiums started to collect and add biological functional data to all annotated genes. Kyoto Encyclopedia of

Genes and Genomes (KEGG) came to existence in 1999 and consists of collections of graphical representations of cellular pathways, including functional information about the interactions between proteins within these pathways (i.e. inhibition, activating, PTMs etc.) [55]. These pathways can be browsed to find your protein of interest and see its function and adjacent proteins and molecules. Around the same time, in 2000, the Gene Ontology (GO) Consortium was established [56]. GO ontologies are based on conservation of genes and their biological roles between organisms. In this case, genes are not grouped in pathways like in KEGG, but in either biological function (e.g. DNA replication or cadherin binding), cellular components (e.g. cytoplasm or nucleolus) or molecular functions (e.g. DNA helicase or kinase). A third pathway database is the Reactome Knowledgebase, which was initiated in 2003 [57]. On the Reactome website it is possible to browser through interlinked cellular pathways, which consist of published and curated information. Furthermore, together with the development of the GSEA in 2005, the Molecular Signatures Database (MSigDB) was established [58]. This database is a collection of published experimental data and previously collected databases, including GO terms and KEGG pathways. Lastly, an open access database is WikiPathways, based on the same principles as Wikipedia, where all users are able to edit and update pathways [59].
Besides identifying enriched pathways or processes, the data needs to be visualised in order to show the results in a paper or presentation. Several tools have been developed to help the researcher organise their data into understandable visualisations. However, it remains a subjective and biased undertaking, subdued to manual interpretation of the data. One of the tools available is STRING database, constructing a protein network of the proteins given as input, based on protein-protein associations derived from public resources, including experimental data, public databases, text mining and data regarding ontology [60]. In addition, the tool provides pathway enrichment analysis and is able to divide the proteins in K-means clusters based on connectivity. Another useful visualising tool is the free software package Cytoscape [61]. Different plugins can be installed within the software to augment data analysis, e.g. PathVisio [62], PhosphoPath [63], GO enrichment analysis and STRING. An example of a commercial software is Ingenuity Pathway Analysis (IPA) from Qiagen Bioinformatics, enabling the researcher to perform various types of enrichment analysis on their data and immediate visualisation of the results.

Special kids on the block: Phospho-proteomics

Cellular pathways are regulated on several different levels, although foremost by post-translational modifications (PTMs), such as phosphorylation. Using a variety of analytical tools it is currently possible to analyse peptides containing a PTM on a large scale using MS. Information about PTM regulation within an MS experiment adds an additional layer of complexity to the subsequent data analysis, since the addition or the removal of a PTM on a protein can have many different biological functions. Although this extra layer of information can discover important cellular processes, the main difficulty is that for most PTMs their biological implications are not known. Therefore, the data analysis of PTM-centric data needs an alternative approach compared to protein-centric data. Since phosphorylation is the most studied

PTM, and all of the data presented in this thesis are considering phosphoproteomics, the following section will only discuss analysis of this particular PTM.

Similar to protein-centric data analysis, several databases exist collecting publicly available data regarding phosphorylation, including PhosphoSitePlus [64], PHOSIDA [65], Phospho.ELM [66], HPRD [67,68] and dbPTM [69]. Although the identification of a large amount of phosphorylation events has been documented within these databases, only a minor part has been assigned to a biological function. Despite the lack of this critical piece of data, it is possible to infer biological meaning from phospho data by data mining and usage of knowledge about kinase-substrate interactions and motif analysis. Furthermore, this data gap can be overcome by analysing trends in phospho data instead of examining individual phosphosites, similar to pathway enrichment analysis in proteomics data. An example of this is the analysis of the enrichment of sequence motifs that can be linked to a specific kinase to indirectly identify its activity. The identification of a sequence motif can, for example, be done by the online tool MotifX [70]. Further examples of algorithms using this approach are KSEA (kinase-substrate enrichment analysis) [71] and IKAP (Inference of kinase activities from phosphoproteomics) [72]. Furthermore, KinomeXplorer [73] is a platform that combines the database NetPhorest [74] and a tool called NetworKIN [75] to predict kinase-substrate pairs.

An alternative method to analyse phospho-proteomics data is to directly analyse phosphoproteins and their interactions. An example of a useful algorithm is PHOTON (PHOsphoproteomics dissecTiOn using Networks) [76]. In this approach, phosphosites are being transformed into regulations on the phosphoprotein level, making the data less complex and easier to interpret. Another tool for pathway analysis and visualisation is the Cytoscape plugin PhosphoPath [63]. This application is able to perform enrichment analysis based on several databases and visualise interactions between both phopsho proteins as well as kinase-substrate pairs. In addition, the tool can be used to visualise time course data in an orderly fashion.

ONE

Drug resistance in cancer

Cancer therapy

Cancer treatment through the ages

The mentioning of cancer as a disease goes back to the time of the ancient Egyptians, dating back to 1500 BC, where the disease was treated by either palliative methods or surgical removal. The word “cancer” originates from the Greek word “karkinos”, meaning “crab”, as described by Hippocrates around 400 B.C. “Karkinos” was changed into the Latin word for crab, cancer, in 47 A.D. by the philosopher Celsus. Over the next centuries, researchers remained to be intrigued by the disease, as many different hypotheses and theories have come to pass in the attempt to explain the phenomenon, including imbalance in bodily fluids, malignant cells and the believe of the spread of cancer cells via parasites. The current knowledge of how cancer originates and develops is based on the discovery of the helical structure of the DNA as discovered by Watson and Crick, and the subsequent discoveries of genetic mutations, oncogenes and tumour suppressor genes in the 1970s.

The most ancient, but still most often used method of cancer treatment is surgical removal. Unfortunately, tumour metastasis can cause impartial removal of a tumour and reoccurrence of the disease after surgery. This was already stated in 1829 by the gynaecologist Joseph Récamier, and later further described by the surgeon Stephen Paget around 1900. Additionally, tumours existing in a non-solid state, like leukaemia, cannot be removed by surgery. One of the first alternative cancer treatments that came into existence was therapy by radiation. Radiotherapy could be developed due to the invention of X-rays at the end of the 19th century, although subsequently, researchers found out that radiation also induces cancer. Following, in the beginning of the 20th century, chemotherapy was developed. The observation of altered bone marrow cells in first world war soldiers that had been exposed to nitrogen mustard gas contributed to this discovery, making it the first chemotherapeutic agent, treating lymphomas.

As the field of medicine evolved rapidly over the last decades, so did the field of cancer therapy. Due to the increasing understanding of the body, other methods than direct targeting of tumour cells could be developed, of which a great example is immunotherapy. Already in the late 19th century the observation of the remission of tumours after bacterial infections gave rise to the idea to use the patients own immune system against the cancer cells. Unfortunately, the idea did not gain ground until the 1970s because of lack of knowledge and fear of unnecessarily infecting cancer patients. Nowadays, there is a large body of research addressing the priming of immune cells *ex vivo*, the discovery of tumour antigens and the development of cancer vaccines [77]. Recently in 2018, the Nobel Prize in Physiology or Medicine has gone to James P. Allison and Tasuku Honjo for their work in cancer immunotherapy.

Another concept that has changed in the last few decades is the moving from the brutal force of killing all fast growing cells by chemo- or radiotherapy, towards the more subtle, targeted approach, with methods that solely target the cancer cells, and do this with great specificity. Two highly successful examples of this approach are Trastuzumab (Herceptin), a monoclonal antibody against HER2 used in breast can-

cer, and Imatinib (Glivec), a small molecule against BRC/ABL used in chronic myelogenous leukaemia (CML) (further discussed in the "Targeted Therapies" section). Apart from the development and research of individual therapies, almost every study considers combining several treatment strategies. An illustrating example is a trial describing the combination of a vaccine against a p53-specific antigen and subsequent chemotherapy, where the majority of people that received the vaccine showed great response to second-line chemotherapy, whereas non-vaccinated patients did not [78].

Due to an explosive amount of cancer research and the subsequent accumulation of new treatments over the last few decades, cancer survival rates are now rising. According to the American Cancer Society [79], the estimated drop in overall cancer mortality over the last 20 years is 26%. Though, contributing to this number are also the improved detection methods, which result in early detection and therefore better survival chances.

Researches are divided over the question if we will ever cure cancer. Although for some types of cancer survival rates are close to 100%, for example for prostate cancer, for others survival is decreasing, for example for uterine corpus tumours. Although cancer treatments will advance further in the next years, it will remain an extremely challenging disease to treat, due to the large variation of cancer types that exist, inter-patient variation and the development of resistance.

Targeted therapies

Since the first targeted therapy was approved for use in the clinic in 1970 (Tamoxifen, used in breast cancer), many more have followed. As of today, more than 30 compounds have been approved by the FDA, and approximately another 150 are currently in clinical trials [80]. Targeted therapies distinct themselves from regular chemotherapy by the fact that they target the cancer cell on the cancer specific defect, where chemotherapy targets all rapid dividing cells. The main advantages of this are more efficient ways of abolishing the tumour cells, since not all tumour cells are dividing rapidly, and less side effects, since other (healthy) rapid dividing cells are no longer targeted. In addition, targeted therapies can be designed as such, that they not only target the responsible oncogene but are also specific for the mutated form. Since this mutated isoform is only expressed in cancer and not in healthy cells, the therapy becomes even more cancer specific.

Targeted therapy can be classified into antibody based or small molecule therapy. The main difference is that antibodies bind to cell surface proteins and receptors, whereas small molecules can enter the cells and inhibit a various number of proteins within cellular pathways. One of the main molecular targets in targeted therapy are protein kinases. These molecules are enzymes that can add a phosphate group onto another protein, causing altered regulation in terms of activity, localisation or other effects. Kinases are attractive anti-cancer targets, mainly due to the fact that they are at the start of various survival and proliferation pathways driving cancer growth, making them crucial proto-oncogenes. In addition, inhibition of a kinase in a healthy cells is often not as toxic as in cancerous cells, which are substantially more dependent on the activity of the kinase.

Although successful, many impediments concerning targeted therapy remain to be solved. Where chemotherapy can be considered too a-specific, targeted therapy might be too specific. Due to heterogeneity within a tumour or between patients, adaptation of cells and presence or development of intrinsic and acquired resistance, the remissive effect of a targeted anti-cancer agent often has an unsatisfactory short life span. Although clinical trials show increased progression free survival and overall survival of the patient, this is often expressed in terms of months, after which relapse of the tumour occurs. Though every extended month of a patient's life is valuable, more fundamental and clinical research is necessary to increase the endurance of new and existing treatments.

Resistance – *In Vivo et in Vitro*

In vivo resistance development

Tumours are entities on their own within the body, depicting distinct tumour specific and clonal evolution throughout tumour progression [81]. Based on this, the main hypothesis addressing development of drug resistance is the presence of intrinsically resistant cells within a heterogeneous tumour which are able to evolve into a fully resistant tumour as the disease progresses (Figure 4). Indeed, it has been shown that a high level of intra-tumour heterogeneity correlates with poor prognosis [82]. Following this hypothesis is the assumption that, due to rapid proliferation in combination with a dysregulated DNA damage response, increased mutagenesis will lead to additional mutations. Under the pressure of the drug treatment, this can eventually result in a new, resistant, tumour, described as acquired resistance (Figure 4) [83]. The main difference between intrinsic and acquired resistance is the dwelling time between the onset of treatment and the emergence of the mutation responsible for resistance. A growing body of literature is now studying this dwelling time, or adaptive resistance (Figure 4). During adaptive resistance, cells are able to persist and will display a slow growing phenotype, while no genomic abolitions yet can be found to explain the resistance, eventually enabling them to acquire additional mutations [84]. Several mechanismsv have been described trying to explain adaptive resistance, including epigenetic changes [84] and PTM and protein expression regulations [85–87]. The main interest of this research is to be ahead of the appearance of a novel oncogenic mutation, rather preventing the development of acquired resistance than cure it.

An alternative hypothesis of development of resistance and the progression of tumours is the existence of cancer stem cells (CSCs) (Figure 4) [88]. These cells have all the abilities of normal stem cells, i.e. they are quiescent and capable of self-renewal and differentiation. The main difference, however, is that they will expand uncontrollably, whereas growth and differentiation of normal stem cells is strictly regulated. In addition, CSCs differ from "normal" cancer cells in several ways, including speed of proliferation, resulting in increased resistance to regular chemo- and radiotherapy [89]. As a result, CSCs are able to renew and maintain tumour growth after eradication of the bulk of tumour cells by anti-cancer treatment. In line with this hypothesis, the only way to cure cancer is to target these CSCs, for these will continue to grow if not dealt with. Over the last couple of decades, numerous CSCs

targeting approaches have been developed of which several have shown success in the clinic [90,91].
Besides the various routes cancer cells can take to resist drug treatment, an additional component to consider is the interaction of tumour cells with the tumour microenvironment (TME) [92]. The TME consists of multiple types of cells (e.g. fibroblasts and immune cells), secreted factors (e.g. cytokines and hormones) and extracellular matrix components (e.g. collagen and fibronectin). The interaction between the TME and the tumour serves multiple purposes including progression and differentiation of the disease [93], metastasis [94], influencing the immune response [95] and the success rate of treatment, including the development of resistance [96]. This adds an additional layer of complexity to cancer research, for the TME cannot be ignored, however it is challenging to mimic the TME-tumour interaction properly *in vitro* [97,98].
As several routes to resistance have been described here, most likely development of resistance *in vivo* is a combination of all of the above [83]. This makes the study of drug resistance in cancer technically difficult, and to establish an efficient and favourable treatment covering all above mentioned processes is a highly challenging task.

In vitro resistance development
Modelling a disease *in vitro* in order to study it inside a laboratory is subjected to making assumptions and compromises. Ideally, all factors that can influence the results are controlled, and a single condition is varied to examine the effect. However, the more controlled the experiment is, the more distinct it is compared to an *in vivo* situation. Therefore, deciding on an experiment is the trade-off between simplistic but artificial, opposing to complex but realistic. The same is true in the study of drug resistance, to which various analytical tools exist that can be used, deviating from simple to complex, all with their own losses and gains.
The most simplified way of studying resistance is to knock out or mutate a single gene and subsequently observe the effect on cancer growth and sensitivity to treatment. Mainly the invention of the CRISPR-Cas9 system [99] allows high-throughput screening of single gene influences on cancer biology, resulting in a large body of literature describing the biological relevance of numerous oncogenes and tumour suppressor genes [100,101]. However, the question that remains is how valid this method is for the study of resistance. It has been shown that both cancer progression as well as development of resistance rarely arises upon a single mutation [102]. In addition, non-mutational influences driving tumour progression cannot be studied or even taken into consideration using this system. Although it is quick, cheap and insightful to study the effects of mutating a single gene in isolated conditions, it might be too simplified to explain *in vivo* developments.
An alternative, very often used method is increased exposure of a drug to a sensitive cell line, mimicking the continuous exposure of a drug to a patient during treatment and subsequent resistance development [103,104]. This approach is regularly combined with clonal expansion of the exposed cells, to isolate resistant clones to study them in more detail [105]. The advantage of this method is that the cells themselves can “decide” which resistance mechanism route to take, possibly making it biological

more relevant. This is in contrast with the mutating strategy described above, where cells are being pushed into being resistant by alteration of the DNA by the researcher. In addition, this approach is discovery-based, meaning that no prior knowledge is necessary to produce the resistance. However, one might argue that culturing cells in a 2D environment, without any flow, extracellular matrix or adjacent cells, and in culture medium that is far from physiological [106], this system is highly artificial, possibly resulting in false findings due to technical aspects. However, similar to CRISPR-Cas9, it is cheap, quick, high-throughput, and ethically easy to use.

Moving away from the simplistic 2D models in cell culture, several 3D and *in vivo*-like, models exist, for example organoids and patient-derived xenografts (PDXs). Organoids were developed by the lab of Hans Clevers, in 2008 [107]. These 3D models come to existence by growing stem cells into a self-organising organ in a dish, representing the original architecture and differentiation of the tissue of the patient. They started with the study of the small intestine, but currently it has become an important and much used model system to study a great variety of tissues [108]. Also in cancer research organoids are now being used in order to study cancer development and resistance [109]. Since organoids are 3D models representing an entire organ in a dish, they reside between the fields of *in-vitro* cell culture models and *in-vivo* animal based models. By taking up this particular position, it has both the advantages of being cheaper and less ethically loaded since no animals are used,

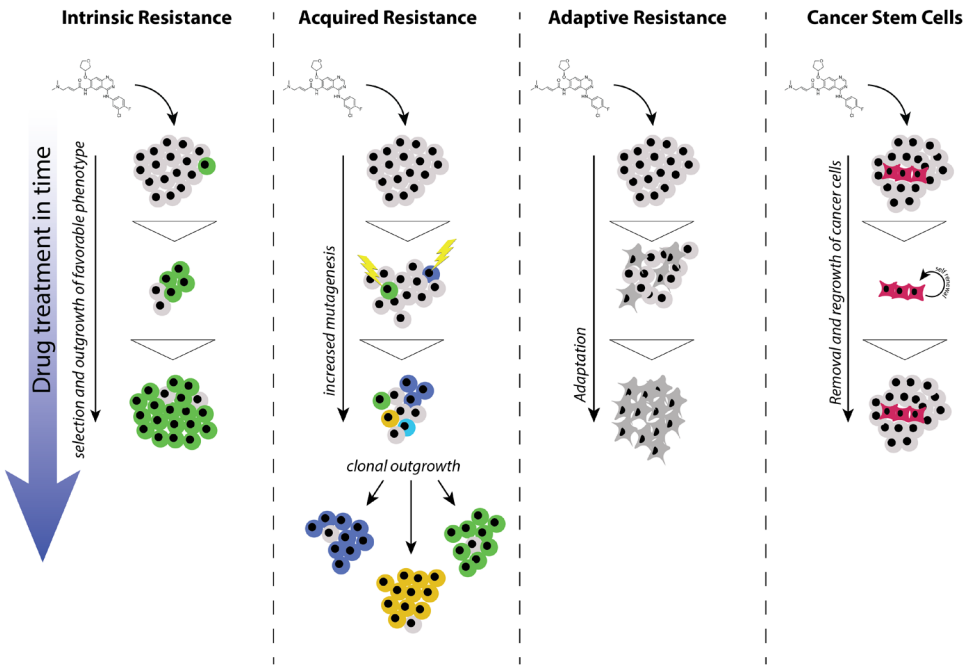


Figure 4 – Resistance Development

while maintaining a 3D structure, making it biological more relevant. The main disadvantage of the use of organoids is the need of a highly equipped lab and experience in the production of organoids.

Models that represent the human *in-vivo* situation the most are animal models. An elegant method of studying human tumours within an animal model is the use of patient derived xenographs (PDXs). PDX models consist of surgical removed human tumour derived tissue, that subsequently is being transplanted into a mouse, creating a cancer model maintaining the patients characteristics, including existing drug sensitivity or resistance [110]. Although the establishment of PDXs can take up to several months, it has been shown to be a beneficial model to study a various amount of cancer types en phenotypes [111]. The main disadvantage of this type of modelling is the use of live animals, which increases the cost, time and ethical weight of the experiments.

In general, both organoids and PDXs are used to extract tumour tissue from a patient and maintain and study the phenotype inside the laboratory. Therefore, it is less suitable to study the transformation of drug sensitive cells towards drug resistant cells, since the modelled tumours represent a specific phenotype (i.e. sensitive or resistant). However, these models are excellent to use as pre-clinical screening to reveal effectiveness and possible biomarkers for resistance [112]. Despite the higher costs and workload, by studying *in vivo*-like models like organoids and animal models before moving towards the clinic, the success rates of expensive and ethically difficult clinical trials can greatly be enhanced, which is shown to be extremely low when only using pre-clinical *in vitro* models that poorly represent human conditions [113–115].

ONE Exploring Cellular Signalling

Although cancer and cancer research is as old as time, many questions regarding this disease remain to be solved. One of the more recent findings is that development of both cancer and subsequent drug resistance is a changeable and dynamic process. This increasing complexity of this pressing problem asks for a system-wide, discovery based approach in order to reveal yet unknown oncogenic cellular mechanisms and new treatment opportunities. A method of choice to analyse cancer models of different complexities is shotgun proteomics. The high-throughput and discovery based approach of proteomics is able to display the regulation of entire pathways on several levels of regulations in a single experiment, without the need of prior knowledge. This allows the researcher to investigate a cancer model as a whole in an unbiased manner, thereby arriving to new and yet unrevealed oncogenic processes, taking the field of cancer research a step further.

In this thesis we have explored short-term adaptive responses of lung- and breast cancer cells in response to targeted therapy with the use of mass spectrometry based (phospho)proteomics. Along the way we came across several technical and biological challenges to solve before arriving at a satisfactory conclusion.

In **Chapter 2** we reviewed several approaches to increase the focus within a MS experiment regarding the discovery of drug function. By zooming in on a specific part of the proteome a clearer picture can be obtained, leading to more concise answers.

In **Part 2** of this thesis, different approaches are investigated in order to increase the amount of data available to increase coverage and molecular detail. In **Chapter 3** different proteases were used to discover understudied parts of the (phospho)proteome. By using complementary protocols for the same biological sample, a larger area of the phosphoproteome could be covered, resulting in increased depth into the phosphoproteome. In **Chapter 4** several publically available phosphoproteomic datasets were re-analysed in order to obtain novel discoveries and to decrease data loss, adding onto existing knowledge. Additional data processing transformed the phosphodata into an easy to use format, enabling the combined analysis of several datasets. Through increasing the amount of data per observed event, differential conclusions could be made in addition to the previous documented conclusions.

In **Part 3** all of these methods were applied to the biological model of resistance to targeted therapy in lung cancer. Sensitive non-small cell lung cancer cells (NSCLCs) were treated with epidermal growth factor inhibitors for different lengths of time and their proteome, phosphoproteome and kinome was explored to explain observed phenotypes. In **Chapter 5** cells were treated for no longer than 5 days, although great phenotypic and morphological differences could be observed. Using a multi-omics approach, the proteomic, phosphoproteomic and kinome landscape of early adaptation of EGFR inhibition was mapped. It was further validated that the reorganisation of the cytoskeleton and increased adhesion rescued the cells of apoptosis upon drug

treatment. The removal of extracellular calcium prevented this, inducing the cells to cease to survive. In **Chapter 6** cells were treated for up to several weeks after which multiple resistant clones could be generated. Extensive phosphoproteomic analysis revealed large homogeneity and cross-sensitivity across the separate clones.

In **Part 4** some words were spend on discussing the most pressing difficulties and the future of the field of (phospho)proteomics and anti-cancer drug resistance.

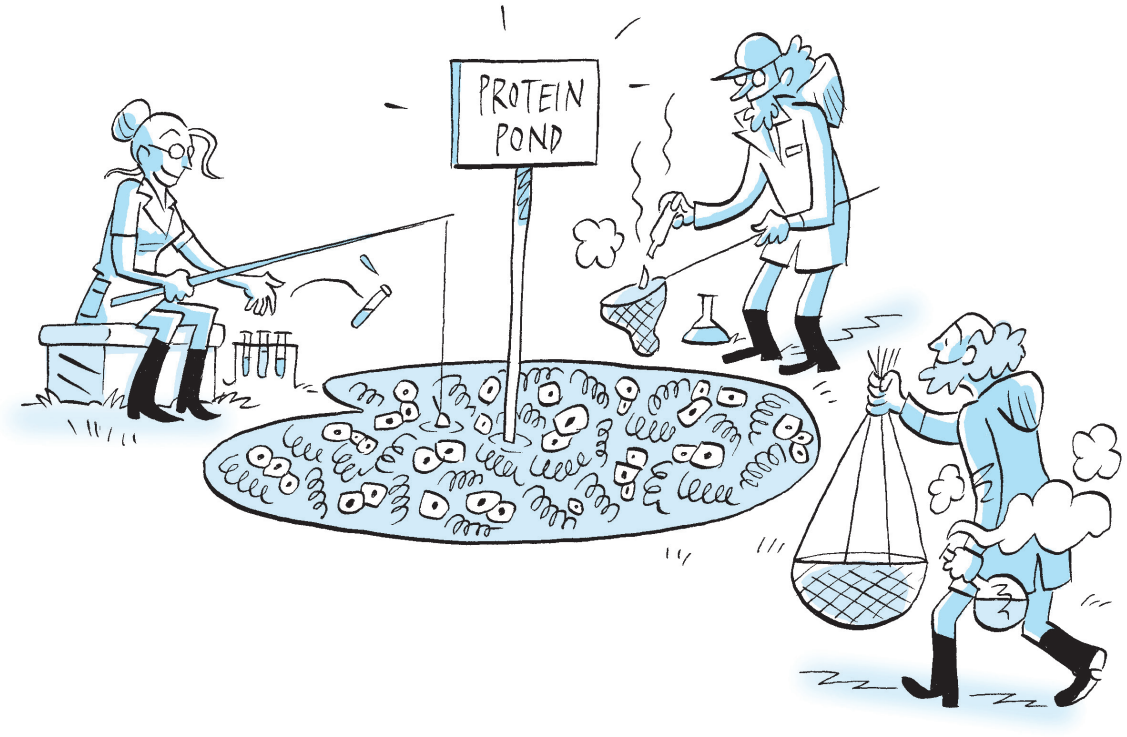
References

1. Wasinger VC, Cordwell SJ, Cerpa-poljak A, Gooley AA, Wilkins MR, Duncan MW, Williams KL, Humphery-smith I, Street V: **Progress with gene-prodct mapping of the Mollicutes : Mycoplasma genitalium**. *Electrophoresis* 1995, **16**:1090–1094.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans C a, Holt R a, et al.: **The sequence of the human genome.** *Science* 2001, **291**:1304–51.
3. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, et al.: **Mass-spectrometry-based draft of the human proteome**. *mass Spectrom Rev* 2013, **32**:5453–465.
4. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, et al.: **A draft map of the human proteome**. *Nature* 2014, **509**:575–581.
5. Altelaar AFM, Munoz J, Heck AJR: **Next-generation proteomics: Towards an integrative view of proteome dynamics**. *Nat Rev Genet* 2013, **14**:35–48.
6. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R: **Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial**. *Mol Syst Biol* 2018, **14**:e8126.
7. Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, Wenschuh H, Aebersold R: **High-throughput generation of selected reaction-monitoring assays for proteins and proteomes**. *Nat Methods* 2010, **7**:43–46.
8. Cristobal A, Marino F, Post H, Van Den Toorn HWP, Mohammed S, Heck AJR: **Toward an Optimized Workflow for Middle-Down Proteomics**. *Anal Chem* 2017, **89**:3318–3325.
9. Lössl P, van de Waterbeemd M, Heck AJ: **The diverse and expanding role of mass spectrometry in structural and molecular biology**. *EMBO J* 2016, **35**:2634–2657.
10. Mermelekas G, Zoidakis J: **Mass spectrometry-based membrane proteomics in cancer biomarker discovery**. *Expert Rev Mol Diagn* 2014, **14**:549–563.
11. Zhao M, Wei W, Cheng L, Zhang Y, Wu F, He F, Xu P: **Searching Missing Proteins Based on the Optimization of Membrane Protein Enrichment and Digestion Process**. *J Proteome Res* 2016, **15**:4020–4029.
12. Moore SM, Hess SM, Jorgenson JW: **Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics**. *J Proteome Res* 2016, **15**:1243–1252.
13. Potel CM, Lin M-H, Heck AJR, Lemeer S: **Defeating major contaminants in Fe 3+-IMAC phosphopeptide enrichment**. *Mol Cell Proteomics* 2018, **17**:1028–1034.
14. Vandermarliere E, Mueller M, Martens L: **Getting intimate with trypsin, the leading protease in proteomics**. *mass Spectrom Rev* 2013, **32**:453–465.
15. Guo X, Trudgian DC, Lemoff A, Yadavalli S, Mirzaei H: **Confetti: A Multiprotease Map of the HeLa Proteome for Comprehensive Proteomics**. *Mol Cell Proteomics* 2014, **13**:1573–1584.
16. Giansanti P, Aye TT, van den Toorn H, Peng M, van Breukelen B, Heck AJR: **An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas**. *Cell Rep* 2015, **11**:1834–1843.
17. Riley NM, Coon JJ: **Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling**. *Anal Chem* 2016, **88**:74–94.
18. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M: **Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics**. *Mol Cell Proteomics* 2002, **1**:376–386.
19. Ong S-E, Mann M: **A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC)**. *Nat Protoc* 2007, **1**:2650–2660.
20. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C: **Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS**. *Anal Chem* 2003, **75**:1895–1904.
21. Hsu JL, Huang SY, Chow NH, Chen SH: **Stable-Isotope Dimethyl Labeling for Quantitative Proteomics**. *Anal Chem* 2003, **75**:6843–6852.
22. Krüger M, Moser M, Ussar S, Thievessen I, Luber CA, Forner F, Schmidt S, Zanivan S, Fässler R, Mann M: **SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function**. *Cell* 2008, **134**:353–364.
23. Sury MD, Chen J-X, Selbach M: **The SILAC Fly Allows for Accurate Protein Quantification *in Vivo***. *Mol Cell Proteomics* 2010, **9**:2173–2183.
24. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M: **Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ**. *Mol Cell Proteomics* 2014, **13**:2513–2526.
25. Li Y, Collins M, An J, Geiser R, Tegeler T, Tsantilas K, Garcia K, Pirrotte P, Bowser R: **Immunoprecipitation and mass spectrometry defines an extensive RBM45 protein–protein interaction network**. *Brain Res* 2016, **1647**:79–93.
26. Mulvey CM, Breckels LM, Geladaki A, Britovšek NK, Nightingale DJH, Christoforou A, Elzek M, Deery MJ, Gatto L, Lilley KS: **Using hyperLOPIT to perform high-resolution mapping of the spatial proteome**. *Nat Protoc* 2017, **12**:1110–1135.
27. Bantscheff M, Eberhard D, Abraham Y, Bastuck S, Boesche M, Hobson S, Mathieson T, Perrin J, Raida M, Rau C, et al.: **Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors**. *Nat Biotechnol* 2007, **25**:1035–1044.

28. Doll S, Burlingame AL: **Mass spectrometry-based detection and assignment of protein posttranslational modifications**. *ACS Chem Biol* 2015, **10**:63–71.
29. Guerreiro ACL, Penning R, Raaijmakers LM, Axman IM, Heck AJR, Altelaar AFM: **Monitoring light/dark association dynamics of multi-protein complexes in cyanobacteria using size exclusion chromatography-based proteomics**. *J Proteomics* 2016, **142**:33–44.
30. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP: **Large-scale characterization of HeLa cell nuclear phosphoproteins**. *Proc Natl Acad Sci* 2004, **101**:12130–12135.
31. Batth TS, Francavilla C, Olsen J V.: **Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics**. *J Proteome Res* 2014, **13**:6176–6186.
32. Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, Shen Y, Monroe ME, Lopez-Ferrer D, Reno T, et al.: **Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells**. *Proteomics* 2011, **11**:2019–2026.
33. Mann M, Wilm M: **Electrospray Mass Spectrometry for Protein Characterisation**. *Trends Biochem Sci* 1995, **20**:219–224.
34. Makarov A: **Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis**. *Anal Chem* 2000, **72**:1156–1162.
35. Schwartz JC, Senko MW, Syka JEP: **A two-dimensional quadrupole ion trap mass spectrometer.** *J Am Soc Mass Spectrom* 2002, **13**:659–669.
36. Shukla AK, Futrell JH: **Tandem mass spectrometry: Dissociation of ions by collisional activation**. *J Mass Spectrom* 2000, **35**:1069–1090.
37. Nagaraj N, D'Souza RCJ, Cox J, Olsen J V., Mann M: **Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation**. *J Proteome Res* 2010, **9**:6786–6794.
38. Porras-Yakushi TR, Sweredoski MJ, Hess S: **ETD Outperforms CID and HCD in the Analysis of the Ubiquitylated Proteome**. *J Am Soc Mass Spectrom* 2015, **26**:1580–1587.
39. Yu Q, Shi X, Feng Y, Kent KC, Li L: **Improving data quality and preserving HCD-generated reporter ions with EThcD for isobaric tag-based quantitative proteomics and proteome-wide PTM studies**. *Anal Chim Acta* 2017, **968**:40–49.
40. Senko MW, Remes PM, Canterbury JD, Mathur R, Song Q, Eliuk SM, Mullen C, Earley L, Hardman M, Blethrow JD, et al.: **Novel parallelized quadrupole/linear ion trap/orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates**. *Anal Chem* 2013, **85**:11710–11714.
41. Cottrell JS: **Protein identification using MS/MS data**. *J Proteomics* 2011, **74**:1842–1851.
42. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing**. *J R Stat Soc B* 1995, **57**:289–300.
43. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data**. *Electrophoresis* 1999, **20**:3551–3567.
44. Eng JKJK, McCormack ALL, Yates III JR, Yates JR: **An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database**. *J Am Soc Mass Spectrom* 1994, **5**:976–989.
45. Karimpour-fard A, Epperson LE, Hunter LE: **A survey of computational tools for downstream analysis of proteomic and other omic datasets**. *Hum Genomics* 2015, **9**.
46. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response**. *Proc Natl Acad Sci* 2001, **98**:5116–5121.
47. Giuliani A: **The application of principal component analysis to drug discovery and biomedical data**. *Drug Discov Today* 2017, **22**:1069–1076.
48. Draisma HHM, Reijmers TH, Meulman JJ, Van Der Greef J, Hankemeier T, Boomsma DI: **Hierarchical clustering analysis of blood plasma lipidomics profiles from mono-and dizygotic twin families**. *Eur J Hum Genet* 2013, **21**:95–101.
49. Masuda H, Qi Y, Liu S, Hayashi N, Kogawa T, Hortobagyi GN, Tripathy D, Ueno NT: **Reverse phase protein array identification of triple-negative breast cancer subtypes and comparison with mRNA molecular subtypes.** *Oncotarget* 2017, **8**:70481–70495.
50. Murphy JP, Pinto DM: **Temporal proteomic analysis of IGF-1R signalling in MCF-7 breast adenocarcinoma cells.** *Proteomics* 2010, **10**:1847–1860.
51. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al.: **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles**. *Proc Natl Acad Sci* 2005, **102**:15545–15550.
52. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, Moulton HM, Dejesus P, Che J, Mulder LCF, et al.: **Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding**. *Cell Host Microbe* 2015, **18**:723–735.
53. Jiao X, Sherman BT, Huang DW, Stephens R, Baseler MW, Lane HC, Lempicki RA: **DAVID-WS: A stateful web service to facilitate gene/protein list analysis**. *Bioinformatics* 2012, **28**:1805–1806.
54. Henry VJ, Bandrowski AE, Pepin AS, Gonzalez BJ, Desfeux A: **OMICtools: an informative directory for multi-omic data analysis**. *Database (Oxford)* 2014, **2014**:1–5.
55. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: Kyoto encyclopedia of genes and genomes**. *Nucleic Acids Res* 1999, **27**:29–34.
56. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.: **Gene ontology: Tool for the unification of biology**. *Nat Genet* 2000, **25**:25–29.
57. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, et al.: **Reactome: A knowledgebase of biological pathways**. *Nucleic Acids Res* 2005,

33:428–432.
58. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP: **Molecular signatures database (MSigDB) 3.0**. *Bioinformatics* 2011, **27**:1739–1740.
59. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, Mélius J, Cirillo E, Coort SL, DIgles D, et al.: **WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research**. *Nucleic Acids Res* 2018, **46**:D661–D667.
60. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, et al.: **The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible**. *Nucleic Acids Res* 2017, **45**:D362–D368.
61. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks**. *Genome Res* 2003, doi:10.1101/gr.1239303.metabolite.
62. van Iersel MP, Kelder T, Pico AR, Hanspers K, Coort S, Conklin BR, Evelo C: **Presenting and exploring biological pathways with PathVisio**. *BMC Bioinformatics* 2008, **9**:1–9.
63. Raaijmakers LM, Giansanti P, Possik PA, Mueller J, Peeper DS, Heck AJR, Altelaar AFM: **PhosphoPath: Visualization of Phosphosite-centric Dynamics in Temporal Molecular Networks**. *J Proteome Res* 2015, **14**:4332–4341.
64. Hornbeck P V., Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E: **PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations**. *Nucleic Acids Res* 2015, **43**:D512–D520.
65. Gnad F, Gunawardena J, Mann M: **PHOSIDA 2011: The posttranslational modification database**. *Nucleic Acids Res* 2011, **39**:253–260.
66. Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ, Diella F: **Phospho.ELM: A database of phosphorylation sites-update 2011**. *Nucleic Acids Res* 2011, **39**:261–267.
67. Goel R, Harsha HC, Pandey A, Prasad TSK: **Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis**. *Mol Biosyst* 2012, **8**:453–463.
68. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, et al.: **Human Protein Reference Database - 2009 update**. *Nucleic Acids Res* 2009, **37**:767–772.
69. Huang KY, Su MG, Kao HJ, Hsieh YC, Jhong JH, Cheng KH, Huang H Da, Lee TY: **dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins**. *Nucleic Acids Res* 2016, **44**:D435–D446.
70. Schwartz D, Gygi SP: **An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets**. *Nat Biotechnol* 2005, **23**:1391–1398.
71. Casado P, Rodriguez-Prados J-C, Cosulich SC, Guichard S, Vanhaesebroeck B, Joel S, Cutillas PR: **Kinase-Substrate Enrichment Analysis Provides Insights into the Heterogeneity of Signaling Pathway Activation in Leukemia Cells**. *Sci Signal* 2013, **6**:rs6-rs6.
72. Mischnik M, Sacco F, Cox J, Schneider HC, Schäfer M, Hendlich M, Crowther D, Mann M, Klabunde T: **IKAP: A heuristic framework for inference of kinase activities from Phosphoproteomics data**. *Bioinformatics* 2015, **32**:424–431.
73. Horn H, Schoof EM, Kim J, Robin X, Miller ML, Diella F, Palma A, Cesareni G, Jensen LJ, Linding R: **KinomeXplorer: An integrated platform for kinome biology studies**. *Nat Methods* 2014, **11**:603–604.
74. Miller ML, Jensen LJ, Diella F, Jorgensen C, Tinti M, Li L, Hsiung M, Parker SA, Bordeaux J, Sicheritz-Ponten T, et al.: **Linear Motif Atlas for Phosphorylation-Dependent Signaling**. *Sci Signal* 2008, **1**:ra2-ra2.
75. Linding R, Jensen LJ, Pasculescu A, Olhovsky M, Colwill K, Bork P, Yaffe MB, Pawson T: **NetworKIN: A resource for exploring cellular phosphorylation networks**. *Nucleic Acids Res* 2008, **36**:695–699.
76. Rudolph JD, Graauw M De, Water B Van De, Geiger T: **Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Report Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Networks**. *Cell Syst* 2016, **3**:585–593.
77. Oiseth SJ, Aziz MS: **Cancer immunotherapy : a brief review of the history , possibilities , and challenges ahead**. *J cancer metastasis Treat* 2017, **3**:250–61.
78. Antonia SJ, Mirza N, Fricke I, Chiappori A, Thompson P, Williams N, Bepler G, Simon G, Janssen W, Lee JH, et al.: **Combination of p53 cancer vaccine with chemotherapy in patients with extensive stage small cell lung cancer**. *Clin Cancer Res* 2006, **12**:878–887.
79. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2018**. *CA Cancer J Clin* 2018, **68**:7–30.
80. Bhullar KS, Lagarón NO, McGowan EM, Parmar I, Jha A, Hubbard BP, Rupasinghe HPV: **Kinase-targeted cancer therapies: Progress, challenges and future directions**. *Mol Cancer* 2018, **17**:1–20.
81. Davis A, Gao R, Navin N: **Tumor evolution: Linear, branching, neutral or punctuated?** *Biochim Biophys Acta - Rev Cancer* 2017, **1867**:151–161.
82. Mroz EA, Tward AD, Pickering CR, Myers JN, Ferris RL, Rocco JW: **High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma**. *Cancer* 2013, **119**:3034–3042.
83. Hata AN, Niederst MJ, Archibald HL, Gomez-Caraballo M, Siddiqui FM, Mulvey HE, Maruvka YE, Ji F, Bhang HEC, Radhakrishna VK, et al.: **Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition**. *Nat Med* 2016, **22**:262–269.
84. Sharma S V., Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, McDermott U, Azizian N, Zou L, Fischbach MA, et al.: **A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations**.

Cell 2010, **141**:69–80.
85. Mulder C, Prust N, van Doorn S, Reinecke M, Kuster B, van Bergen en Henegouwen P, Lemeer S: **Adaptive Resistance to EGFR Targeted Therapy by Calcium Signaling in NSCLC Cells**. *Mol Cancer Res* 2018,
86. Ma P, Fu Y, Chen M, Jing Y, Wu J, Li K, Shen Y, Gao J: **Adaptive and Acquired Resistance to EGFR Inhibitors Converge on the MAPK Pathway**. 2016, **6**:1232–1243.
87. Tricker EM, Xu C, Uddin S, Capelletti M, Ercan D, Ogino A, Pratilas CA, Rosen N, Gray NS, Wong K, et al.: **Combined EGFR / MEK Inhibition Prevents the Emergence of Resistance in EGFR -Mutant Lung Cancer**. *Cancer Discov* 2015, **5**:960–971.
88. Batlle E, Clevers H: **Cancer stem cells revisited**. *Nat Publ Gr* 2017, **23**:1124–1134.
89. Palomeras S, Ruiz-Martínez S, Puig T: **Targeting Breast Cancer Stem Cells to Overcome Treatment Resistance**. *Molecules* 2018, **23**:1–16.
90. Jin X, Jin X, Kim H: **Cancer stem cells and differentiation therapy**. *Tumor Biol* 2017, **39**:1–11.
91. Takebe N, Miele L, Harris PJ, Jeong W, Bando H, Kahn M, Yang SX, Ivy SP: **in cancer stem cells : clinical update**. *Nat Publ Gr* 2015, **12**:445–464.
92. Shen M, Kang Y: **Complex interplay between tumor microenvironment and cancer therapy**. *Front Med* 2018, **12**:426–439.
93. Bhowmick NA, Neilson EG, Moses HL: **Stromal fibroblasts in cancer initiation and progression**. *Nature* 2004, **432**:332–7.
94. Wu W, Zaal EA, Berkers CR, Heck AJR, Lemeer S: **CTGF / VEGFA-activated Fibroblasts Promote Tumor Migration Through Micro-environmental Modulation**. *Mol Cell Proteomics* 2018, **17**:1502–1514.
95. Atretkhany KN, Drutskaya MS, Nedospasov SA, Grivennikov SI, Kuprash D V: **Chemokines , cytokines and exosomes help tumors to shape inflammatory microenvironment**. *Pharmacol Ther* 2016, **168**:98–112.
96. Wu T, Dai Y: **Tumor microenvironment and therapeutic response**. *Cancer Lett* 2017, **387**:61–68.
97. Byrne MB, Leslie MT, Gaskins HR, Kenis PJA: **Methods to study the tumor microenvironment under controlled oxygen conditions**. *Trends Biotechnol* 2014, **32**:556–563.
98. Ayuso JM, María V, Lacueva A, Lanuza PM, Allison SJ, Phillips RM, Pardo J, Fernandez LJ: **Development and characterization of a microfluidic model of the tumour microenvironment**. *Sci Rep* 2016, **6**:1–16.
99. Knott GJ, Doudna JA: **CRISPR-Cas guides the future of genetic engineering**. *Science (80-)* 2018, **362**:866–869.
100. Pan D, Kobayashi A, Jiang P, Andrade LF De, Tay RE, Luoma AM, Tsoucas D, Qiu X, Lim K, Rao P, et al.: **A major chromatin regulator determines resistance of tumor cells to T cell – mediated killing**. *Science (80-)* 2018, **395**:770–775.
101. Neggers JE, Kwanten B, Dierckx T, Noguchi H, Voet A, Bral L, Minner K, Massant B, Kint N, Delforge M, et al.: **Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes**. *Nat Commun* 2018, **9**:1–14.
102. Gerstung M, Davies H, Stratton MR, Campbell PJ, Raine KM, Gerstung M, Dawson KJ, Haase K, Loo P Van: **Universal Patterns of Selection in Cancer and Somatic Tissues**. *Cell* 2017, **171**:1029–1041.
103. Jensen NF, Stenvang J, Beck MK, Han B, Gupta R, Rasmussen MH, Tarpgaard LS, Pfeiffer P, Bosman F, Tejpar S, et al.: **Establishment and characterization of models of chemotherapy resistance in colorectal cancer : Towards a predictive signature of chemoresistance**. *Mol Oncol* 2015, **9**:1169–1185.
104. Sagawa Y, Fujitoh A, Nishi H: **Establishment of three cisplatin-resistant endometrial cancer cell lines using two methods of cisplatin exposure**. *Tumor Biol* 2011, **32**:399–408.
105. Terai H, Soejima K, Yasuda H, Nakayama S, Hamamoto J, Arai D: **Activation of the FGF2-FGFR1 Autocrine Pathway : A Novel Mechanism of Acquired Resistance to Gefitinib in NSCLC**. *Mol Cancer Res* 2013, **11**:759–768.
106. Synthase UMP, Cantor JR, Abu-remaileh M, Kanarek N, Louissaint A, Lewis CA, Sabatini DM, Cantor JR, Abu-remaileh M, Kanarek N, et al.: **Physiologic Medium Rewires Cellular Metabolism and Reveals Uric Acid as an Endogenous Inhibitor of Article Physiologic Medium Rewires Cellular Metabolism and Reveals Uric Acid as an Endogenous Inhibitor of UMP Synthase**. *Cell* 2017, **169**:258–272.e17.
107. Sato T, Vries RG, Snippert HJ, Wetering M Van De, Barker N, Stange DE, Es JH Van, Abo A, Kujala P, Peters PJ, et al.: **Single Lgr5 stem cells build crypt – villus structures in vitro without a mesenchymal niche**. *Nature* 2009, **459**:262–265.
108. Clevers H: **Review Modeling Development and Disease with Organoids**. *Cell* 2016, **165**:1586–1597.
109. Drost J, Clevers H: **Organoids in cancer research**. *Nat Rev Cancer* 2018, **18**:407–418.
110. Mattar M, Mccarthy CR, Kulick AR, Qeriqi B, Guzman S: **Establishing and Maintaining an Extensive library of Patient-Derived Xenograft Models**. *Front Oncol* 2018, **8**.
111. Tentler JJ, Tan AC, Weekes CD, Jimeno A, Leong S, Pitts TM, Arcaroli JJ, Messersmith WA, Eckhardt SG: **Patient-derived tumour xenografts as models for oncology drug development**. *Nat Rev Clin Oncol* 2012, **9**:338–350.
112. Einarsdottir BO, Karlsson J, Söderberg EM V, Lindberg MF, Funck-brentano E, Jespersen H, Brynjolfsson SF, Bagge RO, Carstam L, Scobie M, et al.: **A patient-derived xenograft pre-clinical trial reveals treatment responses and a resistance mechanism to karonudib in metastatic melanoma**. *Cell Death Dis* 2018, **9**.
113. Hutchinson L, Kirk R: **High drug attrition rates—where are we going wrong?** *Nat Publ Gr* 2011, **8**:189–190.
114. Begley CG: **preclinical cancer research**. *Nature* 2012, **483**:531–533.
115. Caponigro G, Sellers WR: **Advances in the preclinical testing of cancer therapeutic hypotheses**. *Nature* 2011, **10**:179–87.


PROTEIN
POND

CHAPTER TWO

Proteomic tools to study drug function

Celine Mulder[1], Niels Leijten[1]
and Simone Lemeer[1]

Current Opinion in Systems Biology, 2018, 10:9-18

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands

Introduction

TWO

An increasing number of small molecule inhibitors are being approved by the FDA, calling for methods enabling the dissemination of drug function at a molecular level. Especially drug target engagement studied in a cellular context is extremely important as it will provide a link between drug and cellular phenotype and could explain possible side-effects or resistance mechanisms. Genetic approaches are well suited to identify novel drug targets and genes which play a role in drug function and development of resistance [1]. However, genetic approaches only reveal drug-target interactions indirectly. A change in drug sensitivity upon mutation or overexpression of a specific gene merely shows a possible influence of the gene, but does not prove direct drug-target interaction [2]. The same limitations hold true for transcriptomic data. Even though the extent of gene transcription can be determined, the correlation between mRNA levels and protein abundance is still a subject of debate [3]. More importantly, both genomics and transcriptomics lack the ability to determine the role of often short lived and reversible events like post-translational modifications (PTMs), protein localisation and protein-protein interactions. On the

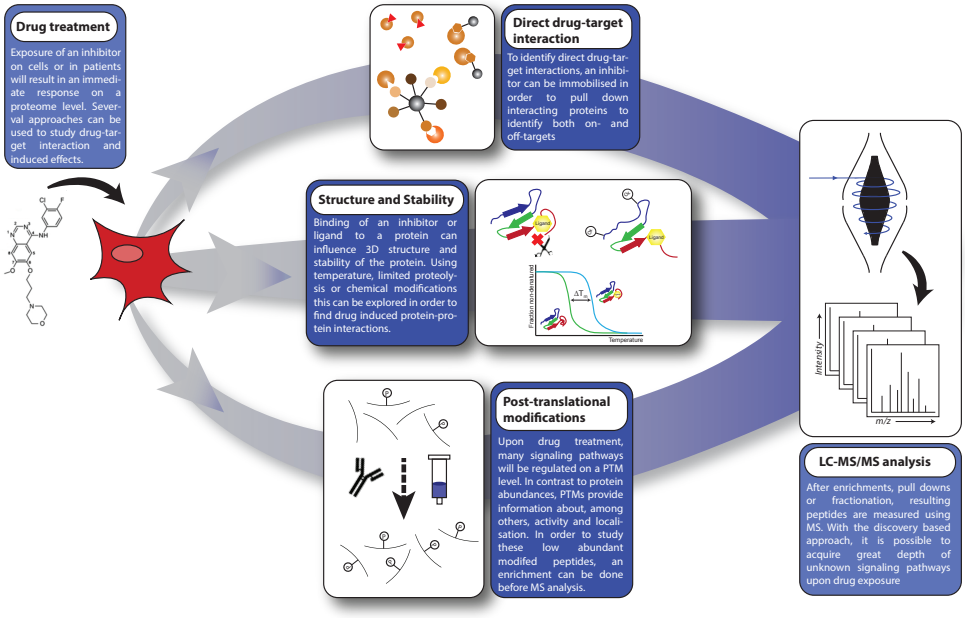


Figure 1 – Overview of discussed topics
Mass spectrometry is a suitable technique to study drug function in complex systems. Moreover, several strategies of sample preparation preceding MS analysis can be used to focus on a particular part of the proteome. Here, we highlight methods to study drug-induced effects on the proteome, with the focus on pull downs to study direct drug-target interactions, methods to study structural changes induced by ligand binding and complex formation and enrichment strategies to study post-translational modifications.

contrary, this information is retained within proteomics data, providing essential information on the regulation of cellular signalling during drug treatment which might explain off-targets and side effects and provide clues to the development of drug resistance which is not caused by mutations. Shotgun proteomics focusses on proteins in a discovery-based manner, resulting in an unbiased representation of the phenotype and molecular landscape [4]. Recent advances made in the proteomics field, such as the development of the Orbitrap mass analyser and Fusion Tribrid instruments, have made it possible to measure and quantify the proteome with high accuracy and great depth. By using optimised sample preparation, fractionation and enrichment strategies preceding LC-MS/MS analysis, different aspects of the proteome can be explored.
In this review, we will list and discuss different ways of fractionation and enrichment before LC-MS/MS analysis to study both drug function and perturbation of cellular signalling induced by small molecule inhibitors, with a focus on direct drug-target interactions, structural changes and PTMs (see figure 1).

Direct drug-target interaction

The classical *in-vitro* workflow used to identify drug-target interactions is an affinity-based protein profiling (AfBPP) approach. This approach is based on a single specific compound immobilized on a solid support, which is incubated with a complex cell lysate. After elution, interacting proteins are detected with LC-MS/MS analysis (see figure 2) [5]. These methods are suitable to screen for both on- and off-targets, to evaluate drug specificity (by performing a competition experiment in which increasing concentrations of free drug that are titrated into the lysates or on cells), and to detect changes in binding affinity or interactome between different conditions [6]. Examples of the successful implementation of this compound centric approach are the target profiling of the chronic myeloid leukaemia drug Dasatinib, originally designed to target Bcr-Abl and Src [7] and Gefitinib, a first generation EGFR inhibitor [8] (More examples are listed in Table 1). Even though this approach is straightforward, introduction of a chemical modification, which is needed to immobilize the compound to a support, might influence the bioactivity of the small molecule. In addition, compounds that cannot be chemically immobilised cannot be analysed using a compound centric approach. An elegant solution to this problem is exemplified by approaches like the kinobead [9], mixed inhibitor bead (MIB) [10] or HDAC probes [11] in which a mixture of broad spectrum inhibitors is immobilized on beads. Using these beads, an entire enzyme class can be enriched from a complex cell lysate. Similar to a compound centric approach, these broad-spectrum binding probes can then be used in a competition experiment with a free drug. However, in contrast to a compound-centric approach, these probes are suitable for high-throughput screening, since it is no longer necessary to produce

a compound-matrix structure for each individual compound of interest because the same probe can be used for each experiment [12]. In addition, these broad binding probes can be used as global enrichment tools for enzyme classes of interest. Several enzyme classes, like kinases or HDACs, are often low abundant, despite their essential role in modulating cellular signalling, and therefore difficult to study using shotgun proteomics. By enriching a subset of enzymes, rather than profiling the proteome, changes in protein abundances and alterations of PTMs which are induced by drug treatment can now be studied in more depth. Some studies using broad-spectrum kinase enrichment have claimed a correlation between affinity and activity [10], however this has recently been opposed [13].
A similar method to study low abundant enzymes and therefore drug function is the use of an activity-based pull down strategy (ABPs). Here a warhead binds a class of enzymes based on their specific enzymatic activity. The warhead is bound to a fluorescent or radioactive probe, which can be visualised on a denaturing gel [14]. With the introduction of MS, ABPs tags were re-designed, in which the reactive group is combined with an affinity tag like biotin. This allows the interactors to be purified and subsequently analysed from complex samples, an approach

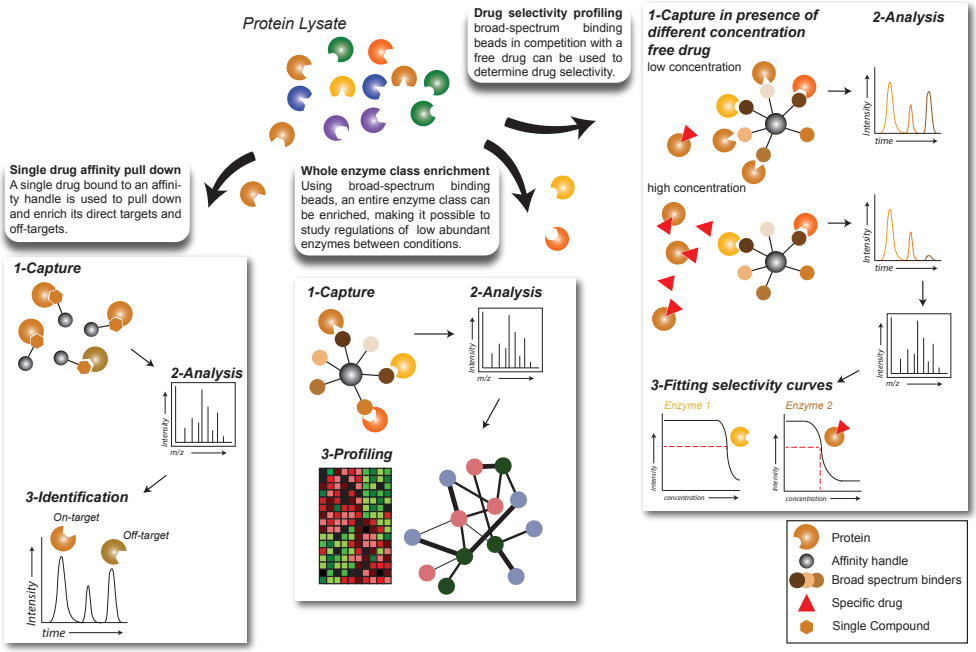


Figure 2 – Pull down Approaches
Overview of discussed pull down approaches to study direct drug-target interactions. An inhibitor immobilised on an affinity handle can be used to purify direct interactors from a complex lysate, which are subsequently analysed with MS. This approach can be used to identify both on- and off-targets. An alternative approach is the use of a broad-spectrum binding probe, which can be used to study an entire enzyme class, or in a competition assay combined with a free drug to determine drug selectivity.

which is called activity-based protein profiling (ABPP) (see figure 2) [15]. Like broad-spectrum binding probes, as described, ABPPs can be used as a global enrichment strategy to study regulation of low abundant, but crucial, enzymes, or in a competition experiment with a free drug. Theoretically, an ABPP probe can be made for every enzyme class, for example methyltransferases [16], proteases [17] hydrolases [18] and kinases [9,10,19,20] (summarised in Table 1). Out of all enzyme classes, kinases are of special interest as they have been implicated in many human diseases [21], therefore several different approaches have been developed to study this enzyme class. Despite enriching the same proteins, these methods all use slightly different methodologies, making them partially complementary [22]. In either method, a consecutive PTM enrichment makes it possible to study modifications determining the regulation of the enzyme with high confidence [23].

Structural changes as evaluation of drug-target interaction

Besides affinity based chemical proteomic approaches, other techniques can be used to measure drug target engagement. Drug-target interactions will cause changes in structure and stability of the target protein, thereby changing its behaviour, which can be detected using limited proteolysis, chemical modifications or due to thermal denaturation (see figure 3).

Limited Proteolysis: DARTS and LiP-MS

In the Drug Affinity Responsive Target Stability assay (DARTS), a low concentration of broad specificity protease (e.g. thermolysin or Pronase) is added to a cell lysate, which causes partial proteolysis at exposed regions of the proteins [24]. When a protein binds to a ligand or compound, this protein might be stabilized and protected from proteolysis, giving rise to a distinct proteolytic pattern of peptides. By comparing the intensity of proteins on a denaturing gel or the intensity of peptides by MS, proteins which are stabilized by the compound can be identified [25,26].

An improvement of this technique is to combine the partial proteolysis step with a tryptic digestion step and subsequently measure these peptides using MS [27]. This allows proteins to be measured on the proteome-wide level, instead of only detecting high abundant proteins using gel-based strategies. Others have used the same approach but focussed the MS analysis on the peptide level, called limited proteolysis mass spectrometry (LiP-MS) [28,29]. By comparing the intensity of specific peptides in the limited proteolysis condition versus a trypsin-only control, the exact region in which a structural change occurs can be determined.

A limitation of DARTS or LiP-MS is the possible laborious optimization of the partially proteolytic conditions for every experiment. In addition, it might be difficult to detect the proteolytic protected peptides in the complex background of unaffected peptides, although this could be circumvented using fractionation. Furthermore, limited proteolysis is only

applicable to soluble proteins due to the native extraction conditions.
An attractive and straightforward alternative is called Pulse Proteolysis [30][31]. In Pulse Proteolysis, the lysate is shortly incubated with a broad specificity protease in the presence or absence of a ligand and at different concentrations of denaturant. By only shortly incubating the lysate with protease, folded proteins will be unaffected while unfolded proteins will be digested. Subsequently, the lysate is analysed on a gel and by comparing the intensities of the bands at different denaturant concentrations, a conclusion can be made concerning the stability of the protein. Bands of interest can then be subjected to in-gel digestion and identified using mass spectrometry. This technique can also be used to measure the stability of membrane proteins [32].

Chemical modification: SPROX and SMTA
An alternative method of determining protein structure and stability is Stability of Proteins from Rate of Oxidation (SPROX) [33,34]. In SPROX, hydrogen peroxide induced oxidation of globally protected methionine residues at increasing concentrations of chemical denaturant is used as a measure for stability. The addition of chemical denaturant will unfold the protein, making previously protected methionine residues susceptible for oxidation. By MS-based quantification of the rate of oxidation at increasing concentration of denaturant, a conclusion can be made on the stability of the protein in either the presence or absence of drugs [35–37]. Subsequently, it is possible to measure binding affinity and domain-based stabilization. A limitation of this technique is the sole application to proteins containing protected methionine residues. In addition, SPROX needs a large amount of input material. Lastly, SPROX is not amendable to membrane proteins, where the chemical induced denaturation is limited due to the lipid bilayer. To increase the scope of SPROX to peptides which do not contain methionine residues, s-methyl thioacetimidate (SMTA) can been used, which causes amidination of buried lysine residues [38]. Since SPROX and SMTA protocols target different amino acids, they are complimentary, thereby improving the effectiveness of this technique [38].

Temperature based: CETSA and TPP
When a protein is subjected to high temperatures, it will start to irreversibly unfold, revealing the hydrophobic core and precipitate. However, when a ligand or compound is bound, the protein stabilizes and the melting point of the complex increases. Consequently, an increased melting point of a protein is indicative of complex formation. In the Cellular Thermal Shift Assay (CETSA), aliquots of lysate or intact cells, which are treated with a drug or vehicle, are subjected to a range of temperature points after which the remaining soluble protein amount is determined using western blot [39,40]. Drug-target interactions can be detected by a difference in protein intensity in vehicle versus drug conditions. The lim-

itation of this method lies in the antibody based detection, as it requires prior knowledge of the target to be analysed as well as the availability of specific antibodies for the target.
In order to deal with these limitations and to allow full proteome probing, this technique has been combined with quantitative MS, called thermal proteome profiling (TPP) [41]. The use of TMT 10plex for ten different temperature points makes it possible to measure a complete melting curve in a single LC-MS/MS run. A big advantage of temperature based profiling is that it can be used on both intact cells and lysates. When drug treatment is performed on intact cells, the cellular environment remains intact. Therefore, it might be possible to not only detect the direct target of the compound but also downstream processes due to (in)activation of the target protein.
However, this method has some limitations. Firstly, TPP does not achieve peptide level resolution, when compared to LiP-MS, therefore it is difficult to determine thermal sensitive of a specific protein domain. Also, in the original protocol, only soluble proteins were analysed while insoluble proteins, such as membrane proteins, were removed before analysis. Later however, it was shown that by adding a mild detergent to the lysis buffer membrane proteins are solubilized, making them amendable for analysis [42]. Furthermore, some proteins do not show thermal stabilization due

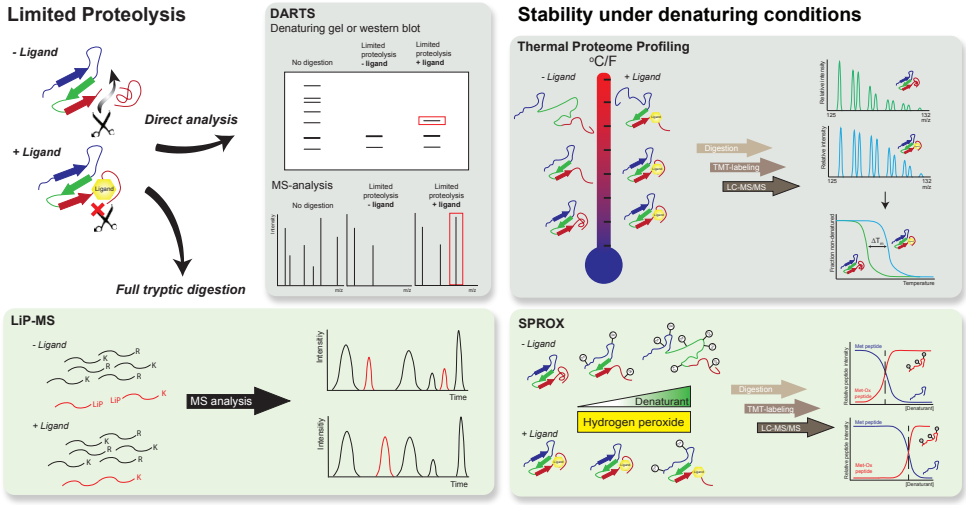


Figure 3 – Structural changes as evaluation of drug-target interaction
Overview of the different principles to evaluate protein-drug target engagement effects. In limited proteolysis a broad specificity protease is used to cleave exposed regions of the protein. This altered digestion pattern is detected using denaturing gel in DARTS, after which the protein can be identified using MS. In LiP-MS the limited proteolysis step is followed by a total digestion using trypsin, followed by detection using MS. In TPP, the ligand stabilizes the protein against thermal denaturation, which can be detected by a shift in melting curve. This method can be used to identify drug-ligand interactions on a protein level. Similarly, in SPROX the rate of unfolding due to a chemical denaturant is determined by oxidizing exposed methionine residues by hydrogen peroxide. This method is gain detailed information of the structural changes due to protein-protein interactions on a peptide level.

to ligand binding, limiting the detection of target engagement effects of these proteins. Lastly, the huge amount of data generated by measuring the whole proteome at the different temperature points can make it difficult to extract valuable information.
Numerous variations on TPP exist. Savitski *et al.* [41] developed a TPP-concentration range assay, which can be used to measure the affinity between ligand and proteins. Becher *et al* [43] developed 2D-TPP, in which they combined 12 temperature points with 5 different drug concentrations. The 2D-TPP assay is more sensitive and can be used to directly measure affinities of stabilized proteins. Lastly, Leuenberger *et al* [44] combined the possibilities of TPP with LiP-MS, which they used to measure thermostability on a proteomic scale with domain-level resolution. Recently, Tan *et al* [45] developed a variant of TPP which they called thermal proximity coaggregation, which can be used to monitor intracellular protein complex dynamics.

Post-translational Modifications
Many small molecule inhibitors that are being developed target enzymes that catalyse post-translational modification (PTMs) of effector proteins. A multitude of possible PTMs affect protein function in terms of turnover, activity, localisation and complex formation, thereby changing a variety of signalling pathways in the cell [46]. Therefore, it is equally important to study direct compound-target interactions as well as the indirect effects of drugs on the PTM landscape upon drug exposure, during treatment and upon development of drug resistance. Peptides carrying PTMs have low stoichiometry compared to their unmodified counterpart, making enrichment before MS analysis a recommended step. PTM enrichments have been developed for several different PTMs, using various approaches [47]. A large body of literature exist on enrichment techniques (a selection has been listed in Table 2), therefore this is not further addressed in this review.
The analysis of PTMs can be used to answer a variety of biological questions in the study to discover drug function. First of all, the discovery based MS approaches are able to determine both direct on-target [48,49], as well as off-target effects [50]. For example, differences in phenotype and additive effect upon treatment with two MEK inhibitors, targeting the same pathway, could be explained by resolving drug induced changes on a PTM level [51]. In addition to detecting immediate effects, changes on a PTM level induced by long-term treatment can explain drug resistance development. For example, the study of the phospho-proteome of a Lapatinib resistant breast cancer cell line led to the discovery of altered signalling pathways responsible for the acquired resistance, in addition to genetic mutations and changes in protein expression [52]. Lastly, since PTMs can determine the activity or function of a protein, they can be used as a biomarker. By integrating data of phospho-proteomes combined

with information on responses, mutation and other metadata, a phospho-proteomic signature reveals hallmarks of disease related signalling [53]. Subsequently, a prediction can be made about the drug response in a specific patient, which can play a role in the determination of the type of treatment.
Some general challenges in PTM enrichment have been described, of which some excellent reviews have been written [47]. A major technical challenge is the large amount of starting material often necessary. Increased enrichment specificity and MS sensitivity are necessary to perform larger scale, high-throughput experiments and/or to perform analysis on clinical material [54]. In addition, some modifications, like cysteine oxidation, are highly unstable, so sample handling should be optimised to preserve these PTMs. In addition to specific enrichment, MS/MS settings can be optimized, fitted to the PTM in question, to get the best quality data [55]. Different fragmentation methods (HCD [56], ETD [57], EtHCD [58], or a combination of techniques [59,60]) are available for optimal fragmentation of (labile) PTMs. These methods allow both the high confidence sequence determination as well as reliable site localization of modified peptides.

Outlook
MS based methods are able to provide detailed knowledge which advances our understanding of unexplored cellular signalling pathways, creating opportunities to resolve drug function. MS-based technology can find novel drug targets, determine off-targets or show an unknown potential in an alternative application than a drug was designed for. Furthermore, proteome wide screening of biological systems can result in understanding of resistance development against an inhibitor or can reveal useful biomarkers. Despite its potential, clinical application of MS technology is still in an early phase. MS experiments are time consuming when analytical depth is required, moreover, a great part of the scientific community does not have access to MS facilities.
Over the last decades, exciting technical advances in the MS field have made it possible to reach further in the proteome than ever before. The current challenge, however, is the vast amount of data produced, risking the true meaning of it to be lost. Therefore, it is important to focus the data acquisition by combining different analytical tools with MS analysis in order to reduce complexity and extract the most useful information from separate parts of a proteome. For this purpose, existing protocols can be altered to make them compatible with MS. Although several examples already exist in literature, more development needs to be done for chemical strategies to be implemented into MS based research. Ultimately, the right combination of inventive methods coupled together with the extensive reach of MS analysis will bring the most unambiguous answers to biological challenges.

Table 1 – Literature overview of pull down approaches

Table of literature describing pull down approaches. Listed are affinity matrix: single or multiple (i.e. single inhibitor/compound or library immobilised on stationary phase), enzyme class targeted by the probe, a description of the design of the affinity or activity compound used for pull down and predicted targets, if listed in the respective paper.

Affinity Matrix	Enzyme Class	Affinity/Activity Compound	Known/predicted Targets	Reference
Single	Kinase	Dasatinib	c-Src; c-Abl	[7]
Single	Kinase	Gefitinib	EGFR	[8]
Single	Methyltransferase	3-Deazaneplanocin A (DzNeP)	*S*-adenosyl-L-homocysteine hydrolase (SAHH); EZH2	[61]
Single	Lipase	Orlistat	Fatty acid synthase (FAS)	[62]
Single	E3 ligase complex	Thalidomide		[63]
Single	PARP1/2 family	XAV939	Wnt pathway	[6]
Single	Kinase	MLN8237	Aurora A Kinase (AKA)	[64]
Single	Kinase	PF-6274484; PCI-32765; Inhibitor 5*, 7*; Afatinib	EGFR; BTK	[65]
Single	Kinases (ATP probe)	Lysine in the ATP pocket		[20]
Single	Proteases	Active site cysteine or serine		[17]
Single	Glycoside hydrolases	2-deoxy-2-fluoroxylobioside		[18]
Single	Methyltransferases	*S*-adenosyl-L-homocysteine (SAH) core		[16]
Single	Serine hydrolases	TAMRA-fluorophosphonate (FP-TAMRA)		[66]
Multiple	Methyltransferases	12 small molecules containing tetrazole library (Gn; *n*=1-12)	DNMT 1	[67]
Multiple	HDACs	SAHA, givinostat		[11]
Multiple	Kinases (CTx-0294885)	Multiple kinase inhibitors		[68]
Multiple	Kinases (kinobeads)	Several inhibitors; beads		[9,12,19]
Multiple	Kinases (MIBs)	Several inhibitors; multi-column		[10]
Multiple	Metalloproteases	Probe library of succinyl Hx scaffolds (HxBPyne probes)		[69]

*Inhibitor 5: (R,E)-1-(3-(4-Amino-3-(4-phenoxyphenyl)-1H-pyrazolo[3,4-d]pyrimidin-1-yl)piperidin-1-yl)but-2-en-1-one
*Inhibitor 7: (E)-N-(4-((3-Chloro-4-fluorophenyl)amino)-7-methoxyquinazolin-6-yl)-4-(dimethylamino)but-2-enamide

Table 2 – Literature overview of PTM enrichment strategies

Many different approaches of PTM enrichment have been developed over recent years. The method of enrichment in largely determinded by the type of PTM, although different approaches are possible, allowing for different applications. A large body of literature exist on the topic, therefore a small sample of PTM enrichment approaches described in literature are listed here as a starting point of further exploration of the desired topic.

Post-translational modification	**Method of enrichment**	**Reference**
Phosphorylation (pTyr)	immunopurification	[70]
Ubiquitination	immunopurification	[71,72]
SUMOylation	IMAC and immunopurification	[73,74]
Lysine and arginine Methylation	immunopurification	[75,76]
Lysine acetylation	immunopurification	[77,78]
O-GlcNAcylation	immunopurification	[79]
Phosphorylation (pHis)	Immunopurification	[80]
Cysteine oxidation	Affinity purification	[81,82]
ADP-ribosylation	PARP/PARG inhibitors combined with affinity-purification	[83][60]
Glycosylation	Leptin magnetic bead array (LeMBA)	[84]
O-GlcNAcylation	biotin alkyne probe combined with click-chemistry	[85,86]
Glycosylation	Boronic acid purification	[87]
Ubiquitination	ThUBD	[88]
Phosphorylation (pSer, pThr)	Fe3+-IMAC	[89]
Phosphorylation (pSer, pThr, pTyr)	TiO4 chromatography; pY antibodies	[90]
Phosphorylation (pSer, pThr)	TiO4 chromatography	[91]

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

SL acknowledges support from the Netherlands Organization for Scientific Research (NWO) through a VIDI grant (project 723.013.008).

References

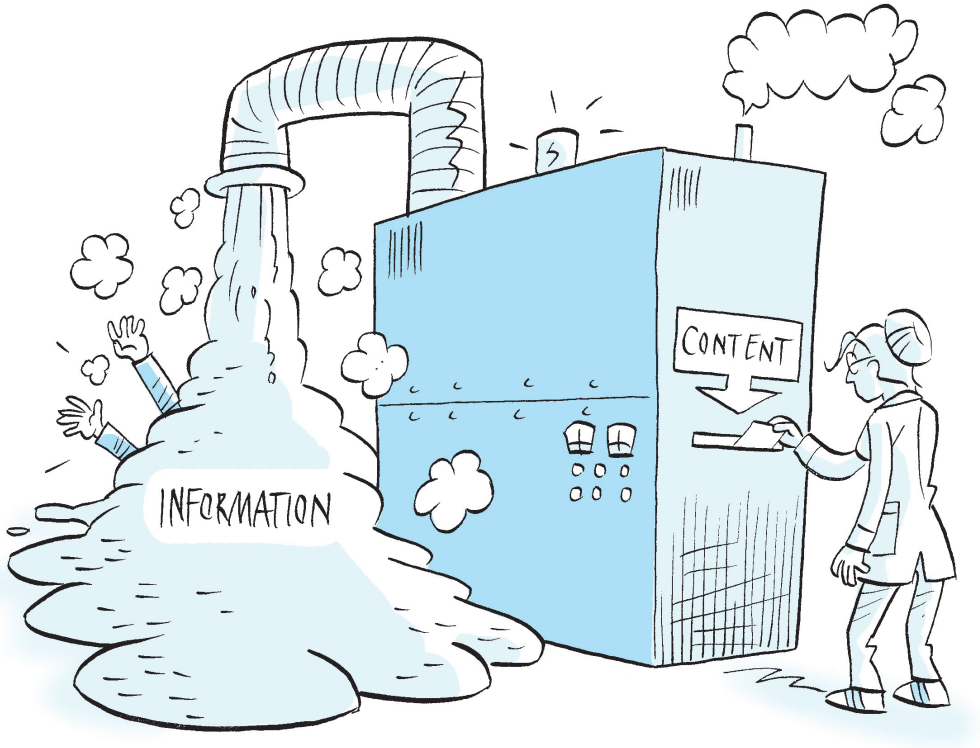
1. Cacace E, Kritikos G, Typas A: **Chemical genetics in drug discovery**. *Curr Opin Syst Biol* 2017, **4**:35–42.
2. Palmer AC, Kishony R: **Opposing effects of target overexpression reveal drug mechanisms**. *Nat Commun* 2014, **5**:1–8.
3. Vogel C, Marcotte EM: **Insights into the regulation of protein abundance from proteomic and transcriptomic analyses**. *Nat Rev Genet* 2012, **13**:227–232.
4. Altelaar AFM, Munoz J, Heck AJR: **Next-generation proteomics: Towards an integrative view of proteome dynamics**. *Nat Rev Genet* 2013, **14**:35–48.
5. Kawatani M, Osada H: **Affinity-based target identification for bioactive small molecules**. *Medchemcomm* 2014, **5**:277.
6. Huang SA, Mishina YM, Liu S, Cheung A, Stegmeier F, Michaud GA, Charlat O, Wiellette E, Zhang Y, Wiessner S, et al.: **Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling**. *Nature* 2009, **461**:614–620.
7. Shi H, Zhang CJ, Chen GYJ, Yao SQ: **Cell-based proteome profiling of potential Dasatinib targets by use of affinity-based probes**. *J Am Chem Soc* 2012, **134**:3001–3014.
8. Brehmer D, Greff Z, Godl K, Blencke S, Kurtenbach A, Weber M, Mu S, Klebl B, Cotten M, Daub H: **Cellular Targets of Gefitinib Cellular Targets of Gefitinib**. *Cancer Res* 2005, **65**:379–382.
9. Bantscheff M, Eberhard D, Abraham Y, Bastuck S, Boesche M, Hobson S, Mathieson T, Perrin J, Raida M, Rau C, et al.: **Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors**. *Nat Biotechnol* 2007, **25**:1035–1044.
10. Duncan JS, Whittle MC, Nakamura K, Abell AN, Midland AA, Zawistowski JS, Johnson NL, Granger DA, Jordan NV, Darr DB, et al.: **Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer**. *Cell* 2012, **149**:307–321.
11. Bantscheff M, Hopf C, Savitski MM, Dittmann A, Grandi P, Michon AM, Schlegl J, Abraham Y, Becher I, Bergamini G, et al.: **Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes**. *Nat Biotechnol* 2011, **29**:255–268.
12. Klaeger S, Heinzlmeir S, Wilhelm M, Polzer H, Vick B, Koenig P-A, Reinecke M, Ruprecht B, Petzoldt S, Meng C, et al.: **The target landscape of clinical kinase drugs**. *Science (80-)* 2017, **358**:eaan4368.
13. Ruprecht B, Zecha J, Heinzlmeir S, Médard G, Lemeer S, Kuster B: **Evaluation of Kinase Activity Profiling Using Chemical Proteomics**. *ACS Chem Biol* 2015, **10**:2743–2752.
14. Evans MJ, Cravatt BF: **Mechanism-based profiling of enzyme families**. *Chem Rev* 2006, **106**:3279–3301.
15. Xiao Y, Wang Y: **Global discovery of protein kinases and other nucleotide-binding proteins by mass spectrometry**. *mass Spectrom Rev* 2016, **35**:601–619.
16. Horning BD, Suciu RM, Ghadiri DA, Ulanovskaya OA, Matthews ML, Lum KM, Backus KM, Brown SJ, Rosen H, Cravatt BF: **Chemical Proteomic Profiling of Human Methyltransferases**. *J Am Chem Soc* 2016, **138**:13335–13343.
17. Fonović M, Verhelst SHL, Sorum MT, Bogyo M: **Proteomics Evaluation of Chemically Cleavable Activity-based Probes**. *Mol Cell Proteomics* 2007, **6**:1761–1770.
18. Hekmat O, Kim YW, Williams SJ, He S, Withers SG: **Active-site peptide "fingerprinting" of glycosidases in complex mixtures by mass spectrometry: Discovery of a novel retaining β-1,4-glycanase in Cellulomonas fimi**. *J Biol Chem* 2005, **280**:35126–35135.
19. Médard G, Pachl F, Ruprecht B, Klaeger S, Heinzlmeir S, Helm D, Qiao H, Ku X, Wilhelm M, Kuehne T, et al.: **Optimized Chemical Proteomics Assay for Kinase Inhibitor Profiling**. *J Proteome Res* 2015, **14**:1574–1586.
20. Patricelli MP, Szardenings AK, Liyanage M, Nomanbhoy TK, Wu M, Weissig H, Aban A, Chun D, Tanner S, Kozarich JW: **Functional interrogation of the kinome using nucleotide acyl phosphates**. *Biochemistry* 2007, **46**:350–358.
21. Wu P, Nielsen TE, Clausen MH: **Small-molecule kinase inhibitors: An analysis of FDA-approved drugs**. *Drug Discov Today* 2016, **21**:5–10.
22. Lemeer S, Zörgiebel C, Ruprecht B, Kohl K, Kuster B: **Comparing immobilized kinase inhibitors and covalent ATP probes for proteomic profiling of kinase expression and drug selectivity**. *J Proteome Res* 2013, **12**:1723–1731.
23. Gridling M, Ficarro SB, Breitwieser FP, Song L, Parapatics K, Colinge J, Haura EB, Marto JA, Superti-Furga G, Bennett KL, et al.: **Identification of Kinase Inhibitor Targets in the Lung Cancer Microenvironment by Chemical and Phosphoproteomics**. *Mol Cancer Ther* 2014, **13**:2751–2762.
24. Lomenick B, Hao R, Jonai N, Chin RM, Aghajan M, Warburton S, Wang J, Wu RP, Gomez F, Loo JA, et al.:

Target identification using drug affinity responsive target stability (DARTS). *Proc Natl Acad Sci* 2009, **106**:21984–21989.
25. Morretta E, Esposito R, Festa C, Riccio R, Casapullo A, Monti MC: **Discovering the biological target of 5-epi-sinuleptolide using a combination of proteomic approaches**. *Mar Drugs* 2017, **15**.
26. Yang Z, Kuboyama T, Tohda C: **A systematic strategy for discovering a therapeutic drug for Alzheimer's disease and its target molecule**. *Front Pharmacol* 2017, **8**:1–13.
27. Lomenick B, Jung G, Wohlschlegel JA, Huang J: **Target Identification Using Drug Affinity Responsive Target Stability (DARTS)**. In *Current Protocols in Chemical Biology*. . 2011:163–180.
28. Feng Y, De Franceschi G, Kahraman A, Soste M, Melnik A, Boersema PJ, De Laureto PP, Nikolaev Y, Oliveira AP, Picotti P: **Global analysis of protein structural changes in complex proteomes**. *Nat Biotechnol* 2014, **32**:1036–1044.
29. Geiger R, Rieckmann JC, Wolf T, Basso C, Feng Y, Fuhrer T, Kogadeeva M, Picotti P, Meissner F, Mann M, et al.: **L-Arginine Modulates T Cell Metabolism and Enhances Survival and Anti-tumor Activity**. *Cell* 2016, **167**:829–842.e13.
30. Chang Y, Schlebach JP, VerHeul RA, Park C: **Simplified proteomics approach to discover protein-ligand interactions**. *Protein Sci* 2012, **21**:1280–1287.
31. Park C, Marqusee S: **Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding**. *Nat Methods* 2005, **2**:207–212.
32. Schlebach JP, Kim MS, Joh NH, Bowie JU, Park C: **Probing membrane protein unfolding with pulse proteolysis**. *J Mol Biol* 2011, **406**:545–551.
33. West GM, Tang L, Fitzgerald MC: **Thermodynamic Analysis of Protein Stability and Ligand Binding Using a Chemical Modification- and Mass Spectrometry-Based Strategy**. 2008, **80**:4175–4185.
34. Strickland EC, Geer MA, Tran DT, Adhikari J, West GM, Dearmond PD, Xu Y, Fitzgerald MC: **Thermodynamic analysis of protein-ligand binding interactions in complex biological mixtures using the stability of proteins from rates of oxidation**. *Nat Protoc* 2013, **8**:148–161.
35. Liu F, Meng H, Fitzgerald MC: **Large-Scale Analysis of Breast Cancer-Related Conformational Changes in Proteins Using SILAC-SPROX**. *J Proteome Res* 2017, **16**:3277–3286.
36. Xu Y, Wallace MAG, Fitzgerald MC: **Thermodynamic Analysis of the Geldanamycin–Hsp90 Interaction in a Whole Cell Lysate Using a Mass Spectrometry-Based Proteomics Approach**. *J Am Soc Mass Spectrom* 2016, **27**:1670–1676.
37. Geer MA, Fitzgerald MC: **Characterization of the Saccharomyces cerevisiae ATP-Interactome using the iTRAQ-SPROX Technique**. *J Am Soc Mass Spectrom* 2016, **27**:233–243.
38. Xu Y, Falk IN, Hallen M a, Fitzgerald MC: **Mass Spectrometry-and Lysine Amidination-Based Protocol for Thermodynamic Analysis of Protein Folding and Ligand Binding Interactions**. *Anal Chem* 2011, **83**:3555–3562.
39. Molina DM, Jafari R, Ignatushchenko M, Seki T, Larsson EA, Dan C, Sreekumar L, Cao Y, Nordlund P: **Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay**. *Science (80-)* 2013, **341**:84–88.
40. Almqvist H, Axelsson H, Jafari R, Dan C, Mateus A, Haraldsson M, Larsson A, Molina DM, Artursson P, Lundbäck T, et al.: **CETSA screening identifies known and novel thymidylate synthase inhibitors and slow intracellular activation of 5-fluorouracil**. *Nat Commun* 2016, **7**:1–11.
41. Savitski MM, Reinhard FBM, Franken H, Werner T, Savitski MF, Eberhard D, Molina DM, Jafari R, Dovega RB, Klaeger S, et al.: **Tracking cancer drugs in living cells by thermal profiling of the proteome**. *Science (80-)* 2014, **346**.
42. Reinhard FBM, Eberhard D, Werner T, Franken H, Childs D, Doce C, Savitski MF, Huber W, Bantscheff M, Savitski MM, et al.: **Thermal proteome profiling monitors ligand interactions with cellular membrane proteins**. *Nat Methods* 2015, **12**:1129–1131.
43. Becher I, Werner T, Doce C, Zaal EA, Tögel I, Khan CA, Rueger A, Muelbaier M, Salzer E, Berkers CR, et al.: **Thermal profiling reveals phenylalanine hydroxylase as an off-target of panobinostat**. *Nat Chem Biol* 2016, **12**:908–910.
44. Leuenberger P, Ganscha S, Kahraman A, Cappelletti V, Boersema PJ, Von Mering C, Claassen M, Picotti P: **Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability**. *Science (80-)* 2017, **355**.
45. Tan CSH, Go KD, Bisteau X, Dai L, Yong CH, Prabhu N, Ozturk MB, Lim YT, Sreekumar L, Lengqvist J, et al.: **Thermal proximity coaggregation for system-wide profiling of protein complex dynamics in cells**. *Science (80-)* 2018, **346**:1–11.
46. Beltrao P, Bork P, Krogan NJ, Van Noort V: **Evolution and functional cross-talk of protein post-translational modifications**. *Mol Syst Biol* 2013, **9**:1–13.
47. Doll S, Burlingame AL: **Mass spectrometry-based detection and assignment of protein posttranslational modifications**. *ACS Chem Biol* 2015, **10**:63–71.
48. Hu C-W, Hsu C-L, Wang Y-C, Ishihama Y, Ku W-C, Huang H-C, Juan H-F: **Temporal Phosphoproteome Dynamics Induced by an ATP Synthase Inhibitor Citreoviridin**. *Mol Cell Proteomics* 2015, **14**:3284–3298.
49. Zhu D, Hou L, Hu B, Zhao H, Sun J, Wang J, Meng X: **Crosstalk among proteome, acetylome and succinylome in colon cancer HCT116 cell treated with sodium dichloroacetate**. *Sci Rep* 2016, **6**:1–12.
50. Kim J-Y, Welsh EA, Fang B, Bai Y, Kinose F, Eschrich SA, Koomen JM, Haura EB: **Phosphoproteomics Reveals MAPK Inhibitors Enhance MET- and EGFR-Driven AKT Signaling in KRAS-Mutant Lung Cancer**. *Mol Cancer Res* 2016, **14**:1019–1029.

51. Stuart SA, Houel S, Lee T, Wang N, Old WM, Ahn NG: **A Phosphoproteomic Comparison of B-RAF ^{V600E} and MKK1/2 Inhibitors in Melanoma Cells**. *Mol Cell Proteomics* 2015, **14**:1599–1615.
52. Ruprecht B, Zaal EA, Zecha J, Wu W, Berkers CR, Kuster B, Lemeer S: **Lapatinib resistance in breast cancer cells is accompanied by phosphorylation-mediated reprogramming of glycolysis**. *Cancer Res* 2017, **77**:1842–1853.
53. Drake JM, Paull EO, Graham NA, Lee JK, Smith BA, Titz B, Stoyanova T, Faltermeier CM, Uzunangelov V, Carlin DE, et al.: **Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer**. *Cell* 2016, **166**:1041–1054.
54. Post H, Penning R, Fitzpatrick MA, Garrigues LB, Wu W, Macgillavry HD, Hoogenraad CC, Heck AJR, Altelaar AFM: **Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons**. *J Proteome Res* 2017, **16**:728–737.
55. Ferries S, Perkins S, Brownridge PJ, Campbell A, Eyers PA, Jones AR, Eyers CE: **Evaluation of Parameters for Confident Phosphorylation Site Localization Using an Orbitrap Fusion Tribrid Mass Spectrometer**. *J Proteome Res* 2017, **16**:3448–3459.
56. Nagaraj N, D'Souza RCJ, Cox J, Olsen J V., Mann M: **Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation**. *J Proteome Res* 2010, **9**:6786–6794.
57. Porras-Yakushi TR, Sweredoski MJ, Hess S: **ETD Outperforms CID and HCD in the Analysis of the Ubiquitylated Proteome**. *J Am Soc Mass Spectrom* 2015, **26**:1580–1587.
58. Yu Q, Shi X, Feng Y, Kent KC, Li L: **Improving data quality and preserving HCD-generated reporter ions with EThcD for isobaric tag-based quantitative proteomics and proteome-wide PTM studies**. *Anal Chim Acta* 2017, **968**:40–49.
59. Zeng WF, Liu MQ, Zhang Y, Wu JQ, Fang P, Peng C, Nie A, Yan G, Cao W, Liu C, et al.: **pGlyco: A pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3**. *Sci Rep* 2016, **6**:1–10.
60. Bilan V, Leutert M, Nanni P, Panse C, Hottiger MO: **Combining Higher-Energy Collision Dissociation and Electron-Transfer/Higher-Energy Collision Dissociation Fragmentation in a Product-Dependent Manner Confidently Assigns Proteomewide ADP-Ribose Acceptor Sites**. *Anal Chem* 2017, **89**:1523–1530.
61. Tam EKW, Li Z, Goh YL, Cheng X, Wong SY, Santhanakrishnan S, Chai CLL, Yao SQ: **Cell-based proteome profiling using an affinity-based probe (AfBP) derived from 3-deazaneplanocin A (DzNep)**. *Chem - An Asian J* 2013, **8**:1818–1828.
62. Yang PY, Liu K, Ngai MH, Lear MJ, Wenk MR, Yao SQ: **Activity-based proteome profiling of potential cellular targets of orlistat - An FDA-approved drug with anti-tumor activities**. *J Am Chem Soc* 2010, **132**:656–666.
63. Teratogenicity T, Ito T, Ando H, Suzuki T, Ogura T, Hotta K, Imamura Y, Yamaguchi Y, Handa H: **Identification of a Primary Target of Thalidomide Teratogenicity**. *Science (80-)* 2010, **327**:1345–1351.
64. Su Y, Pan S, Li Z, Li L, Wu X, Hao P, Sze SK, Yao SQ: **Multiplex imaging and cellular target identification of kinase inhibitors via an affinity-based proteome profiling approach**. *Sci Rep* 2015, **5**:1–10.
65. Lanning BR, Whitby LR, Dix MM, Douhan J, Gilbert AM, Hett EC, Johnson TO, Joslyn C, Kath JC, Niessen S, et al.: **A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors**. *Nat Chem Biol* 2014, **10**:760–767.
66. van Esbroeck ACM, Janssen APA, Cognetta AB, Ogasawara D, Shpak G, van der Kroeg M, Kantae V, Baggelaar MP, de Vrij FMS, Deng H, et al.: **Activity-based protein profiling reveals off-target proteins of the FAAH inhibitor BIA 10-2474**. *Science (80-)* 2017, **356**:1084–1087.
67. Zhu B, Ge J, Yao SQ: **Developing new chemical tools for DNA methyltransferase 1 (DNMT 1): A small-molecule activity-based probe and novel tetrazole-containing inhibitors**. *Bioorganic Med Chem* 2015, **23**:2917–2927.
68. Zhang L, Holmes IP, Hochgräfe F, Walker SR, Ali NA, Humphrey ES, Wu J, De Silva M, Kersten WJA, Connor T, et al.: **Characterization of the novel broad-spectrum kinase inhibitor CTx-0294885 as an affinity reagent for mass spectrometry-based kinome profiling**. *J Proteome Res* 2013, **12**:3104–3116.
69. Sieber SA, Niessen S, Hoover HS, Cravatt BF: **Proteomic profiling of metalloprotease activities with cocktails of active-site probes**. *Nat Chem Biol* 2006, **2**:274–281.
70. Abe Y, Nagano M, Tada A, Adachi J, Tomonaga T: **Deep Phosphotyrosine Proteomics by Optimization of Phosphotyrosine Enrichment and MS/MS Parameters**. *J Proteome Res* 2017, **16**:1077–1086.
71. Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, Sowa ME, Rad R, Rush J, Comb MJ, et al.: **Systematic and quantitative assessment of the ubiquitin-modified proteome**. *Mol Cell* 2011, **44**:325–340.
72. Rose CM, Isasa M, Ordureau A, Prado MA, Beausoleil SA, Jedrychowski MP, Finley DJ, Harper JW, Gygi SP: **Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes**. *Cell Syst* 2016, **3**:395–403.
73. Lamoliatte F, Caron D, Durette C, Mahrouche L, Maroui MA, Caron-Lizotte O, Bonneil E, Chelbi-Alix MK, Thibault P: **Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling**. *Nat Commun* 2014, **5**.
74. Lamoliatte F, McManus FP, Maarifi G, Chelbi-Alix MK, Thibault P: **Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification**. *Nat Commun* 2017, **8**.
75. Wu Z, Cheng Z, Sun M, Wan X, Liu P, He T, Tan M, Zhao Y: **A Chemical Proteomics Approach for Global Analysis of Lysine Monomethylome Profiling**. *Mol Cell Proteomics* 2015, **14**:329–339.
76. Guo A, Gu H, Zhou J, Mulhern D, Wang Y, Lee KA, Yang V, Aguiar M, Kornhauser J, Jia X, et al.: **Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation**. *Mol Cell Proteomics* 2014,

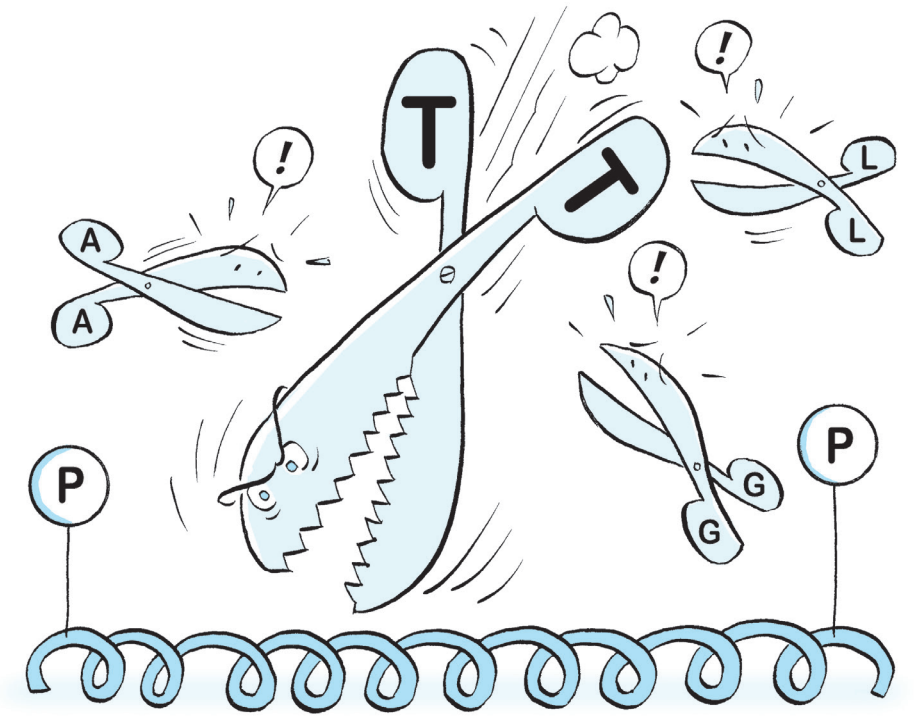
13:372–387.

77. Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen J V, Mann M: **Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions**. *Science (80-)* 2009, **325**:834–841.
78. Svinkina T, Gu H, Silva JC, Mertins P, Qiao J, Fereshetian S, Jaffe JD, Kuhn E, Udeshi ND, Carr SA: **Deep, Quantitative Coverage of the Lysine Acetylome Using Novel Anti-acetyl-lysine Antibodies and an Optimized Proteomic Workflow**. *Mol Cell Proteomics* 2015, **14**:2429–2440.
79. Lee A, Miller D, Henry R, Paruchuri VDP, O'Meally RN, Boronina T, Cole RN, Zachara NE: **Combined Antibody/Lectin Enrichment Identifies Extensive Changes in the O-GlcNAc Sub-proteome upon Oxidative Stress**. *J Proteome Res* 2016, **15**:4318–4336.
80. Fuhs SR, Meisenhelder J, Aslanian A, Ma L, Zagorska A, Stankova M, Binnie A, Al-Obeidi F, Mauger J, Lemke G, et al.: **Monoclonal 1- and 3-Phosphohistidine Antibodies: New Tools to Study Histidine Phosphorylation**. *Cell* 2015, **162**:198–210.
81. Duan J, Gaffrey MJ, Qian W-J: **Quantitative proteomic characterization of redox-dependent post-translational modifications on protein cysteines**. *Mol BioSyst* 2017, **13**:816–829.
82. Guo J, Gaffrey MJ, Su D, Liu T, Camp DG, Smith RD, Qian WJ: **Resin-Assisted enrichment of thiols as a general strategy for proteomic profiling of cysteine-based reversible modifications**. *Nat Protoc* 2014, **9**:64–75.
83. Martello R, Leutert M, Jungmichel S, Bilan V, Larsen SC, Young C, Hottiger MO, Nielsen ML: **Proteome-wide identification of the endogenous ADP-ribosylome of mammalian cells and tissue**. *Nat Commun* 2016, **7**:1–13.
84. Choi E, Loo D, Dennis JW, O'Leary CA, Hill MM: **High-throughput lectin magnetic bead array-coupled tandem mass spectrometry for glycoprotein biomarker discovery**. *Electrophoresis* 2011, **32**:3564–3575.
85. Dehennaut V, Slomianny M-C, Page A, Vercoutter-Edouart A-S, Jessus C, Michalski J-C, Vilain J-P, Bodart J-F, Lefebvre T: **Identification of structural and functional O-linked N-acetylglucosamine-bearing proteins in Xenopus laevis oocyte**. *Mol Cell Proteomics* 2008, **7**:2229–2245.
86. Kupferschmid M, Aquino-Gil MO, Shams-Eldin H, Schmidt J, Yamakawa N, Krzewinski F, Schwarz RT, Lefebvre T: **Identification of O-GlcNAcylated proteins in Plasmodium falciparum**. *Malar J* 2017, **16**:485.
87. Chen W, Smeekens JM, Wu R: **A Universal Chemical Enrichment Method for Mapping the Yeast N -glycoproteome by Mass Spectrometry (MS)**. *Mol Cell Proteomics* 2014, **13**:1563–1572.
88. Gao Y, Li Y, Zhang C, Zhao M, Deng C, Lan Q, Liu Z, Su N, Wang J, Xu F, et al.: **Enhanced Purification of Ubiquitinated Proteins by Engineered Tandem Hybrid Ubiquitin-binding Domains (ThUBDs)**. *Mol Cell Proteomics* 2016, **15**:1381–1396.
89. Ruprecht B, Koch H, Medard G, Mundt M, Kuster B, Lemeer S: **Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns**. *Mol Cell Proteomics* 2015, **14**:205–215.
90. Sharma K, D'Souza RCJ, Tyanova S, Schaab C, Wiśniewski J, Cox J, Mann M: **Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling**. *Cell Rep* 2014, **8**:1583–1594.
91. Thingholm TE, Jørgensen TJD, Jensen ON, Larsen MR: **Highly selective enrichment of phosphorylated peptides using titanium dioxide**. *Nat Protoc* 2006, **1**:1929–1935.


INFORMATION
CONTENT

Part 2

Increasing Proteomic Depth


T
T
!
!
!
A
A
L
L
G
G
P
P

CHAPTER THREE

Increasing (phospho)proteomic coverage by optimized sample preparation

Celine Mulder[1], Sander van Doorn[1] and Simone Lemeer[1]

[1] *Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands*

Introduction

THREE

The study of the cellular proteome, known as the field of proteomics, is still considered to be relatively young, as compared to its 'older brothers' genomics and transcriptomics research. Not surprisingly, technical developments, which are an integral part of the mass spectrometry (MS) based proteomic workflow, constitute the majority of current published proteomic studies. This everlasting quest in scaling up numbers of identified proteins and improving protein coverage, constitutes a large amount of the proteomic research carried out in laboratories around the world [1,2].
A large number of technical tools are at the disposal of the proteomic researcher in order to perform satisfactory optimisation of their experiments. First, the most diverse step in a standard MS workflow is the protocol of cell or tissue lysis and the composition of the used lysis buffer. Significant improvements can be made in peptide detection if, for example contaminants are cleared up [3] or if lysis is adjusted to a desired type of protein [4].
Second, a different set of peptides can be obtained if proteins are digested with different proteases. The use of the highly specific protease trypsin has been the favourable choice in the majority of published proteomic studies [5]. The main reason for this is that tryptic peptides are ideal for MS analysis: peptides have a workable length because of the distribution of lysine and arginine throughout proteins and they have a favourable charge state, due to the charged amino acids at their termini [6]. Consequently, the majority of knowledge on the proteome is based on tryptic peptides. Although the resulting similarities and overlapping results of studies following the use of similar protocols benefits their comparability, most of the available data is exclusively describing trypsin favourable regions, limiting the discovery of the complete proteome. The use of alternative proteases can be a strategy to reveal complementary parts of proteins which cannot be uncovered by trypsin. Others have indeed shown that orthogonal digestion using different enzymes results in the production of distinct types of peptides, covering separate parts of a protein [7–9].
Third, although peptide separation by online liquid chromatography before MS analysis has already become a standard, additional (off- or online) fractionation can be applied to a complex peptide mixture to increase the depth of the identified peptides. Decreasing the complexity of each measured fraction favours the detection of lower abundant peptides, thereby increasing the dynamic range of the experiment. Examples of fractionation are ion exchange chromatography (SCX, SAX) [10], size exclusion chromatography (SEC) [11] or high-pH fractionation [12].
Lastly, the mass spectrometric analysis itself can be optimised. Fitted to the nature of peptides at hand, several parameters can be optimised, for

example the fragmentation type suitable for chemically distinct peptides [13].
An alternative route one can take in order to increase the coverage of the proteome, is to separate distinct parts of the proteome, based on their chemical characteristics, instead of studying all proteins present. A common choice is the study of post-translationally modified proteins or peptides. In this approach modified proteins and/or peptides are separated from their non-modified counterparts and analysed by MS separately. Using this approach, these generally lower abundant and therefore often suppressed modified proteins or peptides, can be readily analysed in the mass spectrometer. Such approaches have been successfully applied to study, amongst others, ubiquitinated, acetylated and glycosylated proteins and peptides, generally using fractionation or affinity chromatographic approaches.
Due to its importance in cellular signalling and its implication in human disease, the study of protein phosphorylation has gained a lot of attention of the last years. Researchers are now capable of identifying thousands of phosphorylation sites from limited amount of samples, which has been aided by the development of several techniques to enrich phosphorylated peptides [14,15], resulting in an continuous growing database of known phosphosites [16].
In this study we used our recently developed sample preparation protocol [3,15], designed for the most optimal phosphopeptide enrichment using trypsin, and employed this to perform digestion with different enzymes, to investigate whether this approach was equally beneficial for other enzymes as well. In addition, we performed high-pH fractionation on a tryptic digest with subsequent phosphopeptide enrichment [12,17] in order to maximise the number of identified phosphopeptides.

Results

Mass spectrometric optimisation

Depending on the type of proteomic experiment and the type of peptides to be analysed, it is crucial to consider the mass spectrometric method and analysis settings one would like to use. Currently, due to the recent advances in Q-Exactive instrumentation and the availability of hybrid instruments such as Fusion and Lumos mass spectrometers, most widely used fragmentation methods for bottom-up proteomics are HCD, ETD or EThcD fragmentation. Specifically, for phosphopeptide analysis, HCD fragmentation has been shown to outperform traditional CID fragmentation, due to higher quality fragment spectra and less phosphate neutral loss during fragmentation, allowing for better site localization [18,19].
The major advantage of ETD and EThcD fragmentation is the preservation of labile PTMs during fragmentation allowing for accurate site localiza-

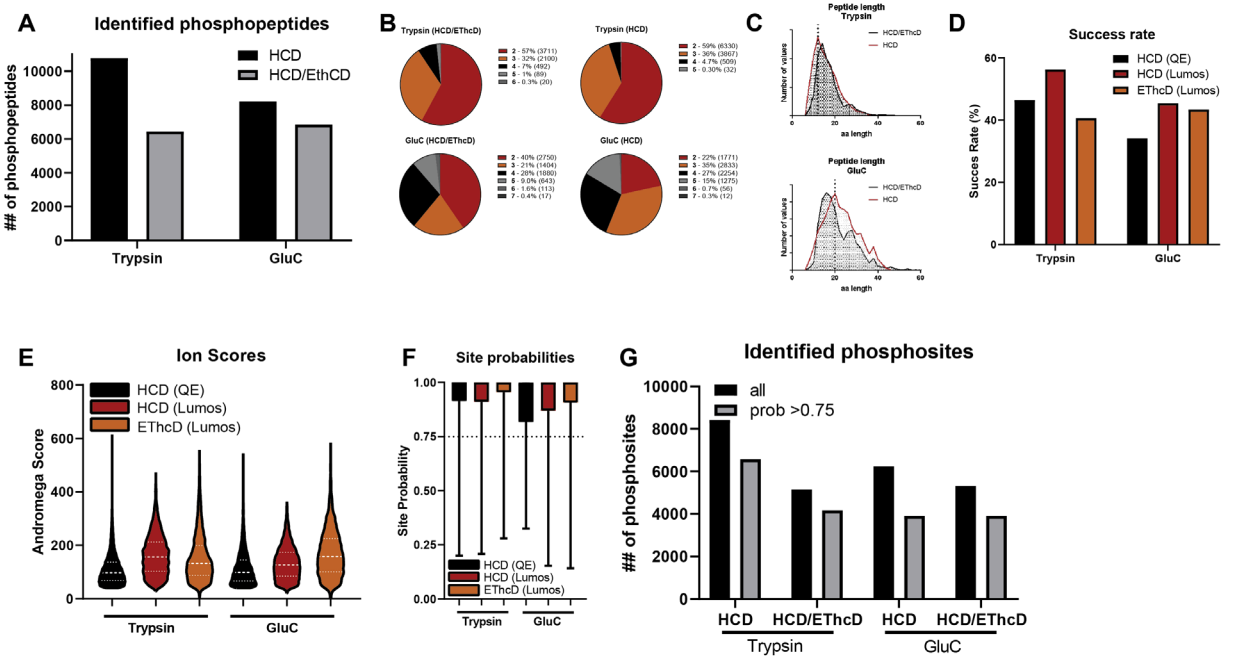


Figure 1 – MS Optimisation

A: The total number of identified phosphopeptides was shown for either a digestion using trypsin or GluC. The black bars indicate the HCD only (Q-Exactive) measurements, the grey bars the HCD/EThcD decision tree (Fusion Lumos) measurements. **B:** In pie plots, the distribution of different charge states was shown for with trypsin (Top) or GluC (Bottom) digested peptides, measured by either HCD/EThcD (left) or HCD only (right). **C**: histograms show the distribution of peptide length for either trypsin (top) or GluC (bottom) digested peptides. The black shaded histograms show the EThcD measurements, whereas the red line indicates the distribution as measured with HCD only. **D, E, F**: For both enzymes the success rate (i.e. the percentage of successful identified MSMS spectra of all MSMS spectra recorded) **(D),** the ion score **(E)** and the distribution of phosphosite localisation probabilities **(F)** was shown for HCD only or, separately, HCD or EThcD MS/MS spectra during the decision tree. **G:** Total number of identified phosphosites was shown for both enzymes for HCD only or HCD/EThcD decision tree measurement. Furthermore, the number of Class I phosphosites were shown (probability score >0.75)

tion. However, EThcD fragmentation is relatively slow and the capture or transfer of low energy electrons does not always disrupt non-covalent interactions, which can result in the product ions staying together, termed ETnoD. Still, EThcD fragmentation has been proven beneficial for the analysis of relatively large peptides and peptides with higher charges due to the larger amount of energy used for fragmentation [20] It has been shown that especially EThcD is well suited, for example, for middle down proteomics [13].

Therefore, prior to conducting a large-scale experiment in which we wanted to test the effect of different digestion enzymes on phosphoproteome coverage, we first examined the performance of HCD and a data dependent decision tree with HCD/EThcD fragmentation on peptides generated by either GluC or Trypsin digestion as a measure for further experiments (see for details materials and method section).

For optimisation experiments, we used 20% of the standard amount of peptide mixture per MS run, in order to save material and IMAC time. Therefore, the number of identified peptides and sites were lower in the optimisation experiments compared to the final experiments. For trypsin, HCD only outperformed decision tree HCD/EThcD in terms of number of identified phosphopeptides, as almost 60% more phosphopeptides were detected using HCD only (Figure 1A). Also, for the GluC digest, HCD outperformed decision tree HCD/EThcD fragmentation, although the difference (17%) is significantly lower compared to the results for trypsin.

For tryptic peptides, both fragmentation methods resulted in the identification of the same type of peptides with similar charge and amino acid length distribution (Fig 1B, C). Paradoxically, for GluC digested peptides, peptides with charges higher than 2+ were favoured for identification with HCD fragmentation compared to decision tree HCD/EThcD fragmentation (Fig 1B). When looking at the amount of HCD or EThcD events triggered for the different charge states of GluC peptides it was clear that almost all 2+ peptides were fragmented by HCD fragmentation. This was indicative for the fact that there was only a minority of small and highly charged peptides (m/z range 350-650) for which EThcD were triggered. For the 3+ peptides half of the spectra were acquired by HCD, the other half by EThcD indicating a distribution of m/z values between 350 and 1200 m/z (Supplementary Figure S1A).

For each charge state the success rate (calculated as the percentage of identified over recorded MS/MS spectra) was investigated. As can be seen, EThcD was extremely unsuccessful for the small amount of 2+ peptides that were selected (Supplementary Figure S1B). Also, for 3+ peptides both HCD and EThcD had a slightly lower success rate when compared to HCD only (Supplementary Figure S1C). Overall the success rate of HCD fragmentation on the Q-Exactive, using our standard settings, was lower than of HCD fragmentation on the Lumos, for both enzymes (Figure 1D). Considering GluC, EThcD had a higher success rate as well. As this cannot

explain the observed higher number of identifications for higher charges in HCD only, we propose that the main explanation for the observed difference in number of identified peptides/sites is the much faster cycle time of HCD as compared to decision tree HCD/EThcD fragmentation.
In contrast, due to the higher level of energy and the production of more complex fragmentation spectra containing multiple ion series by HCD/EthCD it is not surprising that for both enzymes the ion scores (Fig 1E) and site localisation probability (Fig 1F) were higher after HCD/EthCD than after HCD fragmentation. Especially EThcD fragmentation on GluC peptides produced high confident phosphopeptide identifications.
Still, despite the higher scores and higher probabilities, HCD outperformed decision tree HCD/EThcD, when looking at the mere number of identifications. Even after phosphosites which were identified by presumably bad quality spectra were filtered out by selecting only class I phosphosites (defined as sites identified with a localisation probability >0.75) (Fig 1G), HCD outperformed HCD/EThcD for both enzymes, even though less sites were 'lost' for HCD/EThcD. The main explanation for the observed difference in number of identified peptides/sites is consequently still, the much faster cycle time of HCD compared to EThcD fragmentation. Therefore, even for larger and higher charged peptides, the speed of the HCD transcends the superior quality of the EThcD, making HCD the preferred fragmentation method for the current study.

The use of alternative enzymes is more beneficial than fractionation

One of the main reasons for increasing the number of identified phosphopeptides and phosphosites is to extend the coverage of the (phospho) proteome, with the aim to increase biological understanding of a model system. Here we compared the effectiveness of either using alternative proteases or the use of high-pH fractionation, on a tryptic digest with subsequent phosphopeptide enrichment, for the augmentation of the phosphoproteome coverage. We used 4 different proteases, trypsin (cleaving C-terminal of K and R), GluC (C-terminal of D and E), LysC (cleaving C-terminal of K), AspN (cleaving N-terminal N and C) and performed all digests in triplicate. All samples were prepared according to the new lysis protocol and phosphopeptides were enriched using the Fe^{3+}-IMAC system [3, 14]. For high-pH fractionation, 2mg of tryptic digest was fractionated, after which 10 concatenated fractions were enriched using Fe^{3+}-IMAC on a BRAVO system [15].
First of all, for none of the enzymes tested, addition of technical or biological replicates increased the number of unique identifications to a significant extent (Fig 2A, B). Not surprisingly, the use of high-pH fractionation did significantly increase the number of unique identifications when

compared to the unfractionated tryptic digest. Within the high-pH experiment three biological replicates were used and per replicate 10 separate fractions were measured. Each fraction was measured on a 90 minute analysis run, making the total run time 45 hours. This is in comparison with 9 hours per unfractionated digest of 3 replicates using 3 hour runs (Table 1). However, combining all results from the digests using different enzymes, by far the most unique phosphopeptides and phosphosites could be identified, even though the total experimental and analysis time used was lower than for high-pH fractionation. The main reason for this increase was the enormous complementarity of resulting peptides after the use of alternative enzymes. The overlap between single MS runs of the same enzyme was extremely high, whereas the overlap between runs of different enzymes was significantly lower (Fig 2C). To show the full complementarity of the data all identified phosphosites were mapped into a heat map, showing the unique phosphoproteomic signatures of the different enzymes, agreeing with previously published studies (Fig 2D) [8].

Orthology in enzymes increases coverage of the phosphoproteome

Because so far the majority of the proteomic studies have been carried out using trypsin as a protease, publicly available databases have a bias to-

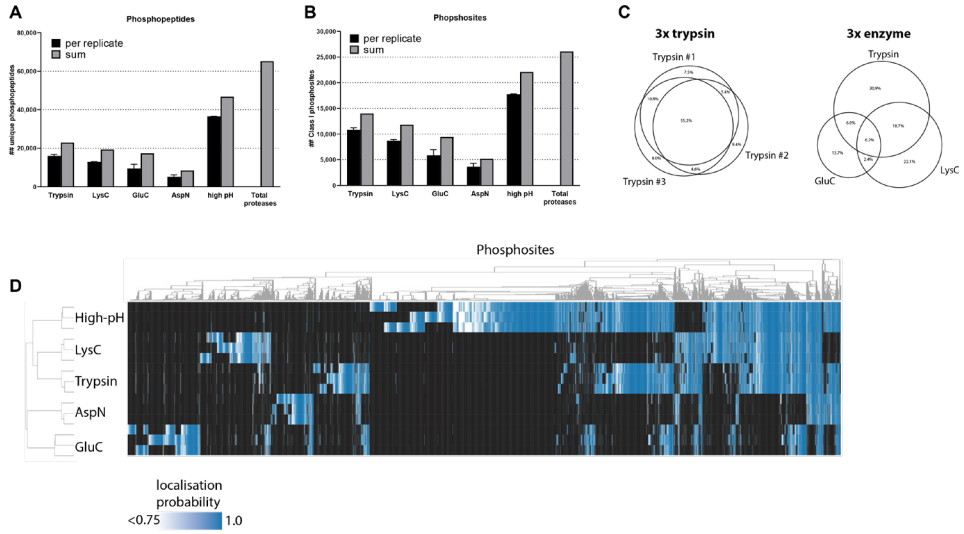


Figure 2 – Comparison of the use of alternative proteases versus high-pH fractionation

A, B: Total number of phosphopeptides **(A)** and phosphosites **(B)** is shown for the separate runs of different enzymes, all enzymes combined and after high-pH fractionation. In black the average number of peptides or sites per condition is shown, in which the error bars represent the different replicates. The grey bar shows the summed amount of all replicates of that condition. **(C)** Overlap of identified phosphosites if three runs of trypsin peptides only (left) or three runs of either trypsin, AspN or GluC peptides (right) were compared. **(D)** All identified phosphosites (vertical clustering) in all replicates (horizontal clustering) were clustered in a hierarchical clustering, based on their phosphosite localisation probability.

wards the collection of phosphosites that are easily detected using tryptic peptides. Indeed, if all sites were categorized as previously identified and documented on PhosphoSitePlus (PSP) [21] or newly identified, a slight increase in the percentage of newly identified sites could be observed when using trypsin as a digestion enzyme (Figure 3A). Surprisingly, the use of high pH fractionation only produced a slight increase in percentage of newly identified sites when compared to trypsin without fractionation (Figure 3A). However, the use of alternative proteases LysC, AspN and GluC, gave a slightly higher increase in the percentage of newly identified sites (Figure 3A). Not surprisingly, the combination of multiple proteases gave the highest proportional increase in newly identified sites, which was higher than both the number and proportional increase in newly identified sites using fractionation (Figure 3A)(Supplementary Figure S2)
Interestingly, although significantly more class I phosphosites were identified when using multiple different digestion enzymes in comparison to high-pH fractionation, only a minority of new phosphoproteins were identified (18% vs 1.6%) (Fig 3B). This discrepancy indicated that by using multiple enzymes the phosphoproteome coverage per protein will strong-

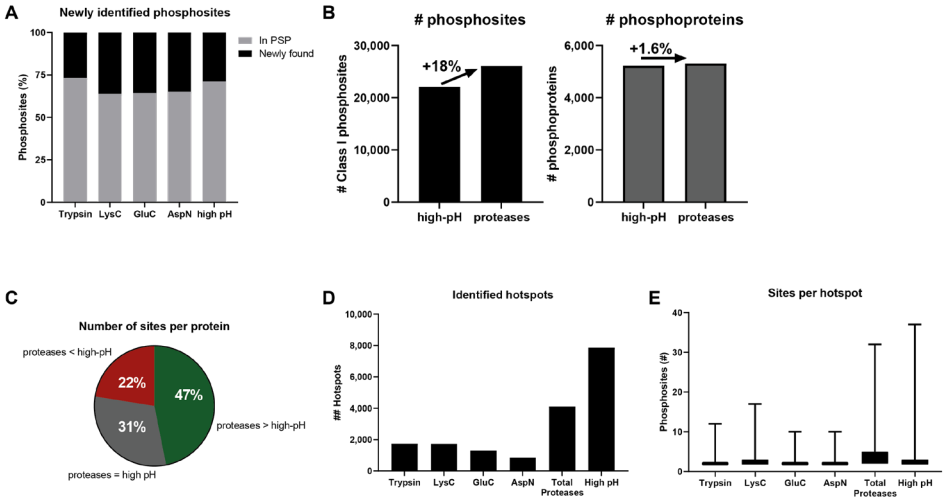


Figure 3 – Alternative proteases increase the phosphosite coverage per protein

A: The percentage of newly identified phosphosites from the total of all identified phosphosites is shown. PhosphoSitePlus (PSP) was used as the reference database and results are depicted for either the enzyme separately or after high-pH fractionation. **B**: The total number of phosphosites (left) or phosphoproteins (right) are shown for either high-pH fractionation or the complete dataset after alternative enzyme digestion. Although with the use of multiple enzymes 18% more phosphosites were identified, they originated from the same set of phosphoproteins, for these only increased with 1.6%. **C**: Of all phosphoproteins, the total number of identified phosphosites was counted, and compared between the use of alternative proteases or high-pH fractionation. Green: more sites per protein were identified using multiple enzymes, grey: the same number was identified, red: more sites per protein were identified using high-pH fractionation. **D**: Of all conditions the total number of phosphorylation hotspots was shown. **E**: Of all conditions the phosphorylation hotspot density (i.e. the number of phosphosites within a single hotspot) was shown.

ly increase compared to the use of trypsin only, despite the additional fractionation step. Indeed, the number of sites per protein was larger in the case of 47% of the phosphoproteins when using multiple proteases, whereas for only 22% of the proteins, the phosphosite coverage was larger after high-pH fractionation (Fig 3C).

It has been proposed that phosphorylation of a single amino acid does not always affect biological activity, but in fact phosphorylation of a cluster of amino acids, called phosphorylation hotspots, causes a protein to change its activity, whether it being change in localization, enzymatic activity or conformation [22–24]. Considering this observation, the detection of a stretch of phosphorylated amino acid could be biological more meaningful compared to the detection of a single site. To elaborate on this idea of hotspots, we explored whether in our data clusters of residues in close proximity could be observed. To do so we extracted phosphorylation hotspots from the data by looking for phosphosites in close proximity of each other (i.e. distance of <8 amino acids between two adjacent sites).

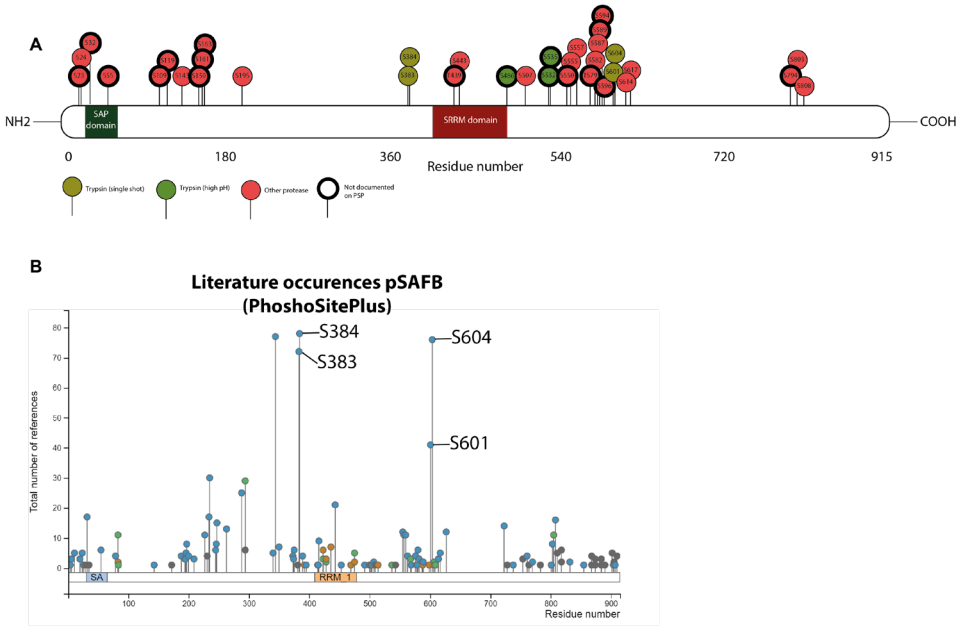


Figure 4 – Phosphorylation site coverage of SAFB greatly benefits from the use of alternative enzymes

A: the sequence of SAFB, in which known protein domains are depicted. Phosphosites identified in either dataset were mapped onto the protein sequence. Within the circle the site is specified. Yellow: identified by trypsin (single shot); green: identified by trypsin with additional high-pH fractionation; red: identified by a peptide produced by an alternative enzyme. **B**: graphical representation from PhosphoSitePlus website, indicating the number of citations per phosphosite, as mapped on the protein sequence of SAFB. The four most often documented sites were annotated to show the overlap with the experimental identified phosphosites.

For all enzymes a significant number of hotspots could be identified, indicating that enzymes generated peptides from overlapping or adjacent regions within the protein (Figure 4A). The numbers for AspN and GluC were lower when compared to trypsin and LysC. This is probably due to a lower number of sites that could be detected using these enzymes. Using a classical shotgun approach, it can be difficult to observe adjacent sites due to under-sampling of the phosphoproteome and the difficulty to identify peptides containing multiple phosphorylated residues. Therefore, it is not surprising that when the data from multiple proteases was combined the number of detectable hotspots massively increased, due to an overall increase in the coverage of individual (phospho)proteins (Figure 2D and 3D). Additionally, while using high-pH fractionation, the number of identified phosphorylation hotspot increased even further (Figure 3E). The reason for this increase in the amount of phosphorylation hotspots for high pH fractionation is probably due to an overall wider sampling of the phosphoproteome in which new hotspots that are separated by more than 8 amino acids will be identified. In contrast, we anticipate that for the different proteases new, adjacent sites within the same proteins and same hotspot regions will be detected. In agreement with this hypothesis, when sampling the overall density of hotspots (i.e. the number of sites within a hotspot) a massive increase in density is observed when using multiple proteases compared to the individual enzymes (Figure 3E). As expected, the increase in density was less elaborate while using a high pH strategy. To illustrate the increased phosphosite coverage of proteins, the phosphoprotein SAFB was studied in more detail and used as an example (Figure 4). In total, 76 individual phosphosites have been documented for SAFB on PhosphoSitePlus. When using trypsin as a digestion enzyme, without any additional fractionation, only 4 phosphosites were detected (Fig 4A), which are also amongst the most often identified sites according to PhosphoSitePlus (Fig 4B). Additional fractionation by high-pH identified three more phosphosites, which were surprisingly not documented on PhosphoSitePlus. The use of three additional enzymes, even without additional fractionation, added a surprising 28 sites, of which 15 were not documented on PhosphoSitePlus (Fig 4). This clearly showcases the overrepresentation of sites that are more readily identified using trypsin as a digestion enzyme on PhosphoSitePlus, illustrating the added value of either fractionation or the use of alternative enzymes.

Orthology in enzymes increases coverage of the proteome

In addition to increasing phosphoproteome coverage, the use of alternative enzymes can also increase proteome coverage. Here, full proteome digests were run in triplicate using a single shot LC-MS/MS run, using a 3 hour gradient.

As can be seen, using trypsin or LysC as a digestion enzyme the highest numbers of protein groups could be identified (Figure 5A). The number

of identified protein groups for GluC and AspN was lower, probably due to inefficient LC-MS/MS workflow settings which were optimized for tryptic peptides. Combination of all protease data did increase the number of protein groups that could be identified, although not to a great extend (Figure 5A). To show the full complementarity of the data, all identified protein groups were mapped into a heat map, showing the signatures of the different digestion enzymes used (Figure 5B). As can be seen, several proteins groups were exclusively identified using a specific enzyme, whereas the majority of protein groups was identified equally well by all enzymes (Figure 5B).

However, despite the little increase in the identification of distinct protein groups, the use of alternative enzymes could provide a larger coverage of individual protein sequences, similarly to what has been observed for the phosphoproteome. Indeed, when evaluating the number of unique peptides per protein one can see that the number of unique peptides per enzyme was rather small, especially for AspN and GluC (Figure 5C). However, when combining data from all different enzymes it becomes immediately clear that the number of unique peptides that could be identified per protein was significantly larger than individually used enzymes, highlighting the complementarity of the enzymes (Figure 5C). When plotting the

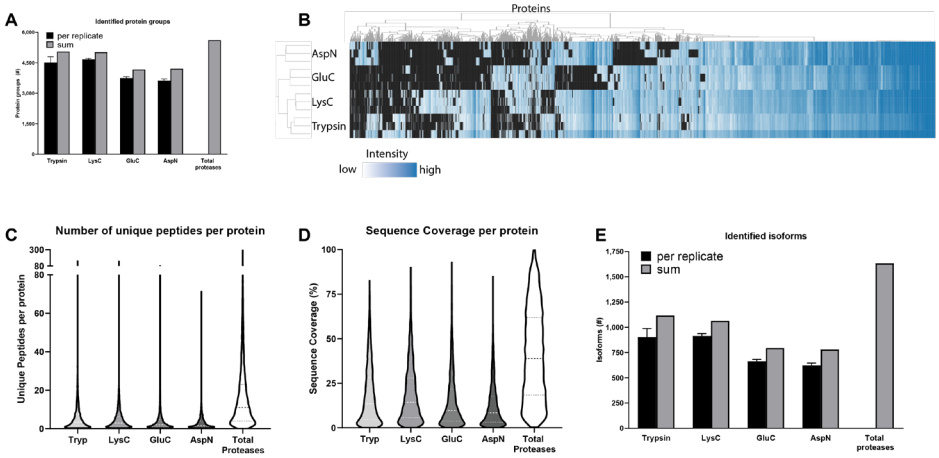


Figure 5 – Proteomic sequence coverage greatly increases by using alternative enzymes

A: Total number of protein groups was shown of separate enzymes as well as the combined dataset. In black the average number of protein groups per condition is shown, in which the error bars represent the different replicates. The grey bar shows the summed amount of all replicates of that condition. **B:** All identified protein groups (vertical clustering) in all replicates (horizontal clustering) were clustered in a hierarchical clustering, based on their normalised intensity. **C, D**: A violin plot shows the distribution of number of unique peptides **(A)** and the sequence coverage **(D)** per condition. **E:** Total number of isoforms that could unambiguously be identified were shown for separate enzymes as well as the combined dataset. In black the average number of isoforms per condition is shown, in which the error bars represent the different replicates. The grey bar show the summed amount of all replicates of that condition.

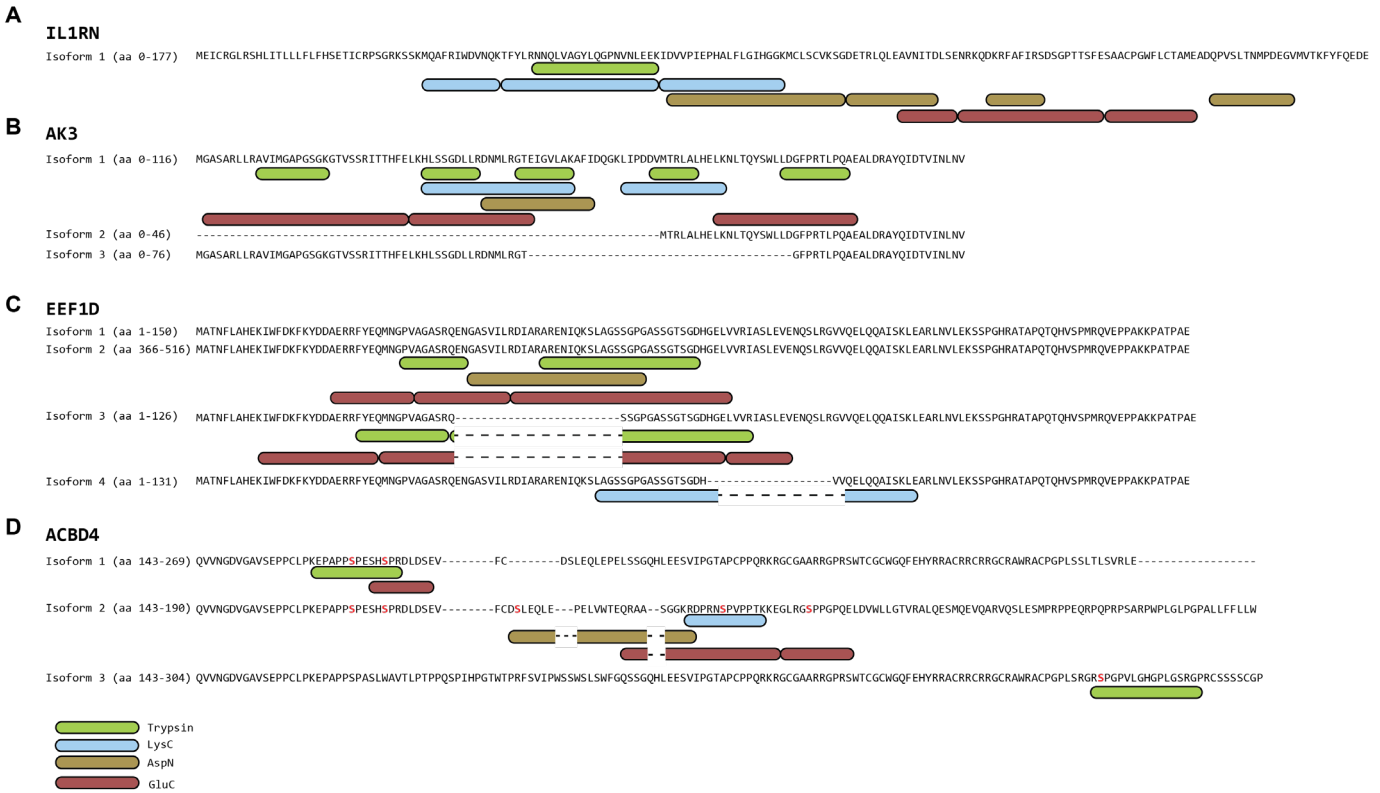


Figure 6 – Detailed view of isoform specific (phospho)peptides.

Protein sequences or parts of protein sequences are shown for IL1RN **(A)**, AK3 **(B)**, EEF1D **(C)** and ACBD4 **(D)**. If applicable, several isoforms were shown, as indicated on the left. Coloured bars indicated identified (phospho)peptides, placed below the corresponding isoform; green: trypsin; blue: LysC; orange: AspN; red: GluC, as indicated in the legend. Dashed lines in protein sequences indicate deletions as compared to one of the other isoforms. Dashed lines in indicated peptides indicate the spanned area of the isoform indicated immediately above the peptide.

sequence coverage per protein, a similar effect could be observed. Whereas the sequence coverage was low for especially AspN, when combining the data from all different enzymes, the sequence per coverage massively increased (Figure 5D). This clearly indicates that when aiming to solve the existence of different proteoforms (defined as a specific molecular form of a protein product arising from a specific gene) a multiple enzyme approach is beneficial and could be used to uncover post-translational modifications beyond phosphorylation over the entire protein sequence.
To illustrate this even further, for each enzyme used, isoform specific peptides were filtered out of the data, allowing for the unambiguous identification of specific isoforms. For each enzyme a significant number of specific isoforms could be detected (Figure 5E). The addition of replicates from the same enzyme caused an additional increase in number of isoforms that could be identified, indicating under sampling in the mass spectrometer. However, the combination of all datasets massively increased the number of specific isoforms to be detected with ~70%.

The use of multiple proteases reveals more detail at the single protein level.

In order to illustrate the increased proteome coverage and the potential to identify specific isoforms a selection of proteins was shown in more detail. For the Interleukin-1 receptor antagonist protein (IL1RN) depicted in Figure 6A, it could be clearly seen that the use of multiple proteases massively increased the sequence coverage that can be achieved for this protein. When using trypsin alone, and performing a single shot analysis, only a single peptide could be identified. The lack of coverage by trypsin could in this case probably be attributed to under sampling in the mass spectrometer, even though in this case the triplicate analysis did not aid in identifying another peptide from this protein sequence. Indeed, when looking at the C-terminal part of the protein, a lack of tryptic cleave sites most likely contributed to the low coverage when using trypsin. In other parts of the protein sequence, trypsin produced peptide fragments that would either be too long or too short to be efficiently analysed. Consequently, in this specific case, especially the use of GluC and AspN significantly enhanced the sequence coverage of this protein.
Besides increasing sequence coverage, the number of specific isoforms that could be detected increased with the combination of proteases used (Figure 5E). As an illustration, a selection of identified isoform using multiple proteases was depicted in Figure 6B-D. For the GTP:AMP phosphotransferase AK3, the use of multiple proteases provided evidence for isoform 1 to be present in the sample, whereas the presence of isoform 2 and 3 cannot be excluded from this data (Figure 6B). For the protein EEF1, especially the use of trypsin and GluC is beneficial to provide evidence for the presence of isoform 1, 2 or both (Figure 6C). In addition, these 2 enzymes were beneficial for providing evidence for the presence

of isoform 3, which has a short deletion in the middle of the sequence. Isoform specific peptides covering this unique sequence provided very strong evidence for the presence of this isoform (Figure 6C). Finally, isoform 4 could be identified solely due to the the use of LysC as a digestion enzyme which generated a quite large peptide that could be used to unambiguously identify this isoform (Figure 6C).
Finally, also isoform specific phosphorylated peptides can be used to distinguish specific isoforms of proteins. As illustrated for the protein ACBD4, phosphopeptides indicated the presence of either isoform 1 or 2 (Figure 6D). However, isoform specific peptides generated by LysC, AspN and GluC indicated a clear presence of isoform 2. An isoform specific peptide for isoform 1 could not be detected thus therefore its presence cannot be unambiguously confirmed. Finally, a tryptic phosphopeptide peptide at the C-terminal end of isoform 3 unambiguously identified its presence in the sample (Figure 6D). Clearly these results indicate and emphasize that the use of multiple proteases are not only beneficial to increase the proteome coverage, with an emphasis on identifying more protein groups, but is also extremely useful to delineate the presence of specific isoforms in a sample. With the growing interest in proteoform specific research, the use of multiple proteases should therefore be considered as a valuable tool and could aid in the elucidation of proteoform specific biological functions.

Discussion

Despite a slow but steady maturation of the field of proteomics, major efforts are still made to enhance the coverage of the 'proteome'. Recent publications on, for example the draft of the human proteome [1,2] highlight these efforts. Still, due to the extremely dynamic nature of the proteome, this quest will probably never end. In addition, due to the way proteomes are analysed, i.e. by measuring peptides and not proteins, the peptide inference problem poses another major challenge on identifying all possible proteins as well as proteoforms within a sample.
Here we have set out to increase the proteome as well as proteoform coverage by using alternative proteases beyond trypsin, or high-pH fractionation. The idea of using multiple proteases to enhance proteome coverage is not new and has been performed by others [8,13]. However, due to a recent optimization of the phosphoproteomic workflow, which includes removal of contaminants and an optimized enrichment, this approach recently became of interest to digestion with multiple proteases as well. Indeed, when compared to a recent paper on this approach (Supplementary Figure S3), our optimized method allowed for the enhanced identification of phosphopeptides and phosphosites.
As mentioned before, researchers are continuously attempting to increase

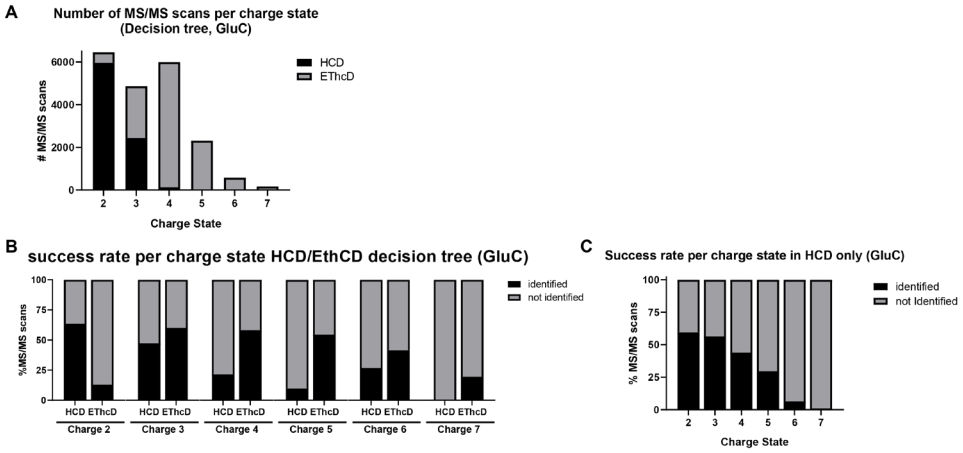
proteome coverage. Often applied methods to increase the number of identifications make use of fractionation techniques at either the protein or peptide level. Although these methods are extremely successful and are relatively easy to implement, they suffer from a massive increase in analysis time and sample input. More importantly, when using multi-dimensional fractionation, researchers should realise that what is lost in the first dimension can never be recovered. As such, the choice of the first dimension of separation is crucial for the outcome of the experiment. Here we compared the use of multiple proteases to the use of high-pH fraction for the comprehensive identification of phosphopeptides. To this end, tryptic peptides were fractionated on a high-pH enrichment with subsequent enrichment of the separate fractions on a Fe^{3+}-IMAC BRAVO system. High-pH fractionation has previously been shown to have excellent orthogonality to reversed-phase separation prior to LC-MS/MS analysis and is as such widely used across laboratories around the world.

However, despite a shorter analysis time, the use of multiple proteases still outperformed fractionation both at the level of phosphopeptides and phosphosites. More than the number of phosphosites that could be detected, the larger coverage of phosphoproteins proved to be a major advantage of the use of multiple proteases. The larger coverage of each phosphoprotein greatly enhances our understanding of the underlying implications of those phosphosites. In favour for the use of multiple proteases it should also be noted that the use of alternative enzymes provides a proportionally larger increase in the detection of 'new', previously not documented sites (referenced to PhosphoSitePlus). As most of the current repositories have a bias towards tryptic peptides, especially the use of AspN and GluC provided a substantial increase in newly discovered sites. Especially in a quantitative experiment, comparing different biological settings, the use of alternative enzymes could reveal, new, previously gone unnoticed regulatory sites. Surprisingly, researchers still preferably use fractionation over the use of alternative enzymes in their quantitative experiments. Here we strongly recommend the use of different proteases for these types of experiments, as they have proven their merit already more than once. In addition to this notion, the higher coverage of individual proteins enhanced the detection and density of phosphorylation hotspots. These hotspots are gaining interest in the last years as researchers become more and more aware that biological function of a protein could be actually regulated by the phosphorylation of multiple residues in a row other than a single site.

Besides being beneficial for phosphoproteomic studies, the use of multiple proteases also proved to be beneficial for enhancing proteome coverage and more specifically the sequence coverage of individual proteins. Here, it should be noted that the results were not compared to high pH fractionation, but we anticipate that a large increase in the number of protein groups would be observed, but a lower increase in the sequence coverage

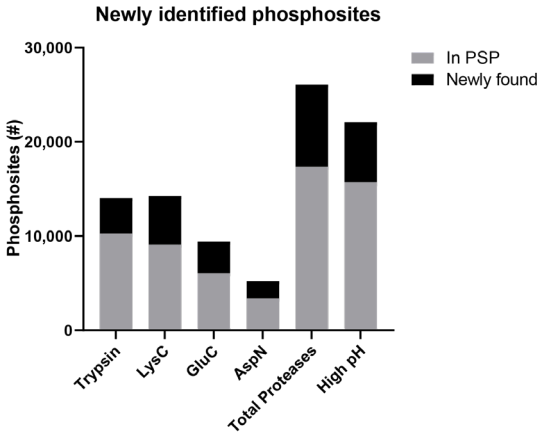
of individual proteins. Future research should be aimed at confirming this hypothesis.

The increase in protein coverage, both at the peptide and phosphopeptide level, allowed us to investigate the presence of isoform specific peptides. Indeed due to the increase coverage, as a consequence of the use of alternative proteases, isoform specific peptides could be detected that allowed us to unambiguously identify specific isoforms in the sample. Again, the growing interest in proteoforms and proteoform specific function could benefit from the use of alternative proteases. Building upon the observations we made in this work, we strongly recommend researchers to look beyond classical fractionation approaches and consider the use of alternative enzymes to enhance proteome coverage.



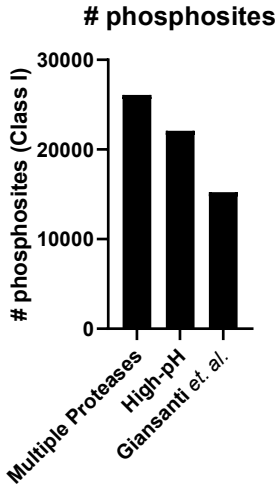
Supplementary Figure S1

A: Number of MSMS scans of GluC digested peptides as shown per charge state. Indicated in black: MSMS spectra where HCD was triggered, indicated in grey: MSMS spectra where EThcD was triggered. **B, C**: Successrate (i.e. percentage of successful identified MSMS spectra) was shown of GluC digested peptides within a HCD/EThcD decision tree **(B)** and GluC only **(C)**. Shown in grey is the percentage of unsuccessful MSMS spectra, in black the percentage of successful identified MSM sepctra.



Supplementary Figure S2

Total number of phosphosites identified after digestion with either enzyme, data from all enzymes combined or high-pH fractionation. In grey are shown the number of phosphosites that already have been documented on PhosphoSitePlus (PSP); in black the number of sites that were not yet documented on PSP.



Supplementary Figure 3

Comparison of number of total identified phosphosites with the use of our optimised protocol and multiple enzyme use, high pH fractionation or as compared to previous studie of Giansanti *et al* [8]

Table 1 – Overview of number of Class I phosphosites identified in specified amount of time, measured peptides per minute MS time, with or without additional fractionation and total mount of material used.

Table 1	Total no identified sites (Class I)	Total minutes MS time	Additional fractionation	Sites per minute	Material used
4x proteases, 3 replicates per enzyme	26,085	2160 (36 hour)	no	12	24 mg (4x3x2mg)
High-pH, 3 replicates	22,097	2700 (45 hour)	High-pH	8,2	6 mg (3x2mg)

Materials & Methods

Cell culture
A431 cell were cultured at 5% CO_2 and 37°C in DMEM (Dubleco's) medium, containing 10% fetal bovine serum (Invitrogen), 1%L-glutamine (Lonza) and 1% penicillin (Lonza). Phosphorylation events were boosted by refreshing the medium 4 hours before harvesting. The cells were harvested by trypsinization and washed two times with PBS-buffer.

Sample preparation
Cell pellets were lysed in lysis buffer containing 7M Urea,1%(v/v) Triton x-100(Sigma-Aldrich) , 5mM Tris(2-carboxyethyl)phosphine hydrochloride(TCEP, Sigma-Aldrich), 30mM chloroacetamide (CAA, Sigma-Aldrich), 1 tablet Complete Mini EDTA-free protease inhibitor Cocktail (Roche), 1 tablet PhosSTOP Phosphatase Inhibitor Cocktail (Roche), 10 U/ml DNase I, 1mM sodium orthovanadate, 1%(v/v) Benzonase(Merck Milipore), and 2mM magnesium chloride(Sigma-Aldrich) in 10mL 100mM Tris(pH 8.5). The cells were disrupted by sonication for 20 minutes (30s on, 30s off) using a Bioruptor Plus (Diagenode). Cell debris was removed by centrifugation at 20.000g for 10 minutes. For incubation of the Benzonase the collected supernatant was kept at room temperature for 2 hours. The protein concentration was determined by BCA assay (Pierce).
The lysate was cleaned up by methanol/chloroform protein precipitation as follows: four volumes of methanol (Sigma-Aldrich) were added to one volume of protein sample, and the mixture was vortexed. One volume of chloroform (Sigma-Aldrich) was then added, and the mixture was vortexed. Three volumes of ultrapure water was added, and the mixture was vortexed. The mixture was centrifuged for 10 min at 5,000rpm at room temperature (RT). The aqueous methanol layer was removed from the top of the samples. Three volumes of methanol were added. After sonication and centrifugation (5.000 rpm, 10 min at RT), the solvent was removed and the precipitate was allowed to dry to the air.
Prior the digestion the samples were reconstituted in digestion buffer containing 100mM Tris-HCl (pH 8.5), 1% SDC, 5mM TCEP and 30mM CAA. Trypsin, LysC, AspN or GluC were added in a ratio of 1:100(w/w) for 4 hours after which an additional ratio of 1:100(w/w) was added. Digestion was preformed overnight at 37°C.

Desalting
Samples were acidified to pH 2 using 10% formic acid (Sigma-Aldrich) and centrifuged at 14000 rpm for 10 min at 4 °C. The supernatant was then loaded on 1cc Oasis HLB-prime (Waters) columns, washed with 2 × 1 mL of 0.1% formic acid and peptides were eluted with 80% acetonitrile (Sigma-Aldrich). Eluted peptides were split in 2 mg peptide aliquots and subsequently dried down using a speedVac centrifuge.

High pH reversed-phase (HpH) for peptide fractionation
2 mg of the desalted tryptic digest was reconstituted in buffer A (10mM ammonium hydroxide, pH 10) and loaded on a Kinetex Evo (Phenomenex) C18 5µm 100x2.1mm column using an Agilent 1100 binary pump. The peptides were eluted using a gradient initiated as follow: 0%B (10mM ammonium Hydroxide in 90% ACN, pH 10) to 14.3%B in 20 minutes, 35%B in 15 minutes and increased to 100%B in 1 minutes at a flow rate of 200 µl/min. In total 40 fractions of 1 minute were collected using an Agilent 1260 infinity fraction collector, and were pooled into 10 fractions using the concatenation strategy [17]. The pooled fractions were dried in a vacuum centrifuge and stored at -80°C for further use.

Fe-Imac High pH samples
The phospho peptides of the high pH fractions were enriched on the AssayMAP Bravo Platform (Agilent Technologies) using Fe (III)-NTA 5µL (Agilent technologies) cartridges. The cartridges were primed with 200 µL of 0.1 TFA in ACN and equilibrated with 200 µL of loading buffer (80%ACN/0.1% TFA). The High pH fractions were reconstituted in 210 µl of loading buffer and loaded at a 5 µl/min flow rate. The columns were washed with 250 µl, after which the phosphopeptides were eluted with 25 µl of 1% ammonia. The samples were dried down and stored at -80°C until analysis by LC-MS/MS

Fe-IMAC
After multi-enzymatic digestion, peptide mixtures were enriched for phosphopeptides by Fe-IMAC. Per enzyme, 3x2mg of dried peptides were reconstituted in buffer A (30% ACN, 0.07% TFA) and loaded on to the Fe-IMAC column (Propac IMAC-10 4x50 mm column) [14] with a flow rate of 0.1 mL/min. After loading, the column was washed with buffer A for 5 minutes with a flow rate of 1 mL/min. Next, peptides were eluted from the column with 50% buffer B (0.3% NH4OH), with a flow rate of 1 mL/min for 1.5 minutes and subsequent 0.5 mL/min for 2.5 minutes. The samples were dried down and stored at -80°C until analysis by LC-MS/MS.

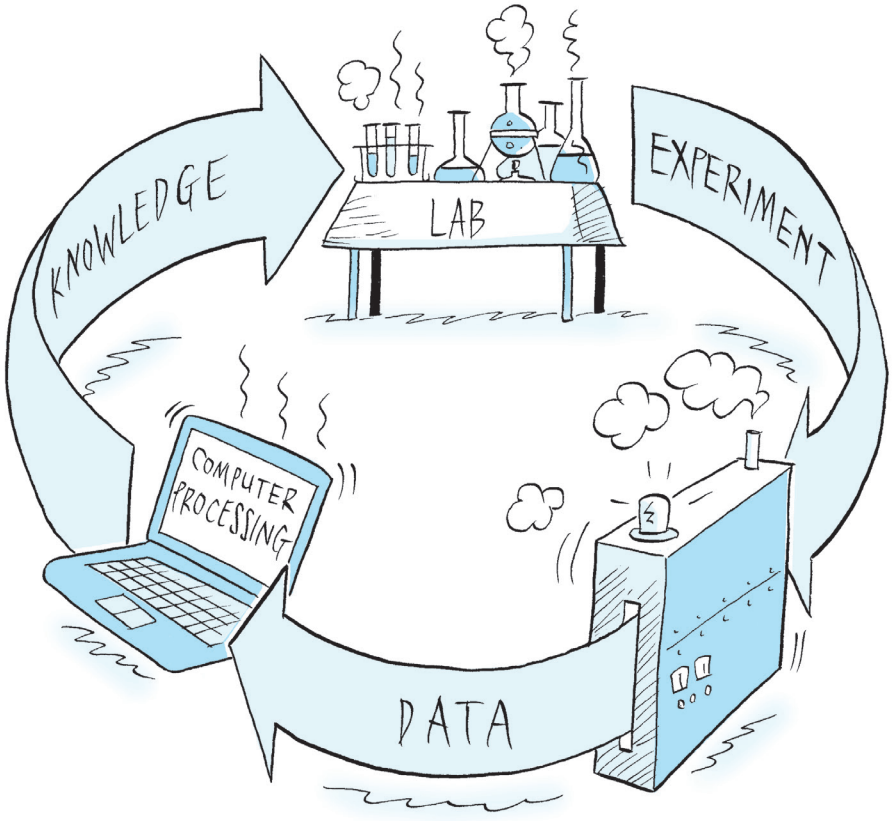
LC-MS/MS
Nanoflow LC-MS/MS was performed by coupling an Agilent 1290 (Agilent Technologies, Middelburg, Netherlands) to an Orbitrap Q Exactive HF (Thermo Scientific, Bremen, Germany) or Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific, Bremen, Germany). The peptides were reconstituted in 10% formic acid and loaded on the trap column (100 µm i.d. × 2 cm, packed with 3 µm C18 resin, Reprosil PUR AQ, Dr. Maisch, Ammerbuch, Germany) with a flow rate of 5 µl/min with solvent A(0.1% formic acid) for 5 minutes. After loading, the peptides were separated on an analytical column (75 µm × 60 cm C18 column Reprosil PUR AQ, 3 µm, Dr. Maisch, Ammerbuch, Germany) at a flow rate of 300nl/min, using a linear gradient initiated at 6% solvent B (80% ACN/0.1% formic acid) increasing to 32% (phosphopeptides) or 10% to 36% solvent B (full proteome) over 155 minutes. For the optimization, 20 percent of the samples was injected and 90 minute methods were used instead of 175 minutes were the gradient was proportionally adjusted.
Q-Exactive HF was used in positive mode with a spraying voltage of 1.9kV and transfer capillary temperature of 275 °C. The instrument was used in data dependent mode with a full scan range of 375-1600 m/z, resolution of 60.000, a maximum injection time of 20ms, and AGC target of 3e6 in MS. The top 15 (top12 for phosphopetides) most intense precursor ions of each cycle were selected for HCD fragmentation. The MS^2 spectra were obtained at a resolution of 30.000 with an AGC target of 1e5 and a maximum injection time of 50ms, 1.4 m/z isolation window, and normalized collisional energy (NCE) of 27.
The Lumos was used in positive mode with a spraying voltage of 1.9kV and transfer capillary temperature of 275 °C. The survey scan range was from 350-1500 m/z at a resolution of 60 000 with an AGC target of 4e5. The most intense precursor ions were selected for fragmentation within a duty cycle of 3 s. The selected precursor ions were fragmented according to a decision tree scheme. Precursor ions in the m/z range of 650-1200 and a charge state of 2 or m/z range of 950-1200 and charge state of 3 were selected for HCD fragmentation, all other ions were selected for EThcD. The MS^2 spectra were obtained at a resolution of 30.000 with an AGC target of 1.0e5 and a 1.6 isolation window. When HCD was applied the collision energy was 35%, and for EThcD the supplemental activation was 40%.

Data analysis
All raw file were processed using the MaxQuant software (version 1.6.2.3) and the spectra were searched against the Uniprot human database (August 2016). The following parameters were used for the Q Exacitve HF raw files, enzyme was set depending on the corresponding enzyme, maximal allowed missed cleavage was set to 2, carbamidomethylation of cysteines as a fixed modification, methionine oxidation and phosphorylation on serine, threonine and tyrosine residues as variable modifications.
For the Lumos raw files the same parameters were used except for the allowed missed cleavages, this was set to 8. The output file evidence was used for determining the number of phosphopeptides, and the peptide characteristics. The Phospho(STY)Sites file was used for determine the phosphosites found. Phosphosites with an localization probability higher than 75% were considered as Class I sites and was used as a cut-off for the Phospho(STY) Sites table. For the comparison of the additional found phosphosites compared to the PhosphoSitePlus database an in-house made R script was used.

References

1. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, et al.: **A draft map of the human proteome**. *Nature* 2014, **509**:575–581.
2. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, et al.: **Mass-spectrometry-based draft of the human proteome**. *mass Spectrom Rev* 2013, **32**:5453–465.
3. Potel CM, Lin M-H, Heck AJR, Lemeer S: **Defeating major contaminants in Fe 3+-IMAC phosphopeptide enrichment**. *Mol Cell Proteomics* 2018, **17**:1028–1034.
4. Moore SM, Hess SM, Jorgenson JW: **Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics**. *J Proteome Res* 2016, **15**:1243–1252.
5. Tsiatsiani L, Heck AJR: **Proteomics beyond trypsin**. *FEBS J* 2015, **282**:2612–2626.
6. Vandermarliere E, Mueller M, Martens L: **Getting intimate with trypsin, the leading protease in proteomics**. *mass Spectrom Rev* 2013, **32**:453–465.
7. Guo X, Trudgian DC, Lemoff A, Yadavalli S, Mirzaei H: **Confetti: A Multiprotease Map of the HeLa Proteome for Comprehensive Proteomics**. *Mol Cell Proteomics* 2014, **13**:1573–1584.
8. Giansanti P, Aye TT, van den Toorn H, Peng M, van Breukelen B, Heck AJR: **An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas**. *Cell Rep* 2015, **11**:1834–1843.
9. Gonczarowska-Jorge H, Loroch S, Dell'Aica M, Sickmann A, Roos A, Zahedi RP: **Quantifying Missing (Phospho)Proteome Regions with the Broad-Specificity Protease Subtilisin**. *Anal Chem* 2017, **89**:13137–13145.
10. Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP: **Large-scale characterization of HeLa cell nuclear phosphoproteins**. *Proc Natl Acad Sci* 2004, **101**:12130–12135.
11. Guerreiro ACL, Penning R, Raaijmakers LM, Axman IM, Heck AJR, Altelaar AFM: **Monitoring light/dark association dynamics of multi-protein complexes in cyanobacteria using size exclusion chromatography-based proteomics**. *J Proteomics* 2016, **142**:33–44.
12. Batth TS, Francavilla C, Olsen J V.: **Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics**. *J Proteome Res* 2014, **13**:6176–6186.
13. Cristobal A, Marino F, Post H, Van Den Toorn HWP, Mohammed S, Heck AJR: **Toward an Optimized Workflow for Middle-Down Proteomics**. *Anal Chem* 2017, **89**:3318–3325.
14. Ruprecht B, Koch H, Medard G, Mundt M, Kuster B, Lemeer S: **Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns**. *Mol Cell Proteomics* 2015, **14**:205–215.
15. Post H, Penning R, Fitzpatrick MA, Garrigues LB, Wu W, Macgillavry HD, Hoogenraad CC, Heck AJR, Altelaar AFM: **Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons**. *J Proteome Res* 2017, **16**:728–737.
16. Hornbeck P V., Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E: **PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations**. *Nucleic Acids Res* 2015, **43**:D512–D520.
17. Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, Shen Y, Monroe ME, Lopez-Ferrer D, Reno T, et al.: **Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells**. *Proteomics* 2011, **11**:2019–2026.
18. Potel CM, Lemeer S, Heck AJR: **Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update**. *Anal Chem* 2019, **91**:126–141.
19. Nagaraj N, D'Souza RCJ, Cox J, Olsen J V., Mann M: **Feasibility of Large-Scale Phosphoproteomics with Higher Energy Collisional Dissociation Fragmentation**. *J Proteome Res* 2010, **9**:6786–6794.
20. Frese CK, Zhou H, Taus T, Maarten Altelaar AF, Mechtler K, Heck AJR, Mohammed S: **Unambiguous Phosphosite Localization using Electron-Transfer/ Higher-Energy Collision Dissociation (EThcD)**. 2013, doi:10.1021/pr301130k.
21. Hornbeck P V., Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E: **PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations**. *Nucleic Acids Res* 2015, **43**:D512–D520.
22. Regev S, Michal L: **Cooperativity within proximal phosphorylation sites is revealed from large-scale proteomics data**. *Biol Direct* 2010, **5**:6.
23. Christian J-O, Braginets R, Schulze WX, Walther D: **Characterization and Prediction of Protein Phosphorylation Hotspots in Arabidopsis thaliana**. *Front Plant Sci* 2012, **3**:1–14.
24. Torres MP, Dewhurst H, Sundararaman N: **Proteome-wide Structural Analysis of PTM Hotspots Reveals Regulatory Elements Predicted to Impact Biological Function and Disease**. *Mol Cell Proteomics* 2016, **15**:3513–3528.

THREE



CHAPTER FOUR

Transformation of phosphosites to phosphorylation hotspots enables data reuse and (re)analysis

Celine Mulder[1] and Simone Lemeer[1]

[1] *Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands*

Introduction

FOUR

Shotgun mass spectrometry (MS) based proteomics can be used to study a biological question, for instance by performing a discovery-based experiment comparing different conditions. The analysis of altered protein abundances and post-translational modifications (PTMs) will reveal a detailed view of the studied models at the molecular level. Furthermore, a discovery-based experiment allows for the detection of new, unknown changes beyond existing biological knowledge. Unfortunately, since it is not yet possible to measure the entire proteome, missing parts of pathways could lead to inaccurate and/or incomplete conclusions. Therefore, a never ending venture in the field of proteomics is increasing the amount of data that can be acquired within a single experiment by for example fractionation [1], the use of complementary protocols [2] and/or optimized MS analysis [3–5]. Paradoxically, downstream data processing of the acquired data and subsequent interrogation of biologically meaningful events remains a major challenge. Unfortunately, wet-lab scientists are often no experts in the field of 'big data' analysis and statistics, whereas bioinformatic researchers usually are not fully aware of the experimental design and underlying biology of the experiment. This miscommunication often causes large parts of data from successfully conducted experiments to remain unexplored.

In practice, researchers aim to answer their research question by mapping proteome data on known pathways, often only using a minority of the generated data. Additionally, due to the enormous amount of data generated in a shotgun proteomic analysis, new questions will be raised, and new hypotheses will be generated. Therefore, validation of shotgun proteomic experiments has become a crucial part of today's proteomic research. Validation by additional orthogonal experiments is probably the most conclusive way of confirming observations. However, validation is time-consuming, expensive and susceptible to failure in case of wrongly interpreted proteomic results. Therefore, correct and comprehensive data interpretation is essential to generate a correct hypothesis after discovery experiments and consequently reduce the chance of failing in the validation phase.

All of the above-mentioned issues, including the need to increase the amount of experimental data, the feasibility of data interpretation, the large amount of data that is lost after processing and the necessity to arrive at the best possible hypothesis for validation can be improved by reusing publicly available data. Unfortunately, data reusing is no common practice in proteomic research, whereas it is regularly applied in genomic research [6–8]. Unlike genomics, proteomics is a relatively young field and therefore pipelines and organization of data repositories is not fully optimized yet [9]. Furthermore, unlike for the genomics field, it is not yet a custom to implement data reuse in biological projects, despite repeated

recommendations of experts in the field [10]. Fortunately, although minor, efforts are made to enable proteomics scientists to engage in data sharing. For most, the foundation of ProteomeXchange [11] was a great step forward, although the amount of data available is yet only a fraction compared to the amount of publicly available genomic data. More and more examples can be found in literature which describe reuse of data for the construction of spectral libraries for targeted MS [12], characterize technical aspects of the MS workflow [13,14], or validate new bioinformatics tools [15,16]. Another type of reuse is exemplified in the collections of entire proteomes, like the first drafts of the human proteome [17,18] and the phosphoproteome of *Arabidopsis thaliana* [19]. Lastly, large protein resources like Uniprot [20] and PhosphoSitePlus [21], use publicly available data.

Although various examples of proteomic data reuse exist, only very few describe the use of quantitative data [22,23]. While successful, they also voice great hindrance in the analysis due to the limited amount of data that turned out to be compatible for their comparison. Most studies that, at first sight, were useful, actually lacked the proper meta-data needed for quantitative comparisons and could therefore not been included. For these reasons, in 2007 the Human Proteome Organization's Proteomics Standards Initiative (HUPO-PSI) developed the Minimal Information about a Proteomics Experiment (MIAPE), attempting to equalize data representation, in order to enforce scientists to report their studies in such a way that data reuse was possible and desired information could easily be extracted [24].

In this study we focus on phosphoproteome data. The addition of a phosphate group to a protein has major biological implications, including changes in localization, degradation, (in)activation or protein-protein binding [25]. Despite the known effects of phosphorylation, the established body of knowledge about the biological functions of specific phosphosites is rather limited, in contrast to the size of the entire phosphoproteome. According to the documentation of the database PhosphoSitePlus [21] at time of writing (December 2018), from the >235,000 ever documented phosphosites, only ~6,900 (2.9%) have been annotated or associated with some form of biological function. This is due to the fact that the identification of phosphorylated amino acids is performed with high throughput approaches (i.e. phosphoproteomics), but the validation and determination of the biological function of a specific phosphosite is achieved with low throughput assays (e.g. knock out or mutational assays). Although a shotgun proteomics experiment will not lead to the discovery of the immediate biological function of any observed site, since correlation does not imply causation, this type of high-throughput profiling can be used to define the best choice of follow up targets.

In this study, data reuse was applied on a set of quantitative phosphoproteomic datasets, previously published and publicly available. In order to

increase the feasibility of combining different datasets which originated from different laboratories, all identified and quantified phosphosites were transformed into phosphorylation 'hotspots'. These hotspots consisted of multiple, adjacent, phosphosites which were considered as one in order to facilitate comparison between different datasets. Indeed, hotspot analysis increased the overlap between different datasets. Subsequently, we investigated whether such a hotspot analysis is a valid method for combining quantitative datasets. Our results indicate that reuse of data provides an unbiased prioritization of downstream targets for validation and in addition provides protein centric information on differential regulation of hotspots, dependent on different experimental conditions.

Results

For this study, 11 distinct datasets, of which 10 were previously published and 1 made in house, were used for further data (re)analysis (Figure 1A). Of these 11 studies, 5 studies were describing distinct biological model systems (including: melanoma, rhabdoid tumor, HeLa cells, leukemia and retina epithelium studies), whereas the other 6 studies described biologically similar HER2/ER+ breast cancer models (Figure 1A). These 6 datasets described a biological model of acquired resistance against targeted therapy, targeting either HER2 or the Estrogen receptor (ER) in breast carcinomas [26–29]. Both receptors are upstream of highly similar growth stimulating pathways and documented resistance mechanisms have shown to be comparable. Although different drug targets and different cell lines were described in the included studies, global regulations were shown to be similar. Co-occurrence of signaling events, especially in disparate experimental conditions, increases the confidence of the biological value of the observed regulation. In addition, combining data from multiple experiments increases the global understanding of the studied models, making it easier to pick the most promising follow-up targets.
In a first round of analysis, the complete set of 11 datasets was used to demonstrate the power of combining datasets. In the second round, the analysis was performed only using the 6 breast cancer studies, in order to investigate whether phenotypic results could be obtained.

Complementarity versus overlap of datasets

Repetitive analysis of a sample, by using more replicates, will not result in a substantial increase in the amount of data, as demonstrated by the accumulation of relative increase (Figure 1B). In contrast, the combination of the 6 breast cancer studies, originating from different laboratories caused the total amount of data to increase with every dataset added, thereby increasing the coverage of the studied model with great extent (Figure 1B). Naturally, the increase in amount of data, when combining complete datasets, was larger compared to the addition of replicates be-

cause there was simply a lot more data within a single dataset (which probably contains several replicates itself). This can be seen by the mere fact that the extent of increase depends on the size of a dataset added. The more data present in the initial dataset, the smaller the increase in total amount of data after combining (Figure 1C). Despite this, it is still obvious that without additional lab work or instrument time, it is possible to massively increase the coverage of a studied model by combining datasets.

The total combined dataset consisted of >100,000 phosphosites (Figure 1C). As expected, the main factor contributing to this large number was the lack of overlap between the datasets. Only ~35% of all identified phosphosites were identified in 2 or more datasets (Figure 1D). Furthermore, of these 100,000 sites, >23,700 (22%) were significantly regulated in at least one condition, of which only ~30% was significantly regulated in 2 or more datasets (Figure 1E). When only the 6 breast cancer relat-

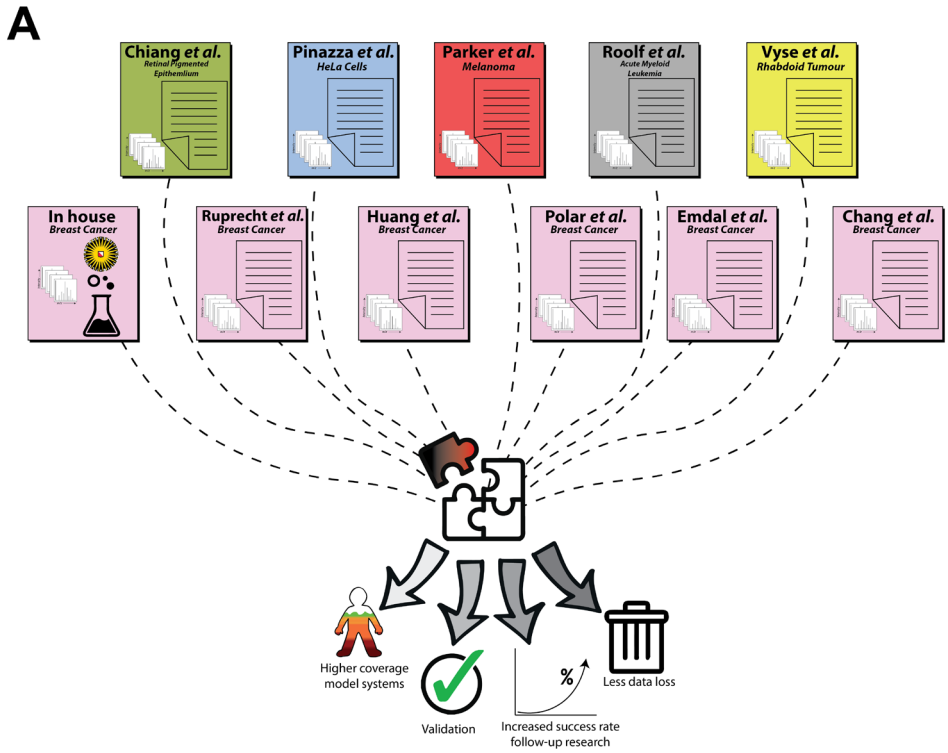


Figure 1 – Data reuse increases amount of data but has low overlap

A: in total 11 datasets were used to generate a large new phosphorylation dataset. 5 studies described various biological models (retinal pigmented epithelium, HeLa Cells, melanoma, AML and rhabdoid tumor cells), whereas the other 6 all described models of drug resistance in breast cancer. *(continued on next page)*

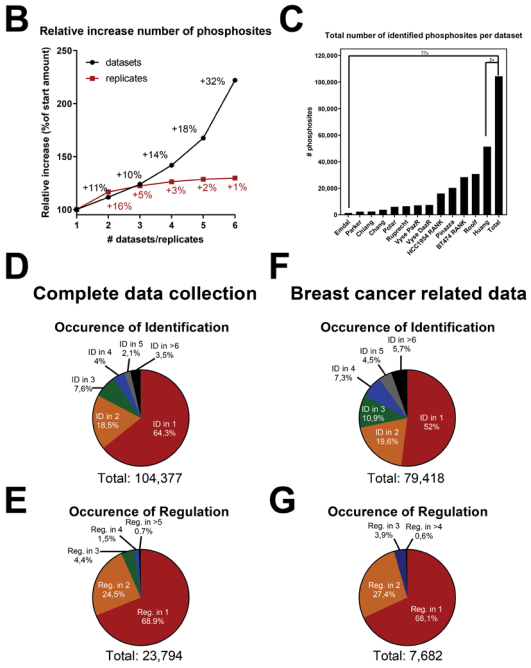


Figure 1 – Data reuse increases amount of data but has low overlap

(continued from previous page) **B**: The dataset of Ruprecht *et al* was used to show the accumulation of data if the number of replicates were increased (red line). The 6 breast cancer datasets were used to describe the accumulation of data if different datasets were used (black line). The first dataset or replicate was set to 100%, after which the percentage of increase of number of phosphosites was calculated after the addition of a replicate or dataset. **C**: Total number of phosphosites within each dataset used was plotted (y-axes), in addition to the total count after all datasets were merged. As an indication, the percentage of increase after merge was calculated for the smallest dataset (Emdal *et al*) or the largest datasets (Huang *et al*), as indicated by the horizontal bars. **D-G**: For each phosphosite included in the total dataset it was counted in how many different datasets a site was identified or regulated. For all pie plots, the total of included sites is shown below the plot, and each section is labelled with its percentage. This was done for the complete set of 11 datasets, counting either identification **(D)** or regulation **(E)**. The same was done, but only considering all breast cancer related studies, counting either identification **(F)** or regulation **(G)**.

ed studies were used, the overlap was higher, despite the fact that the dataset consisted of less datasets (Figure 1F). This can be explained by the fact that the cellular models used in these studies are more alike. Furthermore, since the same cellular pathways were affected due to similar treatments used across the different studies, the same proteins and phosphosites will be regulated (Figure 1G).

Although the complementarity of combined datasets can be used to increase proteomic coverage, quantitative data benefits from overlapping data. Repetitive identification between replicates is beneficial for (label-free) quantification and repeated observed regulation of a protein or phosphosite increases its value and credibility in the process of becom-

ing a favorable follow-up target. Thus, the main added value of combining datasets from different experiments is the increased coverage of the (phospho)proteome of a similar cell line or biological model. However, in order to add biological context to this combined dataset or in order to identify the most important players in specific biological systems, the data will need to be transformed in such a way that the overlap increases and repeated identification as well as quantification of a phosphosite is enabled.

Merging phosphosites into phosphorylation hotspots increases overlap

It has been hypothesized that biological activity which is controlled by phosphorylation is not so much regulated by the modification of a single amino acid, but by the phosphorylation of a cluster of residues in close proximity of each other [30–32]. The study of these so called 'phosphorylation hotspots' is challenging when using a classical shotgun approach, due to under-sampling of the phosphoproteome and the difficulty to identify peptides containing multiple phosphorylated residues. However, by combining multiple datasets, the increase in coverage due to the complementarity of the data (Figure 1A) can be used to the researchers advantage while mapping these hotspots [30].

In this study, 'phosphorylation hotspots' were defined as several phosphosites in close proximity of each other (i.e. distance of <8 amino acids between two adjacent sites). After the identification of the hotspot, the total number of times a site within that hotspot was identified and/or regulated across all included studies was summed to determine the identification or regulation rate, respectively. This was repeated for every phosphoprotein within the entire dataset. Subsequently, a ranking of all hotspots based on either identification or regulation rate demonstrates the relative contribution of a hotspot to the phosphoproteomic landscape (Figure 2A).

After analysis, the total dataset of >100,000 phosphosites was compressed into 17,566 phosphorylation hotspots. Of these, more than 85% contained phosphosites originating from 2 or more different datasets (Figure 2B), compared to 35% overlap of individual phosphosites (Figure 1C). Of all hotspots, 8,057 (~45%) contained sites that were significantly regulated in at least 1 dataset, of which >50% contained regulated sites originating from 2 or more datasets (Figure 2B), compared to 30% overlap of individual phosphosites (Figure 1D). When considering only the breast cancer studies, >79,400 phosphosites in total, were compressed into >12,000 phosphorylation hotspots (Figure 2C). Of all identified hotspots, >3,400 (~25%) consisted of at least one regulated phosphosite (Figure 2C). The overlap between the datasets was slightly lower than for the complete dataset due to the lower number of studies combined (Figure 2B), although it was still robustly more than on the level of phosphosites

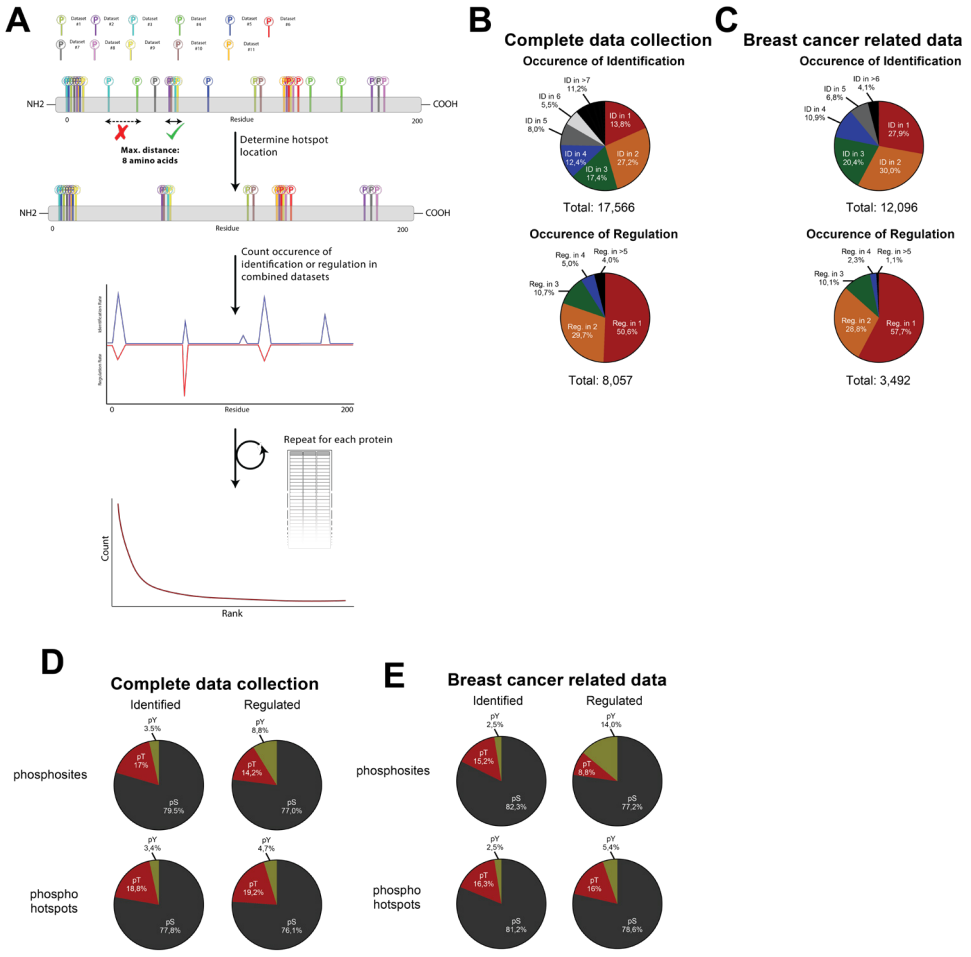


Figure 2 – Phosphorylation hotspots increase overlap but decrease inclusion of tyrosine phosphorylation

A: A schematic representation of the construction of phosphorylation hotspots data is shown. Per protein, the distance between each phosphosite was considered. If the distance was less than 8 amino acids, the site remained within the dataset. Total times of identification or regulation of each phosphosite included in a hotspot was summed, resulting in an identification or regulation rate, respectively. This was repeated for all proteins within the total dataset. Subsequently, each hotspot was ranked based on its identification or regulation rate. **B-C**: For each hotspot, the number of dataset where one or more phosphosites that were retained within that hotspot was identified (top) or regulated (bottom) was counted. For all pie plots, the total of included sites is shown below the plot, and each section is labelled with its percentage. This was done for either all 11 datasets **(B)** or only breast cancer related datasets **(C)**. **D-E**: of either all phosphosites (top) or phosphorylation hotspots (bottom) the percentage of the presence of S-T-Y phosphorylated amino acids was counted. A separation was made between all sites/hotspots (left) or only those that were regulated at least once (right). This was done in either all 11 datasets **(D)** or in breast cancer related studies **(E)**.

(Figure 1F, G).
Due to the merging of phosphosites into phosphorylation hotspots, several sites that did not have a close neighbor were excluded from further analysis. Interestingly, this exclusion was reflected in the distribution of pS/T/Y residues within the data. Both identified phosphosites or hotspots contained 2,5 – 3,5% phosphotyrosine residues (Figure 2D, E). Amongst all regulated phosphosites either 8,8% or 14% (all datasets or breast cancer specific, respectively) of all phosphosites were located on a tyrosine residue, which can be explained by the essential role of tyrosine kinase signaling in cellular biology (Figure 2D, E). Contrastingly, these high percentages changed to 4,7% or 5,4%, respectively, after merging the data into hotspots, demonstrating the disappearance of tyrosine phosphorylation in particular (Figure 2D, E). Indeed, the individual nature and therefore the absence of tyrosine phosphorylation in phosphorylation hotspots has been shown before [31].

Validation of shotgun phosphoproteomic data

A shotgun proteomics experiment results in an extensive collection of data of identified and quantified proteins and PTMs. Subsequent methods for data processing can be described as either the analysis of general trends, with a focus on pathways and processes, or the inspection and discussion of selected proteins of interest, with the focus on the regulation of an individual protein or phosphosite. A list of target proteins or phosphosites is obtained through selective filtering based on set criteria, for example preceding hypotheses, data processing, availability of validation methods or is selected by the researcher him or herself (i.e. statistical relevance, confidence, clustering, availability of antibodies or prior knowledge, either from personal experience or from literature). Needless to say, that when the criteria of this filtering are adjusted, the resulting list of potential follow-up targets will be different. Since it is both impossible as well as unnecessary to follow-up and validate all observations made within a complete study, no harm is done in making a selection. However, consequently, a large part, if not the majority of the data will remain unexplored. Therefore, every published dataset benefits greatly from data reuse, in which, based on different criteria due to a change of vision and circumstances, filtering will be different and alternative findings will emerge. Additionally, if after re-analysis the same proteins or phosphosites are considered to be interesting and worth following up, the confidence in validation research will increase.

Data reuse validates previous results and reveals new insights

Without being influenced by what was discussed in the original papers, several hotspots were selected to illustrate the importance of the use of quantification data. This selection was made based on either the summed

occurrences of identifications in the hotspot or based on the observed difference in hotspot regulation upon a change in the biological context (i.e. all vs breast cancer) (Table 1). Next, this list was cross-referenced with all proteins and phosphosites previously shown and discussed in the original articles. Not surprisingly, ERBB signaling related proteins were often picked for follow-up in the original papers as well as in our analysis. These proteins, including MAPK1/3, EGFR, GAB1, CDK1/2, MAPK14, SHC1, PRCKD and MET, were often regulated due to the direct involvement in pathways related to resistance against HER2/ER inhibition. Fur-

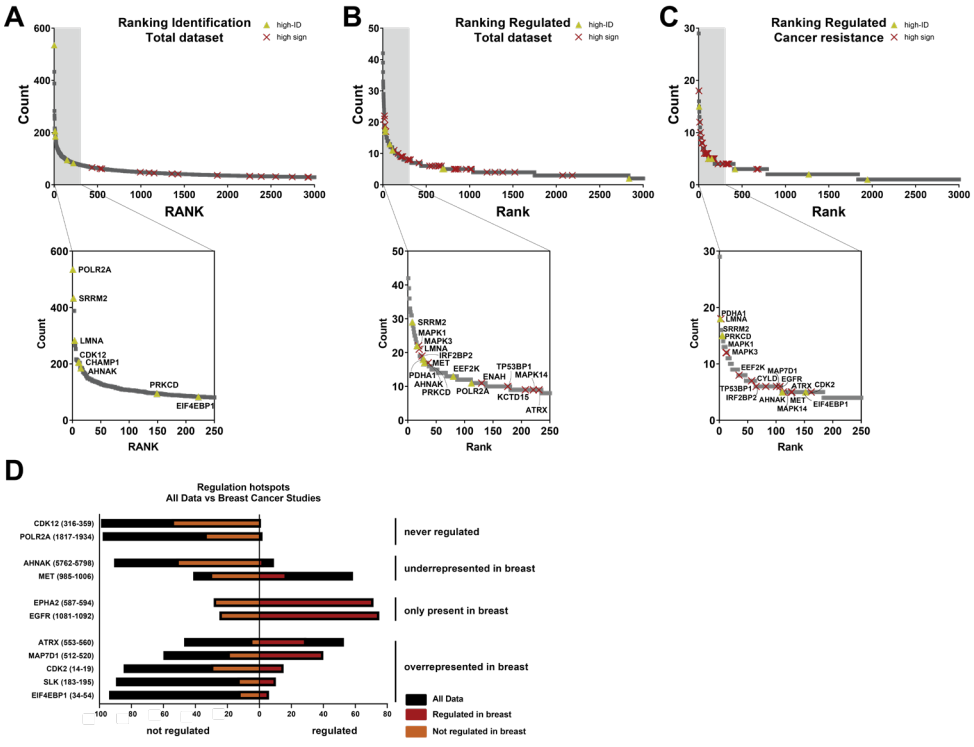


Figure 3 – Biological context influences the phosphorylation hotspot landscape

A-C: Phosphorylation hotspots were ranked (Y-axis) based on total number of occurrences in the complete dataset of identification **(A)** or regulation **(B)** or regulation considering breast cancer specific datasets **(C)**. The first 250 hotspots were shown as a zoom-in for more detail. Interesting hotspots that were amongst the top 250 in **(A)** were indicated with a yellow triangle and their gene name. Hotspot that were found to be interesting and moved up to the top 250 in **(B)** or **(C)** were indicated with a red cross and their gene name. **D**: 11 hotspots were shown in more detail. The black bar represents the total number of identifications in the complete dataset, which is normalized to be 100%. The extent to which this hotspot was regulated is depicted as the horizontal position of the bar; the more to the left, the less this hotspot was regulated. The orange and red bars together represent the portion of identifications originating from breast cancer related studies. Again, the relative horizontal position represents the amount of regulation (orange: not regulated, red: regulated).

thermore, these proteins are well studied in general, resulting in a large body of literature available, reflected in the availability of antibodies and other validation possibilities. Therefore, these proteins are indeed a valid choice when selecting a follow up target of interest. Furthermore, proteins related to the cytoskeleton were often selected to be discussed, due to their high abundance resulting in often high confidence identifications and their known involvement in drug resistance. Several proteins of this class overlapped between the different and our analyses, including AHNAK, ANXA2, CAV1, CTNND1, EPHA2, MAP7D1, PTK2, and PXN.

Additionally, 43 phosphoproteins were selected for further discussion in this study, which were not at all discussed or included in any figure in any of the published studies. Some caught our interest due to the large amount of phosphosites identified but the lack of regulation, including POLR2A, CHAMP and CDK12 (Table 1, Figure 3). Although the absence of regulation explains their absence in the original analyses, the profound presence of these phosphorylation hotspots across a variety of cells and tissues might equally well imply functionality.

Some hotspots were depleted or enriched if only studies with a similar biological model were included in the analysis. For example, EFS, ACP1 and MPP7, which were only identified in the breast cancer related studies, emphasizing their role in this model system. Contrarily, a hotspot that was

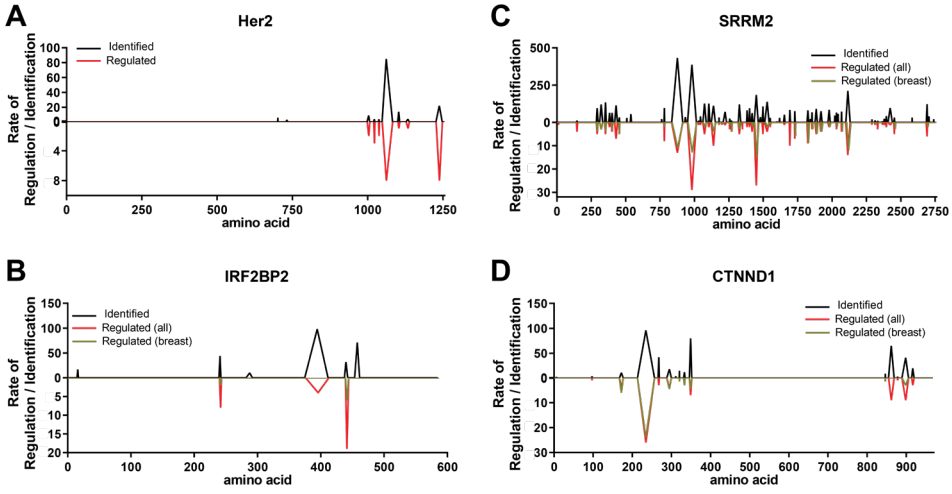


Figure 4 – Comparing identification to regulation rate highlights the most interesting phosphorylation hotspot per individual protein
Four proteins were displayed individually. The X-axis represents the length of each protein. The Y-axis represents either the identification rate (upwards, black line) or regulation rate (downwards, red and brown line). The red line shows the regulation rate of the complete dataset, whereas the brown line is the count of only the breast cancer related studies. Each peak starts and ends at the first and last amino acid within the respective hotspot, respectively. Therefore, the width of the peak indicates the stretch of protein covered with that one hotspot. Shown proteins are Her2 **(A)**, IRF2BP2 **(B)**, SRRM2 **(C)** and CTNND **(D)**.

never identified in these datasets was NDRG1 (aa. 247-255). Interestingly, a second hotspot on the same protein was identified and regulated extensively in all studies (aa. 319-346), indicating possible differentially regulation of this protein in different biological contexts. Another example of this is LMNB2, which was identified in all studies, but solely regulated in non-breast cancer related studies.

FOUR

The use of biologically similar datasets results in phenotype specific phosphorylation hotspot regulation

To provide for a more systematic analysis, all identified hotspots were ranked based on their identification rate within in the total combined dataset (Figure 3A). The most often identified hotspots were indicated with a yellow triangle. These proteins were associated with the cytoskeleton (e.g. LMNA), the spliceosome (e.g. SRRM1), translation (e.g. EIF4EBP1), apoptosis (e.g. PRKCD), cell cycle (e.g. CHAMP1, CDK12, POLR2A), or were extremely large phosphoproteins (e.g. AHNAK).

Next, similar as to for all identified hotspots, all regulated hotspots were ranked, based on their regulation rate within the total combined dataset (Figure 3B). For comparison, the same proteins were annotated in this figure as above. Some proteins that appeared in the top 250 due to altered ranking were annotated with a red cross. This approach showed that when the regulation rate was taken into account, the phosphorylation hotspot landscape changed tremendously. If one focusses on occurrence of identified sites, the most appealing findings will be the high abundant, and easy to find phosphosites. In order to corroborate whether there would be a correlation between intensity and extent of regulation we compared the intensities of regulated sites to the intensities of all identified sites in several datasets. Although the mode of the intensity is the same for both identified and regulated sites, there is an underrepresentation of low abundant phosphosites in the regulated data, which is most likely due to the fact that a certain threshold value is needed for quantification, and low intensity sites increase the chance of having missing values which impairs quantification (Supplementary Figure S1).

When basing the ranking on the occurrence of regulation, we hypothesize that depending on their position within the list, hotspots will describe important biological activities within proteins. Although some proteins remain amongst the highest ranking (e.g. LMNA, SRRM2 or AHNAK), others shift up (e.g. MAPK1/3) or down (e.g. POLR2A), demonstrating their (un) importance in the biological control of the cells. Phosphoproteins that are often identified but less regulated most likely play a role in house-keeping pathways or do not play a role in the specific biological model described in the used studies.

In order to further elucidate this, a well-considered choice of input datasets could enable the discovery of cell or disease specific regulations. Therefore, all phosphorylation hotspots containing regulated phospho-

sites originating from the breast cancer specific studies were ranked as well (Figure 3C). Again, ranking of several hotspots changed in response to the alteration of the analysis. The specific hotspot from aa 289-301 in PDHA1 was the highest ranked hotspot, based on extent of regulation. This hotspot on PDHA1 is responsible for inactivation of this metabolic enzyme, restraining the pyruvate flux into mitochondrial oxidative metabolism, contributing to a switch from glucose oxidation to aerobic glycolysis, also known as the Warburg effect, which in turn contributes to tumor growth, progression and therapy resistance [33]. PRKCD hotspot aa 259-313 was another highly ranked hotspot in the breast cancer data. In correlation with this observation, PRKCD has very recently been identified to play a role in a common adaptive resistance mechanism in HER2 positive breast cancer cells, together with Focal adhesion kinase and Ephrin receptors, for which also several hotspots were regulated in the breast cancer data [34].

When comparing the regulated hotspots in datasets describing either all datasets or breast cancer specific datasets, phenotype specific hotspot regulation can be detected (Figure 3D). Figure 3D highlights several protein hotspots that were never regulated; hotspots for which regulation cannot be found in the breast cancer data, but regulation was observed in the other datasets; hotspots that were only identified and regulated in breast cancer data; hotspots that were more highly regulated in breast cancer data. As shown, selected hotspots in POLR2A and CDK12 were frequently identified in both breast cancer data as well as the cumulative dataset. However, regulation of these hotspots was never observed. In contrast, hotspots in EGFR and EPHA2 were highly regulated in the breast cancer datasets, whereas they were not even identified in the other datasets. Obviously, this observation can be rationalized by the important role these two receptor tyrosine kinases play in breast cancer signaling. In addition, a number of hotspots were overrepresented in the breast cancer data, meaning that hotspots were identified and regulated in all data, but the extent of regulation was proportionally higher in breast cancer related studies. These sites include hotspots on CDK2 and MAP7D1. Finally, some hotspots were underrepresented in the breast cancer data, for example hotspots on AHNAK and MET. Especially the underrepresentation of the phospho hotspot on MET was interesting, considering the suggested role for MET in breast cancer and resistance. However, when looking at the functional implication of phosphorylation on this hotspot, this observation can be rationalized as it is an enzymatic activity inhibiting site, whereas activity is expected to be high in the breast cancer models used [35].

Analysis of hotspots per individual protein reveals most interesting phosphosites

Table 1 shows the 45 most regulated phospho hotspots that were found in this study. The PhosphoSitePlus database graphically indicates per protein

how often a phosphosite has been documented in literature. Within Table 1, it is indicated if the phosphorylation hotspot includes the most often identified site of this protein on PhopshoSitePlus and if a biological function is known. Several proteins showed most of their regulation on sites which are already well studied and well known in literature (e.g. MAPK1-185/187 and MAPK3-p202/204), validating the power of this analysis (Table 1). However, a significant portion of regulated phosphoproteins were found to have very little or no literature available about the biological function of documented site (e.g. CTNND1-p904).

The above described general hotspot ranking of the entire phosphoproteome provides an indication of which proteins could be a potentially interesting follow-up target. However, the analysis of both the most identified as well as the most regulated hotspot *per protein* is an indication of which sites specifically could be biologically interesting. An interesting protein in this respect is HER2, a major player in all included breast cancer studies. Interestingly, in addition to the most identified phosphorylation hotspot (16 phosphosites, spanning aa 1048-1083), a second hotspots attracts the attention by being the most regulated hotspot (4 phosphosites, spanning aa 1227-1248) (Figure 4A). The hotspot spanning aa 1227-1248 has been implicated in regulation the activity of this receptor tyrosine kinase [34]. The aa 1048-1083 hotspot has previously been reported to become phosphorylated in breast cancer cells in response to lapatinib treatment and this phosphorylation regulates HER2 tyrosine kinase activity [36].

IRF2BP2 is another protein with interesting hotspot characteristics. As can be seen in Figure 4B, a hotspot region between aa 376-413 was frequently observed in all datasets, however the site was not often found regulated and never regulated in the breast cancer data. In contrast, the hotspot region aa 439-444 was extensively regulated in both datasets, although less frequently observed. Despite this apparent regulation, no function for IRF2BP2 phosphorylation is known, and even more striking, none of the sites within this hotspot are reported on PhosphoSitePlus. The large RNA splicing protein SRRM2 was highly phosphorylated all over its amino acid sequence. However, much less of the hotspot regions detected in this protein were actually regulated (Figure 4C). Even more striking for none of the identified sites on SRRM2 functional implications are known. Finally, CTNND1 was highly regulated on a hotspot region between aa 213-257, which is known to affect intracellular localization of the protein (Figure 4D). However, for two other extensively regulated hotspot regions, no functional implications are known.

As exemplified above, the reuse approach again clearly illustrates that more information is retained in quantitative data compared to only identification data. In addition, the repetitive observation of regulation across a diverse set of studies highlights the importance of both known and unknown phosphosites. This approach shows that by scrutinizing combined quantitative studies, interesting and novel follow-up targets can be iden-

tified, and already well-known phosphoproteins can be validated for their role in this particular biological setting.

Discussion

A major objective in the field of mass spectrometry based proteomics remains the pursuit to increase the amount of data per experiment and coverage of the measured (phospho)proteome. This has led to an extensive amplification of publicly available proteome data and concomitant biological relevant discoveries. Apart from these successes, a large part of this data remains unexplored. Therefore, following others in parallel omics fields [6–8], this study explored the possibilities of reusing publicly available, quantitative phosphoproteomic datasets. By re-analyzing a variety of phosphorylation data new observation could be made. Observations that were missed or considered insignificant in the original publication became suddenly visible when combined with findings from other studies, resulting in increased understanding of the studied model and a decrease in data loss.

Unfortunately, data reuse is not yet common in the field of proteomics. As a major consequence, available data is often not presented in a usable format or is incomplete. Indeed, as already expressed by others [23], during the collection of suitable datasets for this study, several datasets were found to be interesting, however could not be included. Reasons for this were, amongst others, lack of proper experimental design (i.e. no biological replicates [37–41]), no documentation of processed results at all [38], or documentation of those sites only found interesting by the authors [40,42–44]. Most striking was the fact that less than half of the considered studies had added their RAW data to a public depository, inhibiting reanalysis and preventing further quality control [45,46].

Due to major technical advances in MS based phosphoproteomics, it is now possible to measure tens of thousands phosphosites within a single run, a number that will only keep increasing [5,47]. At the same time, this increase in size of the phosphoproteome raises the question of its functionality, which has led to the hypothesis that a certain level of redundancy exists within the phosphoproteome. This proposition subdues the functionality of an individual site but rather emphasizes the functionality of either bulk phosphorylation [30–32] or the presence of a small number of representing phosphosites, beholding information on the majority of the phosphoproteome [48]. Either way, the importance of phosphosite functionality over identification is stated repeatedly within these propositions. An effective and high-throughput manner to predict or even determine phosphosite functionality is to assess its reoccurred observation throughout different proteomic studies. In this study, combined data from several phosphoproteomic datasets were used for this purpose. Unfortunately, low reproducibility across studies originating from different labora-

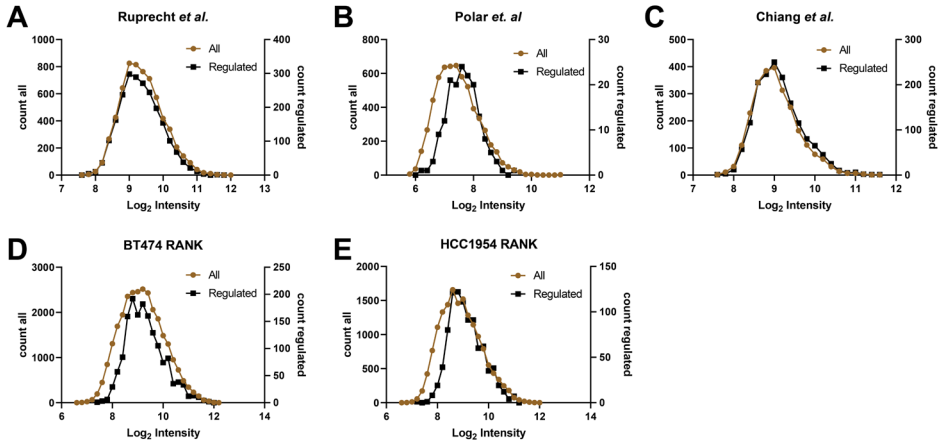
tories restrained reoccurred observation of phosphosites. Not only the use of different cell lines or different protocols caused this lack of overlap, also several technical obstacles hampered the ability to compare datasets. For example, phosphopeptides which cover the same phosphosites can have different masses or charges in different experiments, due to missed cleavages or differences in multiplicity. Consequently, the same site appears as a separate, non-overlapping identification if two data points from two distinct experiments and database searches are combined. Furthermore, in the case of peptides containing several amino acids that can be phosphorylated, the exact localization of the phospho-group can vary between datasets. This can for example be due to differences in site specific fragments within MS2 spectra, resulting in a single phosphosite appearing as two adjacent sites in two different experiments. In order to solve the problem in the lack of overlap and to improve data processing to identify reoccurring phosphosites, the more than 100,000 initial phosphosites included in the analysis were transformed into ~20,000 phosphorylation hotspots. By considering multiple phosphosites as one, these biological and technical artefacts could be overcome. Although absolute quantitative data is lost in this approach (i.e. if a phosphosite increased or decreased under certain conditions) the use of phosphorylation hotspots revealed robust identification of highly interesting phosphosites amongst a large panel of proteomic studies.

An important consideration while merging data from different studies is the type of data used. By varying the nature of the included studies, the resulting conclusions will change accordingly, depending on type of data (i.e. identification or quantification) or biological background (i.e. varied or overlapping). By using phosphoproteomic data from public databases, for example PhosphoSitePlus [21], the mere presence of phosphorylation can be mapped [30]. Following, reoccurring identification [31], alignment and conservation across evolution [32] can be used to predict functional relevance, although not specified to a biological situation of interest. In contrast, by using quantitative studies and therefore taking experimental and phenotypic specific data into consideration, a prediction can be made of the functional relevance of phosphosites in a specific biological situation. As shown in this study, if all used datasets describe a highly similar model system, an enrichment of model specific regulations can be identified. If on the other hand, each dataset describes a unique biological system, the model-specific regulations will not overlap and therefore dilute. Although this complementarity might decrease overlap across datasets, regulations that remain to overlap and are found in a great variety of studies will give insight in important universal events.

In most studies the final part of data processing is the selection of several proteins or phosphosites of interest for follow up. The most used criterion during this process is significant regulation of a phosphosite between bi-

ological relevant conditions. This is an obvious approach, since changing phosphosites can be used to differentiate phenotypes and to identify potential therapeutic windows. However, in contrast, stable, never changing, phosphosites could be of equal interest. Their reoccurred identification but absence of regulation across a large panel of studies could equally imply functionality. Unfortunately, whereas regulation of a phosphosite within an individual study is able to attract interest leading to follow-up research, the absence of regulation of a site will most likely not lead to any further research. Such a phosphosite will only stand out if multiple datasets are combined revealing the level of its stability. Therefore, the added value of data reuse and combining different datasets is not only to add to the existing knowledge of included datasets by discussing additional observations, but also to arrive at observations that would not be possible to be noted when analyzing all studies individually.

In conclusion, in this study we showed that it is possible to combine, and reuse published quantitative phosphoproteomic data. The re-analysis of data adds new knowledge on the existing documented observations. Furthermore, by combining data, observations were made that would have gone unnoticed if all studies were analyzed solely individually. In this study we focused on protein phosphorylation which is an important but, functionally, still understudied biological area. Therefore, more attention should be given to high-throughput identification of phosphosite functionality. It should be noted that it is important to not only expand the phosphoproteome itself but also investigate its implication in order to gain a better understanding of cellular biology and aid future therapeutic developments. This can be achieved by data reuse, as shown in this study. The amount of publicly available data is immense and rich source of unexplored knowledge. If more researchers would invest time in reusing existing data, its value will increase. Although expansion of the (phospho) proteome is certainly necessary, increasing the depth of understanding is a must, for high quality data not to go to waste.



Supplementary Figure S1 – Intensities of all vs regulated phosphosites per dataset
For five separate datasets the distribution of Log2 intensities were plotted for either all identified phosphosites (brown line) or only regulated phosphosites (black line). **A:** Ruprecht *et al*; **B:** Polar *et al.*; **C:** Chiang *et al.*; **D:** BT474 RANK; **E**: HCC1954 RANK.

Gene name	Sites included in hotspot	In original paper	Highest ID in PSP	known function
	S1077, S1078, S1083			
ERBB2	T1227, Y1233, T1240, T1242, Y1248	TRUE	yes	activity
GAB1	S401, Y406	TRUE	2nd	no function
GAB1	Y657, Y659	TRUE	yes	SHP2 activation
IRF2BP2	S439, S441, S444	TRUE	--	--
LMNA	S390, S392, T394, S395, S398, S403, S404, S406, S407, T409, S414, T416, S423, T424, S426, S428, S429, S431	FALSE	2nd	activity
LMNB2	S419, S420, S421, S422, S424, S426	FALSE	no	no function
MAP7D1	S512, S516, S520	TRUE	--	--
MAPK1	T181, T185, Y187, T190	TRUE	yes	activity
MAPK14	T180, Y182	TRUE	yes	activity
MAPK3	T202, Y204, T207	TRUE	yes	increased activity
MET	S985, S988, S990, T992, T993, T995, S997, S1000, Y1003, S1006	TRUE	no	degradation/stabilisation
MPP7	S409, S412, Y417	FALSE	yes	no function
NDRG1	S317, S319, S326, T328, S330, S332, S333, T335, S336, T340, S342, T346	FALSE	yes	carcinogenesis
NDRG1	T247, S249, S251, S252, S255	FALSE	--	--
PDHA1	Y289, S293, S295, S300, Y301	TRUE	yes	activity inhibited
POLR2A	T1817, S1821, S1828, T1835, T1840, S1843, S1845, Y1846, S1847, S1849, S1850, Y1853, T1854, T1856, S1857, Y1860, S1861, T1863, S1864, Y1867, S1868, T1870, S1871, Y1874, S1875, T1877, S1878, T1880, Y1881, S1882, T1884, T1885, S1889, T1891, S1892, T1894, S1896, T1898, S1899, T1903, T1905, S1906, S1910, T1912, S1913, T1915, S1917, T1919, S1920, S1924, T1926, S1927, T1929, Y1930, S1931, T1933, S1934	FALSE	no	no function
PRKCD	S331, T333, Y334	TRUE	2nd	apoptosis
PRKCD	T295, S299, S302, S304, S306, S307, Y313	TRUE	yes	increased activity
PTK2	S568, Y570, Y576, Y577	TRUE	yes	increased activity
PXN	Y118, S119, S126, S130, T132, S137	TRUE	yes	localisation
SHC1	S344, Y349, Y350	TRUE	2nd	inhibited apoptosis
SLK	T183, S189, T193, Y195	FALSE	yes	activity
SRRM2	S834, S836, S837, S838, S839, S846, T848, S854, T856, S857, S864, T866, S871, S875, S876, S883, T885, S887, S890, S892, S894, S895, S901, S902, T903, S908, S910, S912, S913, S914	FALSE	no	no function

Table 1 – Selected hotspots
List of selected and in the text discussed hotspots. Column: Gene names; All included sites (amino acid, residue); Mentioned, discussed or displayed in one of the original studies (True or False); Indicated if one of the included sites is the most often documented site on PhosphoSitePlus (PSP) (yes, no or 2nd); Known function, as documented on PSP.

Gene name	Sites included in hotspot	In original paper	Highest ID in PSP	known function
ACP1	Y132, Y133	FALSE	yes	adhesion; cytoskeleton rearrangements
AHNAK	S5762, S5763, S5769, S5773, S5780, S5782, S5784, S5790, S5793, T5794, T5796, T5798	TRUE	no	no function
ANXA2	S234, Y235, Y238	TRUE	2nd	no function
ANXA2	S184, Y188	TRUE	no	no function
ANXA2	S12, S18, T19, T21, S22, Y24, S26, Y30, T31	TRUE	yes	proliferation; STAT3 induction
ATRX	T553, S556, S558, S560	FALSE	--	--
CAV1	Y6, S9, Y14, T15	TRUE	yes	activity; motility
CDK1	T14, Y15, Y19	TRUE	yes	activity altered
CDK12	S316, S318, Y319, S320, S323, S325, Y327, S332, S333, S334, S338, S341, S343, S345, S349, S355, S357, S359	FALSE	no	no function
DK2	T14, Y15, Y19	TRUE	yes	cell cycle regulation
CHAMP1	S427, S432, S436, S443, S445, S452, T458, S459, S462, S468, S471, S472, S476	FALSE	2nd	no function
CTNND1	Y213, S214, S215, Y217, S219, Y221, S225, Y228, S230, S232, S234, Y241, S244, S245, S248, S251, S252, T254, T256, Y257	TRUE	2nd	localisation
CTNND1	S889, S893, S895, Y898, S899, T900, Y904, S905, T906	TRUE	yes	no function
EFS	S317, S319, S323	FALSE	no	no function
EGFR	S1064, Y1069, S1071	TRUE	no	association and internalisation
EGFR	S1081, T1085, Y1092	TRUE	2nd	increased activity
EGFR	S1166, Y1172	TRUE	yes	increased activity
EIF4EBP1	Y34, S35, T36, T37, T41, S44, T45, T46, T50, Y54	TRUE	yes	transcription; translation
EPHA2	T587, Y588, Y594	TRUE	2nd	motility
EPHA2	S570, Y575, S579	TRUE	no	no function
EPHA2	T771, Y772, S775	TRUE	yes	no function
ERBB2	S1048, S1049, S1050, S1051, T1052, S1053, S1054, S1058, S1063, S1066, S1068, S1070, S1073,	TRUE	no	activity

Acknowledgements

We would like to acknowledge all authors of which we have used data from. We highly appreciate the fact that they have shared their data in a format suitable for re-analysis, providing the opportunity for new discoveries in addition to their own.
Included authors were: Adrián Sanz Moreno and Eva Gonzalez Suarez, Yooijn Chang *et al.*, Cheng-Kang Chiang *et al.,* Kristina B. Emdal *et al.*, Kuan-lin Huang *et al.*, Elena Pinazza *et al.,* Robert Parker *et al*., Yesim Gökmen-Polar *et al*., Catrin Roolf *et al.*, Benjamin Ruprecht *et al*., Simon Vyse *et al*.
References: [26,28,29,49–55].

Materials and Methods

Used Datasets

RANK dataset, in house generated 2018
The RANK data (unpublished) describes two Her2+ cell lines, BT474 and HCC1954, with either the overexpression or knockdown of RANK. RANK is a receptor of the TNF superfamily, and works upstream of several growth stimulating pathways [56]. Despite that RANK has mainly been studied in the context of bone remodeling, the receptor has been linked as well to other processes like immunity and oncogenic transformation [57,58] It has been shown that RANK is able to heterodimerize with Her2, and additionally play a role in Her2 targeted drug resistance in breast carcinomas [59]. Cellular assays showed a significant de- or increase of RANK expression, however only after RANK knockdown a visible phenotypic changed was observed in the form of increased Her2 inhibition sensitivity. Full proteome and phosphoproteome analysis were done to find additional regulations caused by RANK modulation. On a phosphoproteome level, processes associated with adhesion and cellular polarity was regulated.

Ruprecht et al, 2017
Axl is a tyrosine kinase receptor which can be activated by auto phosphorylation, with the result of the downstream activation of Src, Grb2 and MAPK signaling [60]. Aberrant Axl signaling has been linked to increased carcinogenesis in several tissue types [61], including lapatinib resistance in breast cancer [62]. The study of Ruprecht *et al* [28] describes a phosphoproteomic dataset of BT474 breast cancer cells with and without Axl overexpression and subsequent lapatinib resistance. Short term treatment of sensitive cells with lapatinib showed downregulation of ERBB pathway related phosphosites and further downstream targets, including mTOR, JUN, MYC and ERK. Analysis of resistant cells showed both recovery of sites changed by short term treatment, as well as newly regulated sites and proteins, including the reactivation of mTOR signaling and the continuously inhibition of the ERBB-MAPK axis. Furthermore, the authors documented and validated changes in the activity of the spliceosome and the glycolysis.

Chang et al, 2018
The study of Chang *et al* showed a phosphoproteome analysis of lapatinib resistant SKBR3 breast cancer cells [26]. The main documented observations were the regulation of adhesion and cytoskeleton related processes, as annotated by KEGG pathways and STRING-db. Furthermore, PAK2, a kinase which plays a role in oncogenic pathways such as proliferation and invasion, was shown to be increasingly phosphorylated on its activating site (pS141), therefore further validated as a therapeutically target.

Polar et al, 2018
Deleted in Liver Cancer-1 (DLC1) is RhoGAP family member, associated with tumor suppressor activity in ER+ breast cancer. The phosphoproteome of cells with low and high expression levels of DLC1 were analyzed to further determine the role of DLC1 and identify a therapeutic target in cancer types showing low levels of DLC1 [29]. The authors showed that

RhoGEF ECT2 phosphorylation is decreased in the presence of DLC1. They hypothesized that low levels of RhoGAP activity and high levels of RhoGEF activity, due to low DLC1 and high ECT2 activity, respectively, can cause increased oncogenic RhoA activity.

Huang et al, 2017
In order to show resistance mechanisms in a more in vivo situation opposed of the more standard in vitro cultures, patient derived xenografts (PDXs) were produced covering several subtypes of breast cancer and both primary and metastatic tumors [49]. For this study, additional data analysis was done by ANOVA testing, in order to demonstrate differentially regulated phosphosites amongst the produced xenografts.

Emdal et al, 2017
In the study of Embal *et al* [50] the authors employed the use of preclinical tumor xenograft models in order to study acquired resistance against the small molecule osimertinib and the antibody JNJ-61186372 in advanced stage lung cancer. A phosphotyrosine IP approach was used to profile changes upon resistance development and treatment. Between the xenografts a strong correlation existed, although additionally clear tumor specific signatures could be observed. The main conclusion and subject of validation was an interconnection between EGFR and Src family kinases in the outgrowth of resistant tumors.

Chiang et al, 2017
The phosphoproteomic dataset described in the article of Chiang *et al* [51], describes the process of circadian phagocytosis of photoreceptor outer segments (POS) on photoreceptor cells by the retinal pigmented epithelium (RPE). This process is known to be regulated by phosphorylation, although the exact mechanisms are not fully known. A time series of POS exposure was used to map both proteome and phosphoproteome dynamics over time in the human-derived RPE like cell line ARP-19. An ANOVA test was used for the determination of significantly altered phosphosites. The main regulated processes, as identified by GO term enrichment, were associated with cytoskeleton rearrangements and cell-cell junctions. Additional mouse experiments were performed to validate the results found in the cell line based dataset. The authors showed and validated the involvement of PI3K/Akt, MAPK and mTOR pathway in the process of phagocytosis.

Pinazza et al
The study of Pinazza *et al.* [52] describes the application of a previous developed method of peptides based on isoelectric point in a quantitative phosphoproteomic context. They used HeLa cells, including biological conditions of pervanadate treatment or mitotic arrest. Besides the testing of their protocol, they applied machine learning to predict kinase association based on their phosphorylation dataset in order to determine phosphosite functionality.

Parker et al
In the study of Parker *et al.* [53] a model of acquired drug resistance in melanoma has been studied. Cells were made resistant against the BRAF inhibitor PLX-4720 after which phosphoproteomics was used to show reactivation of alternative pathways, including MAPK pathway. Several regulators of the cytoskeleton were found to be significantly altered upon development of drug resistance.

Roolf et al
In the study of Roolf *et al* [54], a phosphoproteomic approach was used to discover alternative modes of actions of acute myeloid leukemia cells with or without FMS like tyrosine kinase 3 (FLT3) mutation, against the multikinase inhibitor Sorafenib. Three models of different FLT3 mutations were used, treated with 2 drug concentrations. The most downregulated pathway were either mTOR or MAPK pathways, depending on mutational status of the FLT3.

Vyse et al
In the study of Vyze *et al* [55], a phosphoproteomic approach was used to study resistance against pazopanib or dasatinib in a model of rhabdoid tumor cells. Possible identified resistance mechanisms were hypothesized to be related to cytoskeleton rearrangements and IGF-1R signaling. Additionally, several transcription factors were shown to be regulated, providing new insights in this yet little studied model and possible novel follow-up targets.

The construction of phospho hotspots and subsequent rankings
From all datasets, the searched and completely processed data as performed in the original publications were used. In addition, the determination of significant regulation across different conditions was performed as described by the methods used by the original authors. No

data considering level of quantification (i.e. up or down, or ratio) was included, since this is increasingly unreliable to compare between experiments from different origin. All data was merged into a single dataset comprised of all phosphosites identified across all studies, with the accompanied data of in which study a site was identified and, if applicable, regulated.
In order to construct phosphorylation hotspots all sites of a single protein were examined and distances between identified sites were determined. Adjacent sites with a distance of less than 8 amino acids between them were considered to be in the same hotspot. Subsequently, the number of identifications as well as the number of times regulation was observed, was summed for all phosphosites included per hotspot. Additionally, the identification of a phosphosite per dataset was multiplied by the number of conditions that were included. In this way, if a phosphosite would be regulated in all included conditions, the summed rate of regulation would be the same as the summed count of identification. Identification as well as regulation rates were calculated for the complete set of 11 datasets as well as the 6 breast cancer related datasets separately.

References

1. Batth TS, Francavilla C, Olsen J V.: **Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics**. *J Proteome Res* 2014, **13**:6176–6186.
2. Giansanti P, Aye TT, van den Toorn H, Peng M, van Breukelen B, Heck AJR: **An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas**. *Cell Rep* 2015, **11**:1834–1843.
3. Abe Y, Nagano M, Tada A, Adachi J, Tomonaga T: **Deep Phosphotyrosine Proteomics by Optimization of Phosphotyrosine Enrichment and MS/MS Parameters**. *J Proteome Res* 2017, **16**:1077–1086.
4. Ferries S, Perkins S, Brownridge PJ, Campbell A, Eyers PA, Jones AR, Eyers CE: **Evaluation of Parameters for Confident Phosphorylation Site Localization Using an Orbitrap Fusion Tribrid Mass Spectrometer**. *J Proteome Res* 2017, **16**:3448–3459.
5. Kelstrup CD, Bekker-Jensen DB, Arrey TN, Hogrebe A, Harder A, Olsen J V: **Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics**. *J Proteome Res* 2018, **17**:727–738.
6. Sweeney TE, Haynes WA, Vallania F, Ioannidis JP, Khatri P: **Methods to increase reproducibility in differential gene expression via meta-analysis**. *Nucleic Acids Res* 2017, **45**:1–14.
7. Sparks R, Lau WW, Tsang JS: **Expanding the Immunology Toolbox: Embracing Public-Data Reuse and Crowdsourcing**. *Immunity* 2016, **45**:1191–1204.
8. Rung J, Brazma A: **Reuse of public genome-wide gene expression data**. *Nat Rev Genet* 2013, **14**:1–11.
9. Griss J, Perez-Riverol Y, Hermjakob H, Vizcaíno JA: **Identifying novel biomarkers through data mining-A realistic scenario?** *Proteomics - Clin Appl* 2015, **9**:437–443.
10. Martens L, Vizcaíno JA: **A Golden Age for Working with Public Proteomics Data**. *Trends Biochem Sci* 2017, **42**:333–341.
11. Deutsch EW, Csordas A, Sun Z, Jarnuczak A, Perez-Riverol Y, Ternent T, Campbell DS, Bernal-Llinares M, Okuda S, Kawano S, et al.: **The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition**. *Nucleic Acids Res* 2017, **45**:D1100–D1106.
12. Lawrence RT, Searle BC, Llovet A, Villén J: **Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry**. *Nat Methods* 2016, **13**:431–434.
13. Vandermarliere E, Mueller M, Martens L: **Getting intimate with trypsin, the leading protease in proteomics**. *mass Spectrom Rev* 2013, **32**:453–465.
14. Barsnes H, Eidhammer I, Martens L: **A global analysis of peptide fragmentation variability**. *Proteomics* 2011, **11**:1181–1188.
15. Lund-Johansen F, De La Rosa Carrillo D, Mehta A, Sikorski K, Inngjerdingen M, Kalina T, Røysland K, De Souza GA, Bradbury ARM, Lecrevisse Q, et al.: **MetaMass, a tool for meta-analysis of subcellular proteomics data**. *Nat Methods* 2016, **13**:837–840.
16. Rudolph JD, Graauw M De, Water B Van De, Geiger T: **Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Report Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Networks**. *Cell Syst* 2016, **3**:585–593.
17. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, et al.: **A draft map of the human proteome**. *Nature* 2014, **509**:575–581.
18. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, et al.: **Mass-spectrometry-based draft of the human proteome**. *mass Spectrom Rev* 2013, **32**:5453–465.
19. van Wijk KJ, Friso G, Walther D, Schulze WX: **Meta-Analysis of Arabidopsis thaliana Phospho-Proteomics Data Reveals Compartmentalization of Phosphorylation Motifs**. *Plant Cell* 2014, **26**:2367–2389.
20. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley M, Bonilla C, Britto R, et al.: **UniProt: The universal protein knowledgebase**. *Nucleic Acids Res* 2017,

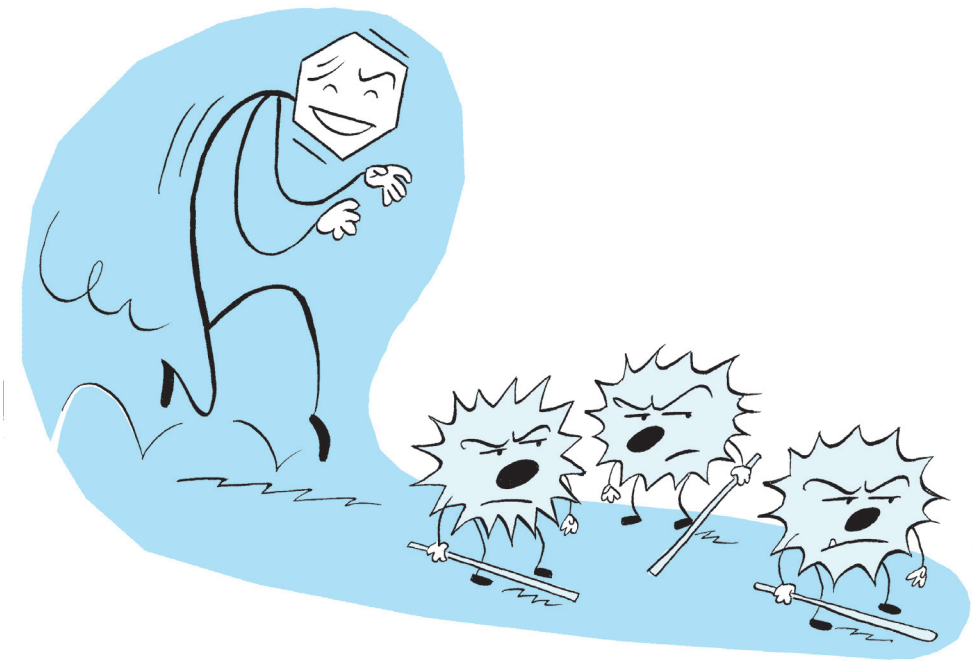
45:D158–D169.
21. Hornbeck P V., Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E: **PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations**. *Nucleic Acids Res* 2015, **43**:D512–D520.
22. Manousopoulou A, Hayden A, Mellone M, Garay-Baquero DJ, White CH, Noble F, Lopez M, Thomas GJ, Underwood TJ, Garbis SD: **Quantitative proteomic profiling of primary cancer-associated fibroblasts in oesophageal adenocarcinoma**. *Br J Cancer* 2018, **118**:1200–1207.
23. Srisawat K, Shepherd S, Lisboa P, Burniston J: **A Systematic Review and Meta-Analysis of Proteomics Literature on the Response of Human Skeletal Muscle to Obesity/Type 2 Diabetes Mellitus (T2DM) Versus Exercise Training**. *Proteomes* 2017, **5**:30.
24. Taylor CF, Paton NW, Lilley KS, Binz P-A, Julian RK, Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW, et al.: **The minimum information about a proteomics experiment (MIAPE)**. *Nat Biotechnol* 2007, **25**:887–893.
25. Humphrey SJ, James DE, Mann M: **Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation.** *Trends Endocrinol Metab* 2015, **26**:676–687.
26. Chang Y, Park KH, Lee JE, Han KC: **Phosphoproteomic analysis reveals PAK2 as a therapeutic target for lapatinib resistance in HER2-positive breast cancer cells**. *Biochem Biophys Res Commun* 2018, **505**:187–193.
27. Wu X, Zahari MS, Renuse S, Nirujogi RS, Kim M-S, Manda SS, Stearns V, Gabrielson E, Sukumar S, Pandey A: **Phosphoproteomic Analysis Identifies Focal Adhesion Kinase 2 (FAK2) as a Potential Therapeutic Target for Tamoxifen Resistance in Breast Cancer**. *Mol Cell Proteomics* 2015, **14**:2887–2900.
28. Ruprecht B, Zaal EA, Zecha J, Wu W, Berkers CR, Kuster B, Lemeer S: **Lapatinib resistance in breast cancer cells is accompanied by phosphorylation-mediated reprogramming of glycolysis**. *Cancer Res* 2017, **77**:1842–1853.
29. Gokmen-Polar Y, True JD, Vieth E, Gu Y, Gu X, Qi GD, Mosley AL, Badve SS: **Quantitative phosphoproteomic analysis identifies novel functional pathways of tumor suppressor DLC1 in estrogen receptor positive breast cancer**. *PLoS One* 2018, **13**:1–21.
30. Regev S, Michal L: **Cooperativity within proximal phosphorylation sites is revealed from large-scale proteomics data**. *Biol Direct* 2010, **5**:6.
31. Christian J-O, Braginets R, Schulze WX, Walther D: **Characterization and Prediction of Protein Phosphorylation Hotspots in Arabidopsis thaliana**. *Front Plant Sci* 2012, **3**:1–14.
32. Torres MP, Dewhurst H, Sundararaman N: **Proteome-wide Structural Analysis of PTM Hotspots Reveals Regulatory Elements Predicted to Impact Biological Function and Disease**. *Mol Cell Proteomics* 2016, **15**:3513–3528.
33. Jin Y, Cai Q, Shenoy AK, Lim S, Zhang Y, Charles S, Tarrash M, Fu X, Kamarajugadda S, Trevino JG, et al.: **Src drives the Warburg effect and therapy resistance by inactivating pyruvate dehydrogenase through tyrosine-289 phosphorylation.** *Oncotarget* 2016, **7**:25113–24.
34. Schwill M, Tamaskovic R, Gajadhar AS, Kast F, White FM, Plückthun A: **Systemic analysis of tyrosine kinase signaling reveals a common adaptive response program in a HER2-positive breast cancer.** *Sci Signal* 2019, **12**:eaau2875.
35. Gandino L, Longati P, Medico E, Prat M, Comoglio PM: **Phosphorylation of serine 985 negatively regulates the hepatocyte growth factor receptor kinase.** *J Biol Chem* 1994, **269**:1815–20.
36. Imami K, Sugiyama N, Imamura H, Wakabayashi M, Tomita M, Taniguchi M, Ueno T, Toi M, Ishihama Y: **Temporal profiling of lapatinib-suppressed phosphorylation signals in EGFR/HER2 pathways.** *Mol Cell Proteomics* 2012, **11**:1741–57.
37. Chen H, Pimienta G, Gu Y, Sun X, Hu J, Kim MS, Chaerkady R, Gucek M, Cole RN, Sukumar S, et al.: **Proteomic characterization of Her2/neu-overexpressing breast cancer cells**. *Proteomics* 2010, **10**:3800–3810.
38. Imami K, Sugiyama N, Tomita M, Ishihama Y: **Quantitative proteome and phosphoproteome analyses of cultured cells based on SILAC labeling without requirement of serum dialysis**. *Mol Biosyst* 2010, **6**:594–602.
39. Nunes J, Zhang H, Angelopoulos N, Chhetri J, Osipo C, Grothey A, Stebbing J, Giamas G: **ATG9A loss confers resistance to trastuzumab via c-Cbl mediated Her2 degradation**. *Oncotarget* 2016, **7**:27599–27612.
40. Rexer BN, Ham AJL, Rinehart C, Hill S, De Matos Granja-Ingram N, González-Angulo AM, Mills GB, Dave B, Chang JC, Liebler DC, et al.: **Phosphoproteomic mass spectrometry profiling links Src family kinases to escape from HER2 tyrosine kinase inhibition**. *Oncogene* 2011, **30**:4163–4174.
41. Li W, Wang H, Yang Y, Zhao T, Zhang Z, Tian Y, Shi Z, Peng X, Li F, Feng Y, et al.: **Integrative Analysis of Proteome and Ubiquitylome Reveals Unique Features of Lysosomal and Endocytic Pathways in Gefitinib-Resistant Non-Small Cell Lung Cancer Cells**. *Proteomics* 2018, **18**:1–16.
42. Yoshida T, Zhang G, Smith MA, Lopez AS, Bai Y, Li J, Fang B, Koomen J, Rawal B, Fisher KJ, et al.: **Tyrosine Phosphoproteomics Identifies Both Codrivers and Cotargeting Strategies for T790M-Related EGFR-TKI Resistance in Non – Small Cell Lung Cancer**. *Clin Cancer*

Res 2014, **20**:4059–74.
43. O'Connell K, Li J, Engler F, Hennessy K, O'Neill F, Straubinger R, Qu J, O'Connor R: **Determination of the Proteomic Response to Lapatinib Treatment using a comprehensive and reproducible ion-current-based proteomics strategy**. *Proteomics Genomics Res* 2013, **1**:27–42.
44. Imami K, Sugiyama N, Imamura H, Wakabayashi M, Tomita M, Taniguchi M, Ueno T, Toi M, Ishihama Y: **Temporal profiling of lapatinib-suppressed phosphorylation signals in EGFR/HER2 pathways.** *Mol Cell Proteomics* 2012, **11**:1741–1757.
45. Ezkurdia I, Vázquez J, Valencia A, Tress M: **Analyzing the first drafts of the human proteome**. *J Proteome Res* 2014, **13**:3854–3855.
46. Foster JM, Degroeve S, Gatto L, Visser M, Wang R, Griss J, Apweiler R, Martens L: **A posteriori quality control for the curation and reuse of public proteomics data**. *Proteomics* 2011, **11**:2182–2194.
47. Potel CM, Lin M-H, Heck AJR, Lemeer S: **Defeating major contaminants in Fe 3+-IMAC phosphopeptide enrichment**. *Mol Cell Proteomics* 2018, **17**:1028–1034.
48. Abelin JG, Patel J, Lu X, Feeney CM, Fagbami L, Creech AL, Hu R, Lam D, Davison D, Pino L, et al.: **Reduced-representation Phosphosignatures Measured by Quantitative Targeted MS Capture Cellular States and Enable Large-scale Comparison of Drug-induced Phenotypes**. *Mol Cell Proteomics* 2016, **15**:1622–1641.
49. Huang KL, Li S, Mertins P, Cao S, Gunawardena HP, Ruggles K V., Mani DR, Clauser KR, Tanioka M, Usary J, et al.: **Proteogenomic integration reveals therapeutic targets in breast cancer xenografts**. *Nat Commun* 2017, **8**.
50. Emdal KB, Dittmann A, Reddy RJ, Lescarbeau RS, Moores SL, Laquerre S, White FM: **Characterization of in vivo resistance to osimertinib and JNJ-61186372, an EGFR/Met bi-specific antibody, reveals unique and consensus mechanisms of resistance**. *Mol Cancer Ther* 2017, **16**:2572–2585.
51. Chiang CK, Tworak A, Kevany BM, Xu B, Mayne J, Ning Z, Figeys D, Palczewski K: **Quantitative phosphoproteomics reveals involvement of multiple signaling pathways in early phagocytosis by the retinal pigmented epithelium**. *J Biol Chem* 2017, **292**:19826–19839.
52. Panizza E, Branca RMM, Oliviusson P, Orre LM, Lehtiö J: **Isoelectric point-based fractionation by HiRIEF coupled to LC-MS allows for in-depth quantitative analysis of the phosphoproteome**. *Sci Rep* 2017, **7**:1–13.
53. Parker J, Xavier D, Molloy MP, Parker R, Cebon J, Amirkhani A, Vella LJ: **Phosphoproteomic Analysis of Cell-Based Resistance to BRAF Inhibitor Therapy in Melanoma**. *Front Oncol* 2015, **5**:1–13.
54. Roolf C, Junghanss C, Murua Escobar H, Schaab C, Dybowski N, Mueller S, Godl K, Sekora A, Knuebel G, Tebbe A: **Phosphoproteome Analysis Reveals Differential Mode of Action of Sorafenib in Wildtype and Mutated FLT3 Acute Myeloid Leukemia (AML) Cells**. *Mol Cell Proteomics* 2017, **16**:1365–1376.
55. Vyse S, McCarthy F, Broncel M, Paul A, Wong JP, Bhamra A, Huang PH: **Quantitative phosphoproteomic analysis of acquired cancer drug resistance to pazopanib and dasatinib**. *J Proteomics* 2018, **170**:130–140.
56. González-Suárez E, Sanz-Moreno A: **RANK as a therapeutic target in cancer**. *FEBS J* 2016, **283**:2018–2033.
57. Bailey ST, Miron PL, Choi YJ, Kochupurakkal B, Maulik G, Rodig SJ, Tian R, Foley KM, Bowman T, Miron A, et al.: **NF-κB activation-induced anti-apoptosis renders HER2-positive cells drug resistant and accelerates tumor growth.** *Mol Cancer Res* 2014, **12**:408–420.
58. Renema N, Navet B, Heymann D: **RANK – RANKL signalling in cancer**. *Biosci Rep* 2016, **36**.
59. Zoi I, Karamouzis M V, Adamopoulos C, Papavassiliou AG: **RANKL Signaling and ErbB Receptors in Breast Carcinogenesis**. *Trends Mol Med* 2016, **22**:839–850.
60. Linger R, Keating A, Shelton Earp H, Graham D: **TAM Receptor Tyrosine Kinases: Biologic Functions, Signaling, and Potential Therapeutic Targeting in Human Cancer**. *Adv Cancer Res* 2008, **100**:35–83.
61. Schoumacher M, Burbridge M: **Key Roles of AXL and MER Receptor Tyrosine Kinases in Resistance to Multiple Anticancer Therapies**. *Curr Oncol Rep* 2017, **19**.
62. Liu L, Greger J, Shi H, Liu Y, Greshock J, Annan R, Halsey W, Sathe GM, Martin A, Gilmer TM: **Novel Mechanism of Lapatinib Resistance in HER2-Positive Breast Tumor Cells : Activation of AXL**. *Cancer re* 2009, **69**:6871–6879.

FOUR

Part 3

Studying Adaptive Resistance



CHAPTER FIVE

Adaptive Resistance to EGFR-Targeted Therapy by Calcium Signaling in NSCLC Cells

Celine Mulder[1], Nadine Prust[1], Sander van Doorn[1], Maria Reinecke[2,3], Bernhard Kuster[2,3], Paul van Bergen en Henegouwen[4], and Simone Lemeer[1]

Molecular Cancer Research; 2018

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands
[2] Chair of Proteomics and Bioanalytics, Technical University Munich, Freising, Gemany.
[3] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.
[4] Division of Cell Biology, Science Faculty, Department of Biology, Utrecht University, Utrecht, the Netherlands.

Introduction

Constitutively active kinases often drive tumour cell survival and proliferation. Consequently, a multitude of small molecule inhibitors targeting these kinases have been developed. About 23 tyrosine kinase inhibitors (TKIs) have been approved to date, showing promising results in the clinic [1,2]. However, despite an initial positive response, most patients eventually develop resistance against the inhibitor. Long term, or acquired, resistance has been studied extensively, mainly concentrating on identifying additional mutations that are either acquired or pre-existing [3]. Simultaneous inhibition of multiple potential resistance mechanisms delays, however does not prevent development of resistance [4,5]. Therefore, the most challenging question remains how cancer cells adapt to and survive initial inhibitor treatment, enabling them to develop acquired resistance.

Currently, two major hypotheses have been formulated trying to explain the development of resistance [6]. First, intrinsically resistant clones may exist through the heterogeneity within the tumour [7]. These pre-existing clones would be positively selected by targeted treatment. Alternatively, resistance could arise as tumour cells collectively adapt under pressure of drug treatment. This would be similar to antibiotic resistance caused by premature ending of a treatment: incomplete eradication of a tumour could strengthen remaining cancer cells [8]. Consequently, resistance can develop when proliferation is inhibited but cell death is not triggered. Unfortunately, small molecule inhibitors are mostly cytostatic, in contrast with the more cytotoxic chemotherapies. Nevertheless, they are highly specific and show great promise in clinical studies, and should therefore not be discarded [9]. In order to improve efficacy, inhibitors are combined with either chemotherapy, immunotherapy or another form of targeted therapy. Compared to mono-treatment, combinational treatment often improves the patients response, but unfortunately also increases toxicity [10,11].

Lung cancer is one of the most prevalent cancer types and has a very high mortality rate in both men and women. Non-small-cell lung cancer (NSCLC) makes up approximately 85% of the lung cancer cases [12] and is currently treated by surgery, radiation, chemotherapy, targeted therapy with small molecule inhibitors or a combination thereof. Epidermal growth factor receptor (EGFR) overexpression and mutations are driving the disease in a substantial percentage (estimated more than 60%) of NSCLC patients [13,14].

Here we set out to identify the initial survival pathways that become activated in NSCLC cells harbouring an EGFR activating mutation, after treatment with the EGFR specific inhibitor afatinib, and subsequently to inhibit these pathways to increase drug efficacy. By targeting the pathways tumour cells use to adapt and survive drug treatment, the specificity of the small molecule inhibitor is further exploited: only cells that are respond-

ing to the first treatment will be sensitive to the second, preferably cytotoxic, agent. This differs from the standard combination with chemotherapy, which targets all rapidly dividing cells. We used mass spectrometry to elucidate the changes in the proteome, kinome and phosphoproteome in the NSCLC cell line PC9, induced by the tyrosine kinase inhibitor (TKI) afatinib, which specifically targets EGFR. Here we show that, upon treatment, NSCLC cells differentially regulate Ca^{2+} signalling and enhance Ca^{2+} dependent cell adhesion. By limiting the access to extracellular calcium during TKI treatment, we increased growth inhibition and induced apoptosis, decreasing the abilities of the cells to develop acquired resistance. These results were confirmed in multiple combinations of NSCLC cell lines with EGFR activating mutations and several EGFR inhibitors.

Results

PC9 cells showed a clear but incomplete response to TKI treatment

With the aim of performing a multi-omic analysis of initial drug resistance, we first characterized the cellular response to small molecule inhibitor treatment. The NSCLC cell line PC9, which harbours an activating deletion in the EGFR (DelE746A750), was highly sensitive to treatment with first (gefitinib, erlotinib) and second generation (afatinib) EGFR inhibitors, as shown by the measurement of viability after 3 days of treatment (Fig. 1A). However, despite the pronounced effect, around 30% of viability persisted. In order to better characterize the properties of the persisting cells, we analysed various hallmarks of cancer, including proliferation, apoptosis and migration. Proliferation analysis showed that untreated cells grew to full confluence within 3 days, where cells treated with 10nM, 100nM or 1μM afatinib, or 1μM erlotinib or gefitinib, needed 6-7 days (Fig. 1B). Although the proliferation rate of treated cells decreased, cells were still able to grow to full confluence despite the continuous inhibition of the primary drug target, as shown by the decreased levels of EGFR-Y1068 phosphorylation (Fig. 1C).

For apoptosis measurements cells were treated with DMSO (negative control), 1 or 10 μM afatinib (cytotoxic concentration as positive control). Untreated cells grew to full confluence, whereas 10 μM afatinib treated cells showed 100% apoptosis, consequently measurements were discontinued for these conditions after 3 days (Supplementary Fig. S1A). Cells treated with 1 μM afatinib showed a significant increase in levels of apoptosis compared to untreated cells (Fig. 1D). However, when compared to the positive control, only one fifth of the response was detected, indicating that the majority of the cells did not go in apoptosis. More surprisingly, after 6 days of afatinib treatment, levels of apoptosis decreased again (Fig. 1D). Finally, complete abolishment of migration after treatment was

shown by a scratch assay (Figure 1E, Supplementary Fig. S1B). This lack of migration caused formation of tight clusters of cells (Fig. 1F and Supplementary Videos). Collectively, these results show a clear response by PC9 cells upon afatinib treatment: it abolished migration and induced both growth inhibition and a small amount of apoptosis. However, a majority of cells persisted. After only 3 days, these cells were increasingly resistant to apoptosis and continued to grow despite the continuous inhibition of the EGFR.

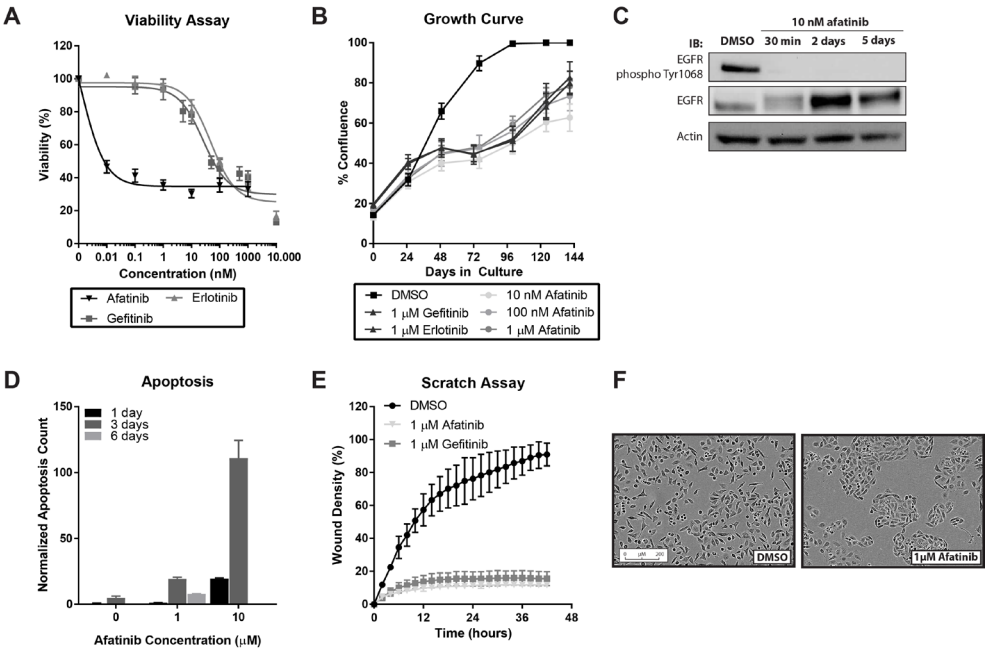


Figure 1: A: To determine drug response, PC9 cells were treated with different concentrations of three inhibitors (afatinib, gefitinib and erlotinib) for 3 days. Viability was measured using AlamarBlue®. Relative viability of cells treated with inhibitors as compared to untreated cells is plotted on the Y-axis. Error bars represent 4 replicate wells within one experiment. **B**: Growth curves of PC9 cells showed incomplete growth inhibition upon EGFR inhibition, independent of the type of inhibitor or concentration. Confluence of cells treated with 10 nM, 100 nM or 1 μM TKI was monitored using the IncuCyte System over a time span of 6 days. Error bars represent quadruple wells within one experiment. Medium in the wells was refreshed every 3 days. **C**: Western blot of PC9 cells treated with DMSO or 10 nM Afatinib for 30 minutes, 2 days or 5 days showed clear inhibition of EGFR activity during our experiments. Antibodies were used against the EGFR protein to show protein expression and against EGFR phosphorylated on Tyr1068 as indication of activity. Actin was used as a loading control. **D**: To monitor apoptosis during EGFR inhibition, PC9 cells were treated with 1 μM Afatinib for 1, 3 or 6 days. Upon treatment, only minor amounts of apoptosis were observed, as compared to a negative control (DMSO) and a positive control (10 μM Afatinib, cytotoxic concentration). Both negative and positive controls were discontinued after 3 days. Error bars represent quadruple replicate wells of which four images were taken as technical replicates. Apoptosis was measured using the Caspase 3/7 green fluorescent probe using the IncuCyte System. Fluorescent count was normalized on well confluence per image. **E**. In order to quantify migration, confluent PC9 cells were treated with DMSO or 1 μM Afatinib or Gefitinib for up to 48 hours. At the beginning of this treatment, a scratch was made in each 96-well and confluence within the scratch was measured using the IncuCyte System. Error bars represent four replicate scratches. **F**: Representative light microscopy pictures of PC9 cells treated with DMSO or 1 μM Afatinib for 3 days to show clear morphological changes upon treatment.

Mass spectrometry-based multi-omic profiling of afatinib response

In order to detect the changes that would be responsible for continued growth in the presence of the drug, PC9 cells were treated with afatinib for 1 to 7 days at a concentration of 10 nM and the proteome, phosphoproteome and kinome were analysed at various time points (1-7 days). The selected concentration represents 10x the concentration at which we start to observe the stable presence of viable cells (Fig. 1A). After treatment, proteins were fractionated using high-pH chromatography for full proteome analysis (Supplementary Table 1). PCA analysis showed clustering of replicates and a clear separation of time points, indicating changes on proteome level over time (Supplementary Fig. S2A). An additional kinobead pulldown was used to detect expression changes on kinome level (Supplementary Table 2) [15]. As shown by the PCA plot, most variation exists between DMSO-24h-48h and 72h-7days, whereas less difference exists within these two groups (Supplementary Fig. S2B).

To detect changes at the signalling level, phosphopeptides were enriched using a Fe^{3+}-IMAC column [16]. Cells were treated with 10 nM afatinib for 30 minutes to determine immediate drug effects, and for 2 and 5 days to detect long term changes (Supplementary Table 3). A PCA plot of all identified peptides showed a clear separation between untreated and 30 minutes, versus 2 and 5 days of treatment (Supplementary Fig. S2C). The collective dataset (Supplementary Tables 1, 2 and 3) comprises quantitative information for > 5,600 proteins, > 270 protein kinases and > 22,900 phosphopeptides (>18,400 unique phosphorylation sites).

An ANOVA test was performed to detect all differentially regulated proteins between time points (Fig. 2A). Up- and down regulated proteins were analysed for GO terms, showing decreased translation and mRNA processing, and increased actin cytoskeleton construction and calmodulin binding.

Hierarchical clustering of phosphoproteome data shows in total 2483 regulated phosphosites with a p-value <0.05 according to an ANOVA test, in at least one time point vs DMSO (Fig. 2B). In order to include reactivated phosphosites, regulated sites between 2/5 days vs 30 minutes were also included in the clustering. Based on the clustering we selected four clusters on which GO term analysis was performed. Interestingly, all clusters showed enrichment for adhesion and cytoskeleton reorganization related pathways, which is in correlation with the proteome analysis. The two clusters showing regulation between 30 minutes vs. 2-5 days of treatment were additionally enriched for cell cycle related pathways.

Unlike proteomics data, the direction of regulation of a phosphopeptide is not always indicative of the direction of the activity of the protein: phosphorylation can either inhibit or activate the protein. Since functional consequences of phosphorylation are often not known, it can be difficult to define changes of functionality based only on measured ratios. To solve

this problem, we transformed peptide level ratios into functionality scores on the protein level using the tool PHOTON [17]. Hierarchical clustering of PHOTON scores indicated that most of the activated sites were enriched for Ca^{2+}-related signalling (Fig. 2C). As expected, cell cycle related pathways showed decreased activity, confirming that afatinib induced growth inhibition. The cluster showing a decrease after 30 minutes, but an increase after 2 and 5 days, was enriched for ErbB Signalling. This implies that initially the ErbB pathway is effectively inhibited, but seems to regain activity in time, despite the effective inhibition of EGFR (Fig. 1C).

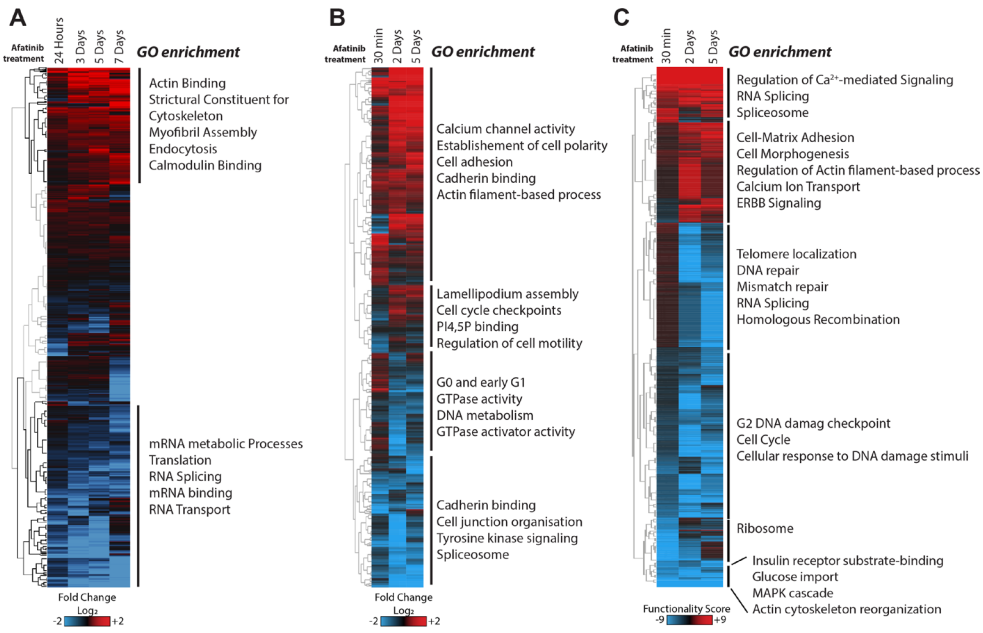


Figure 2: A: Summarized proteome response of PC9 cells treated with 10 nM Afatinib for 1, 3, 5 or 7 days. Hierarchical clustering of the $\log_2$ ratios of the proteins that were significantly distinguished by an ANOVA test between the time points are shown. GO term enrichment analysis, as depicted on the right, demonstrated up-regulation of actin binding processes and down-regulation of mRNA and translation related processes. **B**: Phospho-proteome data of PC9 cells after afatinib treatment showed regulation of cell adhesion, cell cycle and spliceosome related processes. PC9 cells were treated with 10 nM Afatinib for 30 minutes, 2 and 3 days. All phospho sites that were differentially regulated in at least one time point, as compared to untreated cells or compared to 30 minutes, were shown in a hierarchical clustering. Different clusters were analysed for GO term enrichment, as depicted on the right. **C**: Phospho-proteome data was reanalysed using PHOTON, resulting in functionality scores. All phospho-proteins significantly changing in at least one time point were shown in a hierarchical clustering. Different clusters, based on the hierarchical clustering, were analysed for enriched GO terms, as denoted on the right. In addition to the recurring regulation of adhesion and cell cycle signalling, calcium signalling was enriched within the activated phospho-proteins.

Reactivated pathways downstream of EGFR

As our phospho proteomic data indicates, growth activating pathways downstream of EGFR are re-activated within days. Prominent reactivation of PI3K/mTOR signalling after initial inhibition was observed, for example, indicated by recovery of activity-determining S235/236 phosphorylation in RPS6, a major downstream target of mTOR (Fig. 3A). Western blot analysis of RPS6 phosphorylation confirmed its recovery, whereas RPS6 levels remained constant (Fig 3B). In addition, MEK/ERK signalling recovered rapidly, as indicated by regained ERK1/2 phosphorylation of the activity-determining sites (Fig. 3A) but unchanged protein expression, which was again confirmed by western blot analysis (Fig. 3B). Reactiva-

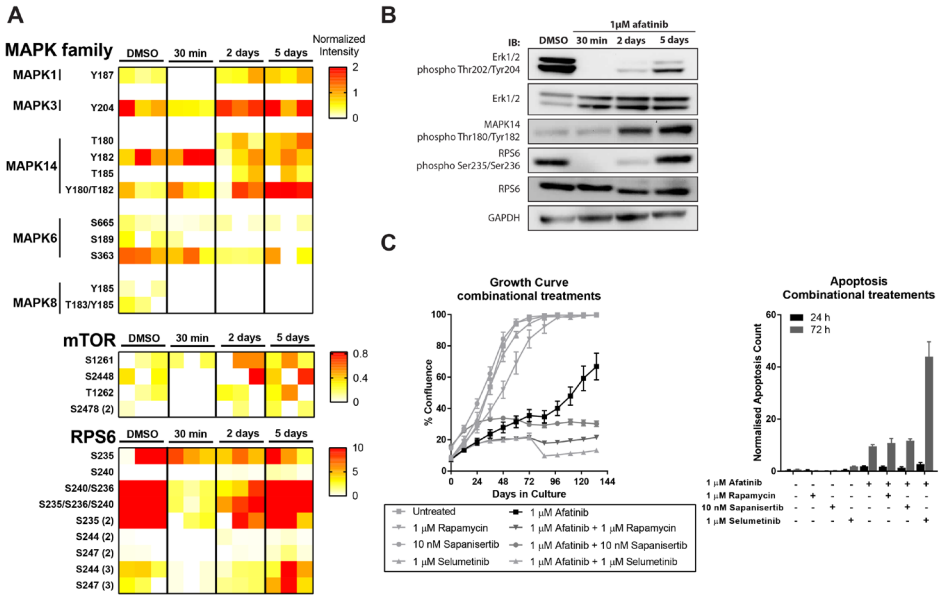


Figure 3: A: MS data of specified phospho proteins showed reactivation of activity determining sites in proteins involved in proliferation stimulating pathways, including ERK1/2, mTOR and RPS6. For each individual heat map, each column represents a replicate of the specified time point and each row an individual phospho peptide. If a row is annotated with more than one site, or a number in brackets is shown, the peptide was singly, doubly or triply phosphorylated. Peptides with no multiplicity shown are singly phosphorylated. The colour is indicative of the normalized intensity. For this, each protein or protein family has his own legend, due to dynamic range differences. **B**: Validation of MS data on western blot showed similar patterns of reactivated phospho-sites. PC9 cells were treated with 1 μM Afatinib for 30 minutes, 2 and 5 days. Antibodies were used against ERK1/2, ERK1/2 phosphorylated on Thr202 and Thr204 respectively, MAPK14 doubly phosphorylated on Thr180 and Thr182, RPS6 and RPS6 doubly phosphorylated on Ser235 and Ser236. GAPDH was used as a loading control. **C**: MEK and mTOR reactivation was validated by monitoring growth and apoptosis during combinational treatment of EGFR inhibition combined with mTOR (Rapamycin: mTOR1; Sapanisertib: mTOR1/2) or MEK (Selumetinib). Growth was monitored in the IncuCyte System (left). Error bars represent triplicate replicate wells within one experiment. Apoptosis was measured using the Caspase 3/7 Fluorescent probe (right). Fluorescent count was normalized on well confluence per image. Error bars represent triplicate replicate wells of which four images were taken as technical replicates. In both combinations growth inhibition was increased, however only the combination of EGFR + MEK inhibition resulted in increased levels of apoptosis.

tion of both pathways has previously been shown to occur after acquired EGFR inhibitor resistance in a subset of clones [4,5,18]. In order to test whether this reprogrammed signalling could pharmacologically be exploited, we performed combination treatments with the mTOR inhibitors rapamycin (mTORC1) and sapanisertib (mTORC1/2) and the MEK inhibitor selumetinib. As shown in Figure 3C, combining rapamycin with afatinib significantly decreased the growth capabilities of the cells compared to afatinib treatment alone, but it did not increase the amount of apoptosis. Treatment with rapamycin alone did not affect cell viability nor apoptosis. Since rapamycin is solely inhibiting mTORC1, we included the pan mTOR inhibitor sapanisertib (or INK128) in our analysis. Sapanisertib in combination with afatinib did not increase apoptosis and showed similar growth inhibition when compared to afatinib/rapamycin combination treatment. These results indicate that mTORC2 signalling does not contribute to reactivated mTOR signalling after afatinib treatment. In addition these results indicate that combined mTORC1/afatinib inhibition mainly induces cytostatic effects and consequently the development of further resistance pathways cannot be excluded. Combination of the MEK inhibitor selumetinib with afatinib also resulted in a prominent reduction of cell growth capabilities. However in contrast to mTOR inhibition and in agreement to what has been shown previously [5], MEK inhibition had a tremendous impact on the amount of apoptosis induced, indicating that such treatment could be beneficial for the eradication of tumour cells (Fig. 3C).

To find the upstream kinases that could be responsible for reactivation of mTOR or ERK1/2 we further scrutinized our data. A well-known acquired resistance mechanism after TKI treatment is the upregulation of different (receptor) tyrosine kinases [19,20] that would re-fuel growth inducing signalling. However, in our (phospho)proteome analysis we find no obvious indications that this would be the case. Also, the kinobead data did not provide evidence that another tyrosine kinases takes over EGFR function (Supplementary Table 2). In the kinobead data, EGFR levels significantly decreased at all time points (Supplementary Fig S2D, Fig 4A). This was in contrast to the full proteome analysis (Fig. 4B) and western blot analysis (Fig. 1C). This discrepancy can be explained by the fact that afatinib irreversibly binds EGFR, which makes EGFR no longer accessible for binding to the kinobeads. This again strongly indicated the continuous presence of the drug and inhibition of the receptor during the course of our experiments.

Ca^{2+} Signalling and Cytoskeleton organizing related proteins were highly affected by treatment

As previously discussed, we detected strong re-activation of PI3K/mTOR and ERK pathways, but could not identify an obvious upstream kinase that would be responsible. However, we noticed a large number of regulated (phospho)proteins to be involved in the reorganization of the actin cyto-

skeleton and Ca^{2+}/Calmodulin related signalling (Fig. 2). For example, the protein levels of Ca^{2+}-dependent protease Calpain-2 (CAPN2) significantly decreased after afatinib treatment (Fig. 4B). This important protease is responsible for degradation of focal adhesions and Ca^{2+} pumps [21,22]. In line with this, we found the calcium transporting ATPases, ATP2B4 and ATP2C1 to be significantly upregulated. In addition, the actin cross-linking proteins MARCKS, MARCKSL1 and Calponin-2 (CNN2) were significantly upregulated after treatment. These are proteins that bind actin, resulting in more stiffness and increased adhesion [23,24]. Spectrin super family members, like α-Actinin, SPTAN1 and SPTBN1 that can connect different components of the cytoskeleton, were also upregulated [24].

Levels of EphA4 increased (Fig. 4A, Supplementary Fig S2D), which has been associated with cytoskeleton rearrangements, inhibiting migration and invasion in NSCLC cell lines and increasing survival in lung cancer patients [25,26]. In contrast, levels of EphA2, a known marker of TKI resistance and stimulator of migration, decreased [27]. The levels of Fyn and SIK1 increased over the course of 7 days. Fyn is a Src family kinase, targeting to adherence junctions and a key player in the apoptosis-resistance of tumour cells [28]. The tumour-suppressing kinase SIK1 phosphorylates and thereby stabilizes p53 to initiate anoikis in case of decreased adhesion [29].

Also, the phosphoproteome data showed regulation of pathways related to Ca^{2+} signalling and actin cytoskeleton organization. After afatinib treat-

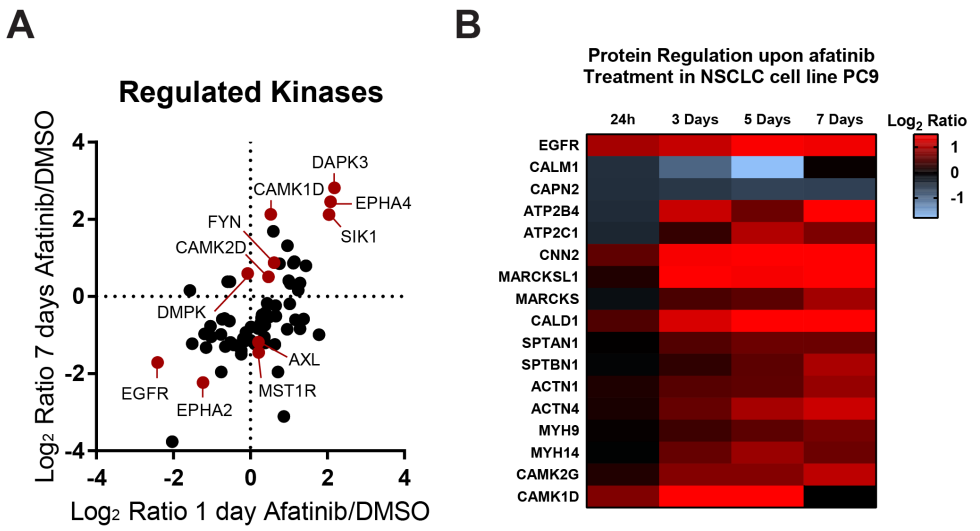


Figure 4: A: Upon afatinib treatment, several adhesion related kinases were differentially regulated, as measured by the kinobead assay. On the X-axis the $\log_2$ ratio after 24 hours of afatinib treatment is plotted, on the Y-axis the $\log_2$ ratio after 7 days of treatment. Highlighted in red are proteins of interest, labelled with their protein name. **B:** Similarly, in the proteomics data several adhesion related proteins were regulated. Log_2 ratios are shown for proteins of interest from the full proteome data set. Colour coding is based on the mean of the three replicate MS runs.

ment, two phosphosites of Focal Adhesion Kinase (FAK), the main kinase in the organisation of focal adhesions, were found to be downregulated in the phospho data: Ser910 and Ser843, whereas protein levels remained constant (Fig. 5A, Supplementary Table 1,3). Decreased phosphorylation of Ser910 has been shown to reduce proliferation [30]. Phosphorylation of Ser843 has been shown to increase upon cellular detachment, implying that the decreased phosphorylation after afatinib treatment is following from increased adhesion [31]. In addition, loss of both Ser910 and Ser834 phosphorylation changes cellular morphology and abolishes migration. Tyrosine phosphorylation is, due to relative low abundance, underrepresented in our phosphoproteomics data. Therefore, in order to complement the hypothesis that afatinib treatment is influencing focal adhesion dynamics, some tyrosine phosphorylation sites with established biological functions were analysed by western blot (Fig 5B) [32]. Indeed Tyr397 and Tyr576/577 of FAK as well as Tyr31 and Tyr118 of Paxillin, sites known to stabilize focal adhesions, showed an increase in phosphorylation upon afatinib treatment (Fig 5B).

The phosphorylation status of Ca^{2+}-binding protein Caldesmon has been linked to its relocation to the cytoskeleton and function to stabilize actin and increased motility [33,34]. As shown by the proteomics data, the protein abundance of Caldesmon increased upon afatinib treatment (Fig. 4B). In addition, phosphorylation of Caldesmon increased (Fig. 5A). The most severely changing site, Ser789, has been shown to be phosphorylated by ERK1/2 [35]. A possible kinase of Caldesmon, CDK14, increased in our kinobead data (Supplementary Fig. S2D).

In non-muscle cells, three genes comply the myosin protein complex: myosin 9, 10 and 14 [33]. Upon phosphorylation, myosins are stabilized, causing increased tension, regulating cell shape, cell polarity and migration by forming actin-binding filaments. Afatinib treatment increased phosphorylation of myosin (Fig. 5), confirming increased levels on a protein level (Fig. 4B, Fig. 5B). Possible kinases of myosin, the DAPK family, were increasingly phosphorylated and therefore activated (Fig. 5A). DAPKs can be phosphorylated by CAMKs, which increased in both abundance (Fig. 4B, Supplementary Fig. S2B) and phosphorylation (Fig. 5A). Another stimulator of myosin phosphorylation is DMPK, via phosphorylating and inhibiting myosin phosphatase MYPT1 [36]. Phosphorylation of MYPT1 on the inhibitory site Thr696 by DMPK was significantly increased after 2 and 5 days of TKI treatment (Fig. 5A).

Depletion of extra-cellular Ca^{2+} increases sensitivity to TKI treatment

Signalling resulting from increased adhesion has previously been shown to be anti-apoptotic [37,38] and adhesion pathways are largely regulated by Ca^{2+} signalling [39]. PC9 cells treated with different TKIs completely abolished migration and formed tight clusters, while resisting apoptosis

(Fig. 1). Consistent with this, proteomic analysis showed regulation of both adhesion related proteins and of Ca^{2+} signalling. Therefore, we hypothesized that the cells use increased adhesion induced by altered Ca^{2+} signalling to resist apoptosis in response to early EGFR inhibition, and that making extracellular calcium unavailable to the cells could inhibit this adhesion and induce apoptosis.

Therefore, we treated PC9 cells with different EGFR inhibitors in the presence or absence of either EGTA, a specific chelator of Ca^{2+} to remove extracellular Ca^{2+}, or Lanthanum(III) Chloride ($LaCl_3$) to block Ca^{2+} uptake. To find a wide applicability we included the first generation EGFR inhibitors erlotinib and gefitinib, in addition to afatinib, in our analysis. Cells treated with TKI alone grew in 6 days from 20% to 90% confluence. In the presence of EGTA or $LaCl_3$ though, cells grew in 6 days from 20% to less than 40% confluence, showing increased proliferation inhibition by the addition of EGTA or $LaCl_3$ (Fig. 6A, left panel). In addition, levels of apoptosis were significantly higher in the presence of EGTA or $LaCl_3$ (Fig. 6A, right panel). Effects were comparable between the three different inhibitors used, in-

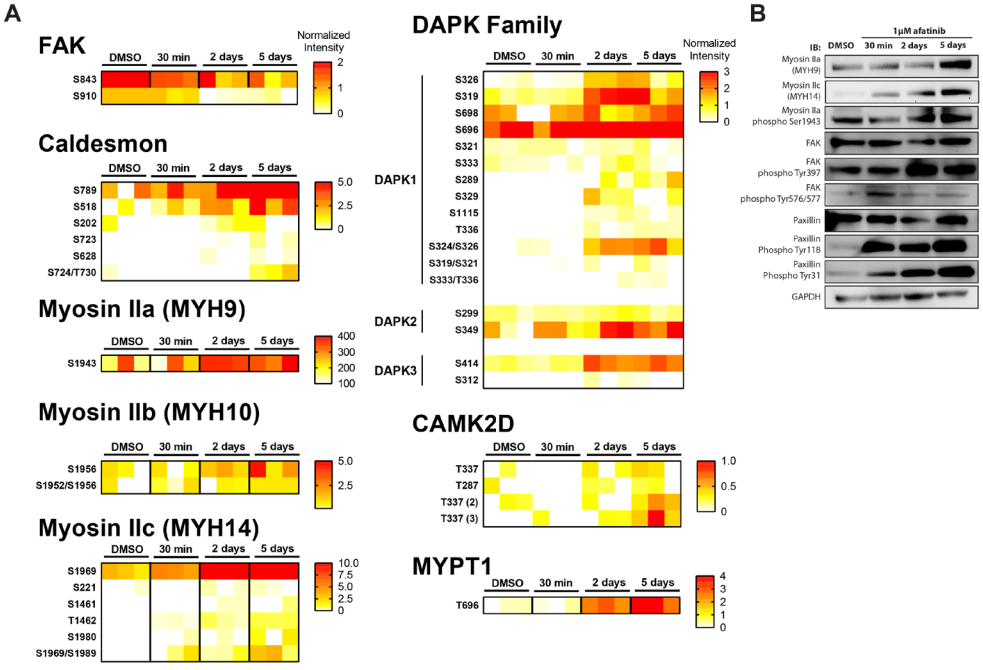


Figure 5: A: Regulation of phospho profiles are shown in detail for several proteins of interest. Each column represents a replicate of the specified time point. Each row represents a phospho peptide. If a row is annotated with more than one site, or a number in brackets is shown, the peptide was singly, doubly or triply phosphorylated. Peptides with no multiplicity shown are singly phosphorylated. The colour is indicative of normalized intensity. For this, each protein or protein family has his own legend, due to dynamic range differences. **B**: Western blot of PC9 cells treated with 1 μM afatinib for 30 minutes, 2 or 5 days showed increased levels of phosphorylation of Myosin IIa and of protein levels of Myosin IIa and IIc. In addition, the western blot analysis shows increase tyrosine phosphorylation of FAK Tyr397 and Tyr576/577 and Paxillin Tyr31 and Tyr118. GAPDH was used as a loading control.

dicating that the development of a Ca^{2+} dependency is not drug specific, but rather EGFR inhibition specific.

To determine whether the observed effect was cell line specific, we tested an additional 3 NSCLC cell lines with different genetic backgrounds and different sensitivities towards the TKIs used (Supplementary Fig. S3A). HCC827 has the exon19del E746-A750 deletion and is highly sensitive to both first (erlotinib and gefitinib) and second (afatinib) generation TKIs. H1975 has both the L858R and the T790M mutation, making it resistant to first, but sensitive to second generation TKIs. H1650 has the exon19del E746-A750 deletion in combination with PTEN loss, resulting in resistance to first generation TKIs, and a response to treatment with afatinib only upon 1 µM and higher. In all cell lines, neither growth nor apoptosis rate was affected by EGTA treatment alone up to 1mM (Fig 6B, Supplementary Fig. S3B). However, both the magnitude of the growth inhibition and the

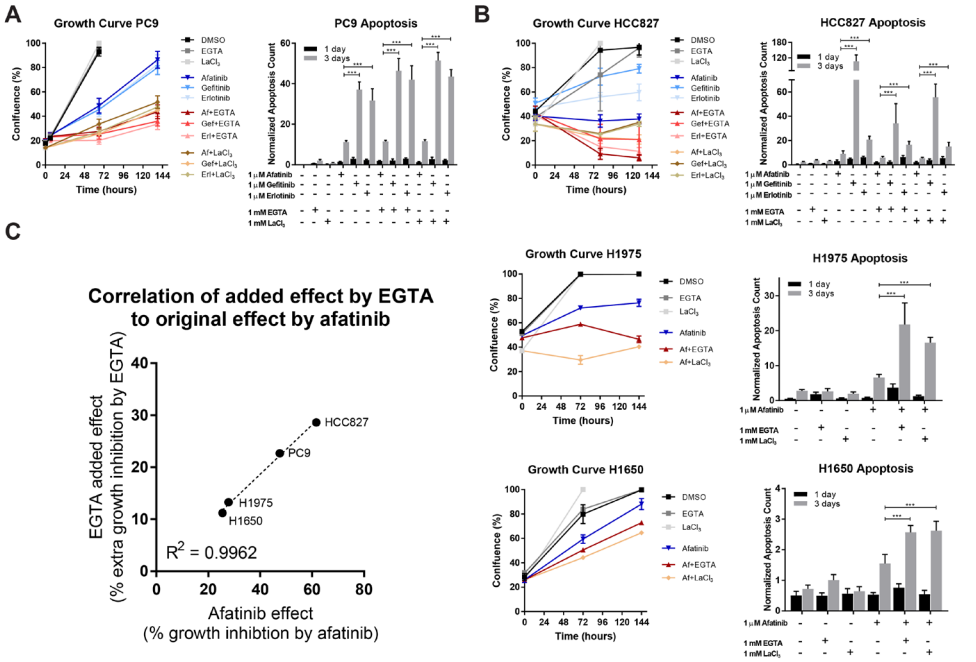


Figure 6: Influence of Ca^{2+} signalling in combination with EGFR inhibition was shown using growth curves and levels of apoptosis. **A**: PC9 cells were grown in the presence or absence of 1 mM EGTA or 1 mM $LaCl_3$, and simultaneously treated with 1 µM EGFR inhibitors (afatinib, gefitinib or erlotinib). Percentage of confluence was measured to show increased growth inhibition and induction of apoptosis, independent of inhibitor used, as measured with green fluorescent Caspase 3/7 Probe. **B**: Combined EGFR inhibition and interference with Ca^{2+} availibility was explored on three different NSCLC cell lines. Confluence and apoptosis was measured in the IncuCyte System. *p*-values are depicted above bars: ***:<0.001. All error bars represent the data of four replicate wells within one experiment. **C**: Growth inhibition induced by afatinib or by afatinib in the presence of EGTA showed great correlation, as shown by a scatter plot. The mean percentage of inhibition after 3 days induced by afatinib is plotted on the X-axis, where the mean percentage of added effect of EGTA in combination with afatinib was plotted on the Y-axis. The dotted line indicates the linear regression, the R^2 is depicted in the graph.

amount of apoptosis significantly increased after interference with Ca^{2+} signalling by EGTA or $LaCl_3$ (Fig. 6B). Interestingly, in all cases apoptosis within the first 24 hours was negligible, but rose significantly in the following 2-3 days.
The effect of the combinational treatment was comparable between first and second generation TKIs. In contrast, however, the degree of the added effect differed between cell lines, demonstrating high correlation with the sensitivity to the drug (Fig. 6C). HCC827 cell line, which was the most sensitive, displayed the largest added effect of addition of EGTA. H1650 was the most resistant to TKI treatment, showing the lowest increase in levels of apoptosis and growth inhibition after addition of EGTA. In general, the more sensitive a cell line was to TKI treatment, the more sensitive it was for addition of EGTA.

Discussion

While targeted approaches in cancer therapy show great promise over conventional chemotherapy, development of resistance remains an important obstacle to overcome [9]. The majority of resistance studies have focused on acquired resistance mechanisms that occur after an extensive period of treatment, usually by screening for additional mutations. Only recently, attention has shifted to early, adaptive resistance, i.e. why sensitive tumour cells initially survive drug treatment, to subsequently gain additional mutations [8,40]. Here we hypothesise that rapid adaptation to drug treatment allows incomplete eradication of the tumour, ultimately resulting in additional mutations and a fully resistant tumour. We used a multi-omics approach to elucidate adaptive resistance mechanisms after EGFR directed TKI treatment with afatinib in NSCLC cells. We demonstrated that NSCLC cells adapt to afatinib inhibition within days and re-activate growth stimulating pathways, despite continues inhibition of the EGFR. Our data indicates that the cells use altered Ca^{2+} and adhesion signalling as a resistance mechanism, attenuating drug-induced cell death. Interfering with adhesion by Ca^{2+} depletion or limiting access to extracellular Ca^{2+} significantly increased cell death in a TKI dependent manner.
Our phospho proteomic data showed that, despite continues inhibition of the EGFR and downregulation of cell cycle related pathways, the growth stimulating mTOR and ERK pathways were reactivated within days. This correlates with the re-activated growth of NSCLC cells and strongly suggests a rapid adjustment to EGFR inhibition. In agreement with this, pharmacological inhibition of mTOR and ERK pathways shows a synergistic effect in combination with afatinib treatment. The ERK pathway is most conventionally activated by upstream tyrosine kinase receptors [41]. However, ERK kinases can also be activated by several other signalling pathways, for example integrin signalling via enhanced adhesion [42]. Indeed, re-activated growth signalling by increased adhesion is supported

by our data in which we find a strong upregulation of adhesion related proteins after EGFR inhibition. Several regulators of the stability of the cytoskeleton were upregulated upon treatment. In addition, the protease calpain-2, responsible for degradation of focal adhesions, was downregulated. It is intriguing to speculate that increased ERK signalling is induced by increased adhesion via Ca^{2+} signalling as the observed effects on growth inhibition and apoptosis upon inhibition of either ERK or Ca^{2+} signalling are very similar.

Cell adhesion is an integral part of the complex and dynamic process of migration. Enhanced adhesion could ultimately result in both increased and decreased migration and the context determines whether it is oncogenic or tumour suppressive. Enhanced adhesion initially results in apoptosis resistance, stimulating the development of drug resistance [40]. However, when cells become increasingly oncogenic due to, for example epithelial to mesenchymal transition (EMT), induction of adhesion is demonstrated to be tumour suppressive [43]. Increased adhesion has been described previously as a resistance mechanism in haematological malignancies, referred to as cell-adhesion mediated drug resistance (CAM-DR) [44]. EphA4, a key player in CAM-DR, was also shown to be upregulated after afatinib treatment in our dat [45]. Inhibition of apoptosis in NSCLC has been shown on a mRNA level to be driven by the focal adhesion pathway [40]. In our data, the main kinase involved in the regulation of focal adhesions, FAK, showed altered levels of phosphorylation, which has been associated with morphologic changes and increased adhesion [30–32]. In addition, the focal adhesion associated component Paxillin showed increased tyrosine phosphorylation, indicating increased stabilisation of adhesive structures.

From the here presented data, regulation of the cytoskeleton was oncogenic: enhanced adhesion and decreased migration initiated by afatinib treatment most likely resulted in apoptosis resistance. Re-organization of the cytoskeleton can, amongst others, be established by altering Ca^{2+} homeostasis, since many adhesion and cytoskeleton remodelling related processes are regulated in a Ca^{2+}-dependent manner [46,47]. Not surprisingly, various Ca^{2+}/Calmodulin related proteins were found to have increased expression. On the contrary, Calmodulin itself was found to be decreased. The latter is difficult to interpret, solely based on the presented data. Future research will be aimed at determining the regulation of Ca^{2+}/Calmodulin signalling during adaptive resistance.

In order to intervene with Ca^{2+} homeostasis, we cultured the cells in the presence of EGTA, a chelating agent with great affinity for Ca^{2+}, or with $LaCl_3$, an inhibitor of plasma membrane Ca^{2+} transporters. In both cases, the sensitivity to the three EGFR inhibitors tested in four different NSCLC cell lines increased, indicating that EGFR inhibition creates a strong dependency on the accessibility of extracellular Ca^{2+} and in addition implicates a role for Ca^{2+} signalling during the development of adaptive

resistance. More importantly cells treated with EGTA or $LaCl_3$ alone did not show any response, highlighting the specificity of the combinatorial treatment. Mechanistically we suspect that under the influence of EGFR inhibition cells either increase Ca^{2+} usage and deplete their intra-cellular Ca^{2+} stores, or intra-cellular Ca^{2+} becomes inaccessible. From literature, it is known that Ca^{2+} concentrations in the cytoplasm can be influenced by RTKs, including EGFR, which modulate PIP_2/IP_3 signalling [48]. Ca^{2+} oscillations in the cell have been shown to be reduced by inhibition of the EGFR [49], therefore, it is not surprising that the magnitude of extracellular Ca^{2+} availability correlated strongly with TKI sensitivity, suggesting a direct link between inhibited EGFR signalling and Ca^{2+} dependency during adaptive resistance. However, since Ca^{2+} signalling has been implicated in many processes within the cell, future research will be aimed at determining the exact role of Ca^{2+} during the observed adaptation.

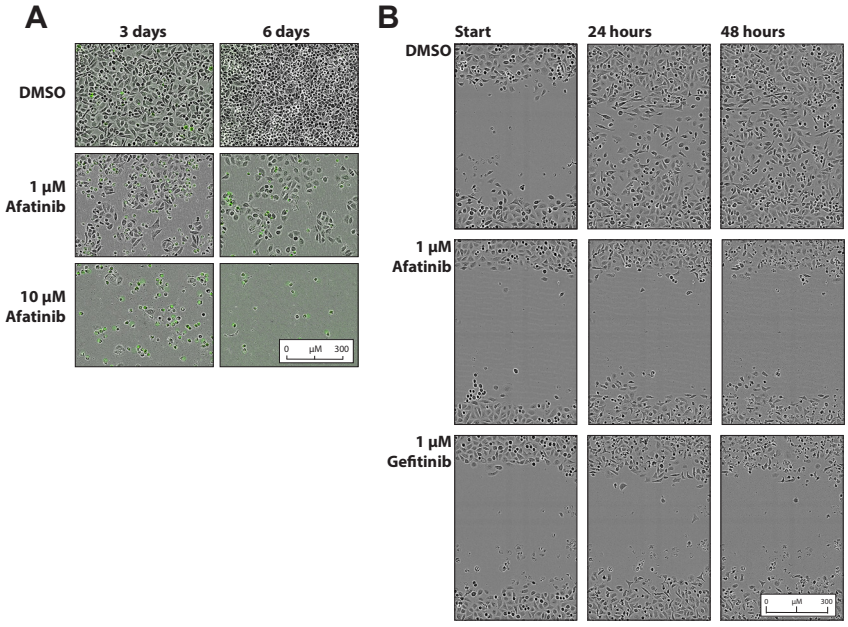
When looking at the extensive added effect Ca^{2+} deprivation, two observations stand out. First, induction of apoptosis, both in the case of TKI alone as well as in combination with EGTA or $LaCl_3$, only occurs 24 hours after the start of the treatment. The fact that the effects induced by an EGFR inhibitor combined with Ca^{2+} deprivation follow the effects induced by the inhibitor alone, is a strong indication that development of Ca^{2+} dependency functions in a TKI dependent manner. Secondly, even though the addition of extracellular Ca^{2+} deprivation caused significantly increased levels of apoptosis, eradication of the cell population is still incomplete. This behaviour, similar to treatment by the TKI alone, demonstrates the adaptive capabilities of cancer cells. Even in harsh circumstances some cells are able to survive, highlighting the need for further research into adaptive resistance mechanisms to targeted therapy.

Preceding clinical research, fundamental research is necessary to find novel targets and therapeutic strategies. Although this study is limited to cell line models, the acquired data could lead to a better understanding of cellular behaviour under the pressure of an already well-studied and clinically applied drug. Using a simplified disease model, our data demonstrates that cells start to modify their signalling directly upon drug exposure, revealing an interesting therapeutic window in the study and treatment of drug resistance. It shows the necessity to increasingly study the short-term and adaptive responses of cells, in contrast with long term, acquired mechanisms. In addition, although on a fundamental level, we revealed the involvement of calcium signalling in the adaptive response of cancer cells. Calcium signalling has been implicated in cancer previously, but to our knowledge we are the first to show a direct link between calcium signalling and adaptive resistance of cancer cells against EGFR inhibition. Several ways exist to target calcium signalling in cancer cells [50], therefore, a logical next step would be exploring the effects of specific small molecule inhibitors combined with existing drugs targeting calcium signalling. In addition, detailed knowledge on the behaviour of key play-

ers in Ca^{2+} signalling, like Calmodulin, will shed more light on the molecular mechanisms underlying the observed effects. Furthermore, findings as presented here, will need to be validated in *in vivo* models in order to be translated into a clinical application.
In conclusion, this study demonstrates that TKI sensitive NSCLC cells use re-organization of their cytoskeleton, regulated by Ca^{2+} signalling, as an adaptive resistance mechanism. Targeting Ca^{2+} signalling sensitizes the tumour cells to TKI treatment in a TKI dependent manner, making the global approach of the Ca^{2+} depletion highly specific. Concentrating on combinational treatment as early as possible after drug exposure in a clinical setting, could improve treatment efficacy by preventing any acquired resistance mechanism from arising, rather than attempting to cure it.

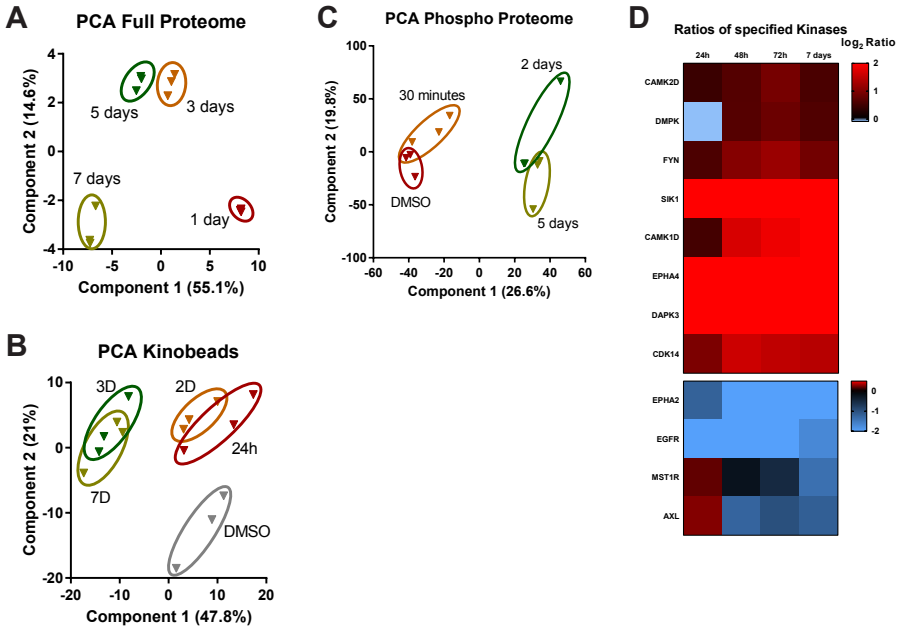
Acknowledgments
SL acknowledges support from the Netherlands Organization for Scientific Research (NWO) through a VIDI grant (project 723.013.008). This work was supported by the Roadmap Initiative Proteins@Work (project number 184.032.201), funded by the Netherlands Organization for Scientific Research (NWO).

ProteomeXchange Accession
The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD008454.



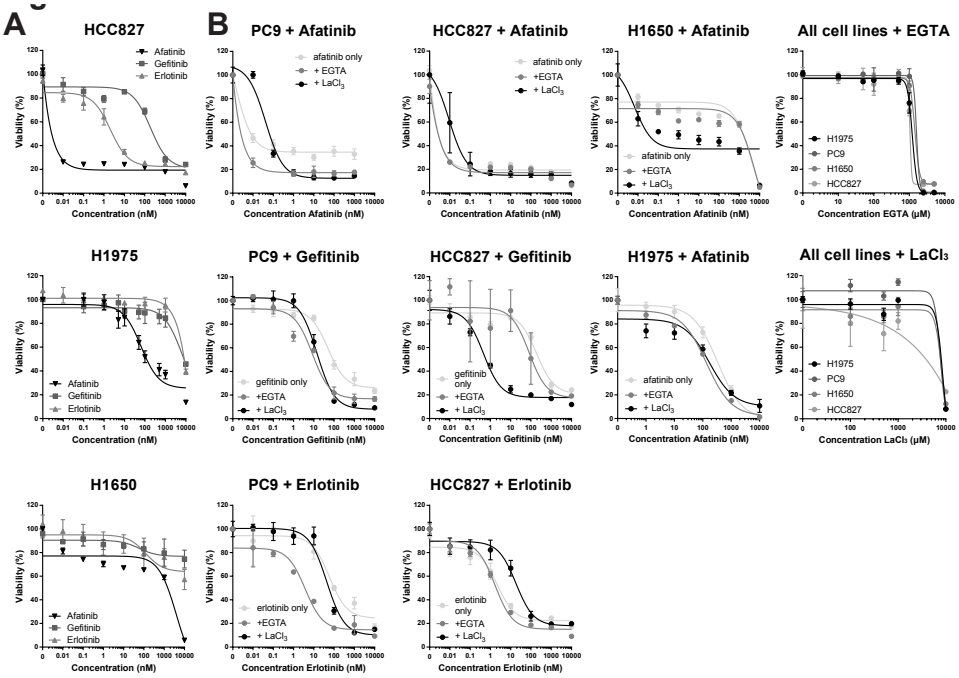
Supplementary Figure S1 – Representative light microscopy pictures of apoptosis and migration

A: Measurements of apoptosis was performed in the IncuCyte system, green fluorescence is indicative of an apoptotic cell. Light microscopic pictures, representing all replicates, show PC9 cells after 3 or 6 days treated with DMSO, growing to full confluence, 1 µM Afatinib, showing few apoptotic cells, or 10 µM afatinib, showing 100% apoptosis. **B.** Decreased migration upon EGFR inhibition was quantified by a scratch assay. Representative light microscopy pictures of DMSO, afatinib or gefitinib treated cells are shown 0, 24 and 48 hour after making the scratch. The same position in the well was depicted for the three time points.



Supplementary Figure S2 – Clear separation of conditions in each MS data set as shown by PCA analysis

Principle component analyses of the different datasets showed clear separation between conditions. **A:** PC9 cells were treated with 10nM Afatinib for 1, 3, 5 or 7 days for full proteome analysis. PCA showed separation of ratios of time points versus DMSO. **B:** PC9 cells were treated with 10 nM afatinib for 1, 2, 3 or 7 days for kinobead pull down. The PCA shows the separation of LFQ intensities of the different time points. **C**: PC9 cells were treated with DMSO or 10 nM afatinib for 30 minutes, 2 or 5 days before phospho-peptide enrichment. The PCA shows separation of LFQ intensities of the different time points. **D**: From the kinobead pull down data, detailed $\log_2$ ratios of kinases of interest are depicted in a heat map to show their regulation as compared to untreated cells. Color coding is based on the mean of the three replicate MS runs.



Supplementary Figure S3 – Behavior of four NSCLC cell lines treated by combined EGFR inhibition and Ca^{2+} deprivation

A: Viability assays are shown of three cell lines used in this study: HCC827, H1975 and H1650, to show the difference in sensitivity to EGFR inhibitors (afatinib, gefitinib and erlotinib). **B**: Viability assays show decreased sensitivity of four different cell lines used in this study to the combination of either EGTA or $LaCl_3$ with Afatinib, Gefitinib or Erlotinib. PC9 and HCC827 are shown for Afatinib, Gefitinib and Erlotinib. H1650 and H1975 are only shown for Afatinib treatment, since they are resistant to Gefitinib or Erlotinib treatment. All cell lines are shown to be insensitive up to 1 mM EGTA or $LaCl_3$ treatment. IC50 curves were made after 3 days of the start of an experiment. All error bars represent the data of four replicate wells within one experiment.

Materials and Methods

Cells and Compounds
Inhibitors Afatinib, Gefitinib, Erlotinib, Selumetinib, Sapanisertib and Rapamycin were purchased from Selleckchem. EGTA and Lanthanum(III) Chloride hydrate were purchased from Sigma-Aldrich. Cell lines NCI-H1975, NCI-H1650, HCC827 were derived from ATCC and PC9 was derived from Sigma. EGFR-pTyr1068, Paxillin-pTyr31 and Actin Antibody were from Abcam. EGFR, Myosin-IIa, Myosin-IIa-pSer1943, Myosin-IIc, Erk1/2, ERK1/2-pThr208/pTyr206, RPS6, RPS6-p-Ser235/Ser236, MAPK14, Paxillin, Paxillin-pTyr118, FAK, FAK-pTyr576/577 and FAK-pTyr397 antibodies were from Cell Signaling. GAPDH was derived from Genetex.

IncuCyte, Scratch Assay, Apoptosis and IC50 curves
Cells were plated 1500 cells (PC9/H1975) or 2500 (H1650/HCC827) per 96-well. An IncuCyte System (Essen Bioscience) was used to monitor cell growth. For the wound scratch assay, scratches were made using the IncuCyte wound maker. Apoptosis was measured using the IncuCyte Caspase-3/7 Apoptosis Assay (Essen Bioscience), according to manufacturer's protocol. Viability was measured using AlamarBlue® Cell Viability Assay (ThermoFisher Scientific), and IC50 curves were fitted using a non-linear fit curve fitting in GraphPad Prism 7.

Western Blot
20-30 µg cell lysates were run on a precast gel (Biorad, 12% SDS). Proteins were transferred to a PVDF membrane in transfer buffer (1x TG buffer (Bio-Rad), 20% methanol). Membranes were blocked with 5% milk or BSA before incubation with primary antibody overnight at 4°C. Membranes were incubated with secondary antibodies (Dako) for 2 hours at room temperature and bands were made visible by Pierce ECL Plus Substrate (Thermo).

Cell Culture
Cells were cultured in a humidified atmosphere and 5% CO_2. Growth medium was RPMI 1640 medium (Lonza), supplemented with 10% Fetal Bovine Serum (Thermo), 1% Pen/Strep (Lonza) and L-glutamine (Lonza). Cells were tested negative for mycoplasma infections on a regular basis.
For proteomics, proteins were labelled with heavy amino acids Arginine-6 and Lysine-4 in DMEM depleted of light amino acids (Buchem BV). Cells were cultured for a minimum of 5 passages to achieve complete labelling.

Cell Lysis and digestion
Before lysis, all cells were washed with cooled PBS (Lonza). For proteome analysis, cells were lysed in 1% SDC, 10 mM TCEP, 100 mM TRIS and 40 mM chloroacetamide (Sigma); protease inhibitors (Complete mini tablet, EDTA-free, Roche), phosphatase inhibitors (PhosSTOP tablet, Roche). Next, lysates were sonicated (15 cycles, 30sec on, 30sec off). For Phospho proteomics cell pellets were lysed in 7M urea, 1% triton, 100 mM TRIS, 5 mM TCEP, 30 mM CAA, 10 u/mL DNAse1, 1 mM Pervanadate, 1% Benzonase, 2.5 mM Mg^{2+}, phosphatase inhibitor, protease inhibitors. Next, proteins were precipitated using methanol/chloroform precipitation. Proteins were overnight digested with Trypsin (Sigma). Peptides were desalted using Seppack C18 column (1cc (50mg), Waters Corp.).
For Kinobead analysis, cells were lysed in lysis buffer containing 50 mM TRIS, 5% glycerol, 1.5 mM $MgCl_2$, 150 mM NaCl, 1 mM Na_3VO_4, 25 mM NaF, 0.8% NP40, protease inhibitors, cocktail of phosphatase inhibitors (see for details supplementary Materials and Methods).

High-pH fractionation, Fe^{3+}-IMAC, Kinobead analysis before LC-MS/MS Analysis
For proteome analysis, 100 µg desalted cell lysate was fractionated on a high-pH gradient. Peptides were loaded in buffer A (10mM NH_4OH, pH 10) on a Phenomenex Gemini C18 3µm 100x1.0 mm column using an Agilent 1200 HPLC system. A gradient of buffer B (10mM NH_4OH in 90% ACN, pH 10) was used for elution. Fractions were analysed separately [51].
For phospho peptide enrichment peptides were loaded on a the Fe^{3+}-IMAC column (Propac IMAC-10 4x50 mm column, Thermo) [16]. After wash with buffer A (30% ACN, 0.07% TFA) peptides were eluted with a gradient of buffer B (0.3% NH_4OH). Using a UV-abs signal the outlet of the column was monitored and one phospho fraction was collected. Phosphorylated peptides were dried in a speedvac and stored at -80°C.
For kinobead analysis 2 mg of cell lysate was enriched for kinases in a 96-well format using the kinobead assay as described previously [15]. After enrichment the eluates were digested using in-gel digestion.
Subsequently, peptides were analysed on an Agilent 1290 LC system (Agilent Technologies) coupled to an Orbitrap Q-Exactive HF (Thermo).

Data Analysis
RAW files were searched with MaxQuant. Further data processing was done using Perseus (v. 1.5.0.0), TIBCO Spotfire (v. 7.6.0.39), Metascape Excel Add-In (v 1.0.0.10) and PHOTON [17]. p-values of ratios were calculated per time point using student un-equal variance t-test for all data sets. In addition for proteomics and phospho proteomics data an ANOVA p-value was calculated. Significance was determined by correcting the p-value for multiple testing using a FDR of 5% (S0 set to 1, 250 randomizations).

Supplementary Figure and Material and Methods
Additional information on used methods, RAW files and supplementary tables and figures are provided online.

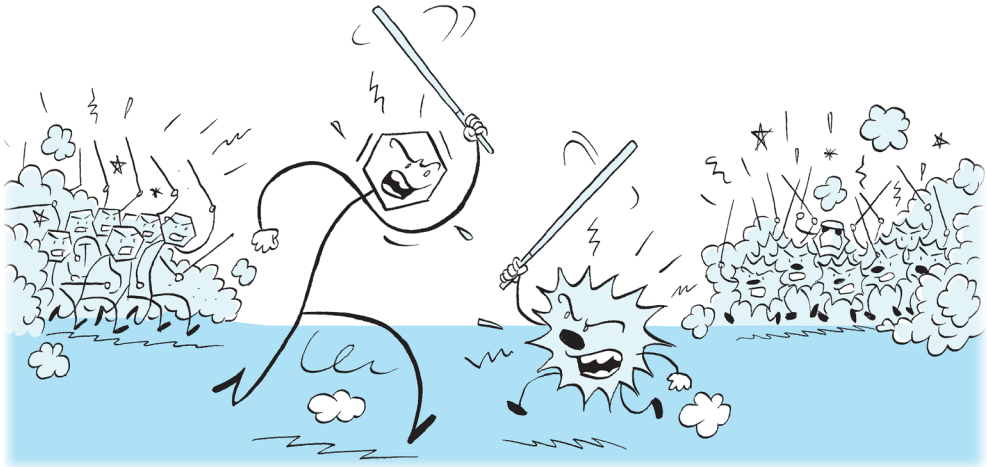
References

1. Minguet J, Smith KH, Bramlage CP: **Targeted therapies for treatment of renal cell carcinoma: Recent advances and future perspectives**. *Cancer Chemother Pharmacol* 2015, **76**:219–233.
2. Hirsh V: **Next-generation covalent irreversible kinase inhibitors in NSCLC: Focus on afatinib**. *BioDrugs* 2015, **29**:167–183.
3. Ramirez M, Rajaram S, Steininger RJ, Osipchuk D, Roth MA, Morinishi LS, Evans L, Ji W, Hsu C, Thurley K, et al.: **Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells**. *Nat Commun* 2016, doi:10.1038/ncomms10690.
4. Ma P, Fu Y, Chen M, Jing Y, Wu J, Li K, Shen Y, Gao J: **Adaptive and Acquired Resistance to EGFR Inhibitors Converge on the MAPK Pathway**. 2016, **6**:1232–1243.
5. Tricker EM, Xu C, Uddin S, Capelletti M, Ercan D, Ogino A, Pratilas CA, Rosen N, Gray NS, Wong K, et al.: **Combined EGFR / MEK Inhibition Prevents the Emergence of Resistance in EGFR -Mutant Lung Cancer**. *Cancer Discov* 2015, **5**:960–971.
6. Burrell RA, Swanton C: **Tumour heterogeneity and the evolution of polyclonal drug resistance**. *Mol Oncol* 2014, **8**:1095–1111.
7. Burrell R a, McGranahan N, Bartek J, Swanton C: **The causes and consequences of genetic heterogeneity in cancer evolution.** *Nature* 2013, **501**:338–45.
8. Sharma S V., Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, McDermott U, Azizian N, Zou L, Fischbach MA, et al.: **A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations**. *Cell* 2010, **141**:69–80.
9. Dolly SO, Collins DC, Sundar R, Popat S, Yap TA: **Advances in the Development of Molecularly Targeted Agents in Non-Small-Cell Lung Cancer**. *Drugs* 2017, **77**:813–827.
10. Wen F, Li Q: **Treatment dilemmas of cetuximab combined with chemotherapy for metastatic colorectal cancer**. *World J Gastroenterol* 2016, **22**:5332–5341.
11. Ahn M-J, Sun J-M, Lee S-H, Ahn JS, Park K: **EGFR TKI combination with immunotherapy in non-small cell lung cancer**. *Expert Opin Drug Saf* 2017, **16**:465–469.
12. Reck M, Heigener DF, Mok T, Soria J-C, Rabe KF: **Management of non-small-cell lung cancer: recent developments**. *Lancet* 2013, **382**:709–719.
13. Hirsch FR, Varella-Garcia M, Bunn PA, Di Maria M V., Veve R, Bremnes RM, Barón AE, Zeng C, Franklin WA: **Epidermal growth factor receptor in non-small-cell lung carcinomas: Correlation between gene copy number and protein expression and impact on prognosis**. *J Clin Oncol* 2003, **21**:3798–3807.
14. Rusch V, Baselga J, Cordon-Cardo C, Orazem J, Zaman M, Hoda S, McIntosh J, Kurie J, Dmitrovsky E: **Differential expression of the epidermal growth factor receptor and its ligands in primary non-small cell lung cancers and adjacent benign lung**. *Cancer Res* 1993, **53**:2379–2385.
15. Ruprecht B, Zecha J, Heinzlmeir S, Médard G, Lemeer S, Kuster B: **Evaluation of Kinase Activity Profiling Using Chemical Proteomics**. *ACS Chem Biol* 2015, **10**:2743–2752.
16. Ruprecht B, Koch H, Medard G, Mundt M, Kuster B, Lemeer S: **Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns**. *Mol Cell Proteomics* 2015, **14**:205–215.
17. Rudolph JD, Graauw M De, Water B Van De, Geiger T: **Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Report Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Networks**. *Cell Syst* 2016, **3**:585–593.
18. Fei S-J, Zhang X-C, Dong S, Cheng H, Zhang Y-F, Huang L, Zhou H-Y, Xie Z, Chen Z-H, Wu Y-L: **Targeting mTOR to overcome epidermal growth factor receptor tyrosine kinase inhibitor resistance in non-small cell lung cancer cells.** *PLoS One* 2013, **8**:e69104.

19. Shetty PK, Thamake SI, Biswas S, Johansson SL, Vishwanatha JK: **Reciprocal Regulation of Annexin A2 and EGFR with Her-2 in Her-2 Negative and Herceptin-Resistant Breast Cancer**. *PLoS One* 2012, **7**.
20. Yoshida T, Zhang G, Smith MA, Lopez AS, Bai Y, Li J, Fang B, Koomen J, Rawal B, Fisher KJ, et al.: **Tyrosine Phosphoproteomics Identi fi es Both Codrivers and Cotargeting Strategies for T790M-Related EGFR-TKI Resistance in Non – Small Cell Lung Cancer**. *Clin Cancer Res* 2014, doi:10.1158/1078-0432.CCR-13-1559.
21. Storr SJ, Carragher NO, Frame MC, Parr T, Martin SG: **The calpain system and cancer.** *Nat Rev Cancer* 2011, **11**:364–374.
22. James P, Vorherr T, Krebs J, Morellij A, Castello G, Mccormickn DJ, Pennistonll JT, Flora A De: **Modulation of Erythrocyte Ca2+-ATPase by selective Calpain Cleavage of the Calmodulin-binding Domain**. *J Biol Chem* 1989, **264**:8289–8296.
23. Alli AA, Bao HF, Liu BC, Yu L, Aldrugh S, Montgomery DS, Ma HP, Eaton DC: **Calmodulin and CaMKII modulate ENaC activity by regulating the association of MARCKS and the cytoskeleton with the apical membrane**. *Am J Physiol Ren Physiol* 2015, **309**:F456–F463.
24. Liu R, Jin JP: **Calponin isoforms CNN1, CNN2 and CNN3: Regulators for actin cytoskeleton functions in smooth muscle and non-muscle cells**. *Gene* 2016, **585**:143–153.
25. Kania A, Klein R: **Mechanisms of ephrin – Eph signalling in development , physiology and disease**. *Nat Rev Mol Cell Biol* 2016, **17**:240–56.
26. Saintigny P, Peng S, Zhang L, Sen B, Wistuba II, Lippman SM, Girard L, Minna JD, Heymach J V, Johnson FM: **Global Evaluation of Eph Receptors and Ephrins in Lung Adenocarcinomas Identifies EphA4 as an Inhibitor of Cell Migration and Invasion**. *Mol Cancer Ther* 2012, **11**:2021–2033.
27. Koch H, Del Castillo Busto ME, Kramer K, Medard G, Kuster B: **Chemical Proteomics Uncovers EPHA2 as a Mechanism of Acquired Resistance to Small Molecule EGFR Kinase Inhibition**. *J Proteome Res* 2015, **1**:2617–2625.
28. Eguchi R, Kubo S, Takeda H, Ohta T, Tabata C, Ogawa H, Nakano T, Fujimori Y: **Deficiency of Fyn protein is prerequisite for apoptosis induced by Src family kinase inhibitors in human mesothelioma cells**. 2012, **33**:969–975.
29. Cheng H, Liu P, Wang ZC, Zou L, Santiago S, Garbitt V, Gjoerup O V, Iglehart JD, Miron A, Richardson AL, et al.: **SIK1 Couples LKB1 to p53-Dependent Anoikis and Suppresses Metastasis**. *Sci Signal* 2009, **2**:1–12.
30. Villa-Moruzzi E: **Targeting of FAK Ser910 by ERK5 and PP1delta in non-stimulated and phorbol ester-stimulated cells.** *Biochem J* 2007, **408**:7–18.
31. Jacamo R, Jiang X, Lunn JA, Rozengurt E: **FAK phosphorylation at Ser-843 inhibits Tyr-397 phosphorylation, cell spreading and migration.** *J Cell Physiol* 2007, **210**:436–444.
32. Mitra SK, Hanson DA, Schlaepfer DD: **Focal adhesion kinase: in command and control of cell motility**. *Nat Rev Mol cell Biol* 2005, **6**:56–68.
33. Vicente-Manzanares M, Ma X, Adelstein RS, Horwitz AR: **Non-muscle myosin II takes centre stage in cell adhesion and migration.** *Nat Rev Mol Cell Biol* 2009, **10**:778–90.
34. Leung WKC, Ching AKK, Wong N: **Phosphorylation of Caldesmon by PFTAIRE1 kinase promotes actin binding and formation of stress fibers**. *Mol Cell Biochem* 2011, **350**:201–206.
35. Jiang Q, Huang R, Cai S, Wang C-LA: **Caldesmon regulates the motility of vascular smooth muscle cells by modulating the actin cytoskeleton stability.** *J Bomedical Sci* 2010, **17**:6.
36. Zhang R, Liu F, Hirano K, Ito M, Epstein HF, Y DJH: **Myotonic dystrophy protein kinase phosphorylates the myosin phosphatase targeting subunit and inhibits myosin phosphatase activity**. 2001, **493**:80–84.
37. Zhong X, Rescorla FJ: **Cell surface adhesion molecules and adhesion-initiated signaling: Understanding of anoikis resistance mechanisms and therapeutic opportunities**. *Cell Signal* 2012, **24**:393–401.
38. Reddig PJ, Juliano RL: **Clinging to life: Cell to matrix adhesion and cell survival**. *Cancer Metastasis Rev* 2005, **24**:425–439.
39. Tsai F-C, Kuo G-H, Chang S-W, Tsai P-J: **Ca2+ signaling in cytoskeletal reorganization, cell migration, and cancer metastasis**. *Biomed Res Int* 2015, **2015**:409245.
40. Hsu M, Tseng C, Chen F, Hsu JT: **Non-small-cell lung cancer cells combat epidermal growth factor receptor tyrosine kinase inhibition through immediate adhesion-related responses**. *Onco Targets Ther* 2016, **9**:2961–2973.
41. Cargnello M, Roux PP: **Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases.** *Microbiol Mol Biol Rev* 2011, **75**:50–83.
42. Schwartz M a, Assoian RK: **Integrins and cell proliferation: regulation of cyclin-dependent kinases via cytoplasmic signaling pathways.** *J Cell Sci* 2001, **114**:2553–2560.
43. Suda K, Tomizawa K, Dds MF, Murakami H, Osada H, Maehara Y, Yatabe Y, Sekido Y, Mitsudomi T: **Epithelial to Mesenchymal Transition in an Epidermal Growth Factor Receptor -Mutant Lung Cancer Cell Line with Acquired Resistance to Erlotinib**. *J Thorac Oncol* 2011, **6**:1152–1161.
44. Hazlehurst LA, Dalton WS: **Mechanisms associated with cell adhesion mediated drug resistance (CAM-DR) in hematopoietic malignancies**. *Cancer Metastasis Rev* 2001, **20**:43–

50.
45. Ding L, Shen Y, Ni J, Ou Y, Ou Y, Liu H: **EphA4 promotes cell proliferation and cell adhesion – mediated drug resistance via the AKT pathway in multiple myeloma**. *Tumor Biol* 2017, **I**–**II**.
46. Giannone G, Rondé P, Gaire M, Beaudouin J, Haiech J, Ellenberg J, Takeda K: **Calcium rises locally trigger focal adhesion disassembly and enhance residency of focal adhesion kinase at focal adhesions**. *J Biol Chem* 2004, **279**:28715–28723.
47. Tsai F-C, Seki A, Yang HW, Hayer A, Carrasco S, Malmersjö S, Meyer T: **A polarized Ca2+, diacylglycerol and STIM1 signalling system regulates directed cell migration.** *Nat Cell Biol* 2014, **16**:133–44.
48. Clapham DE: **Calcium Signaling**. *Cell* 2007, **131**:1047–1058.
49. Bryant JA, Finn RS, Slamon DJ, Cloughesy TF, Charles AC: **EGF activates intracellular and intercellular calcium signaling by distinct pathways in tumor cells**. *Cancer Biol Ther* 2004, **3**:1243–1249.
50. Cui C, Merritt R, Fu L, Pan Z: **Targeting calcium signaling in cancer therapy**. *Acta Pharm Sin B* 2017, **7**:3–17.
51. Wang Y, Yang F, Gritsenko MA, Wang Y, Clauss T, Liu T, Shen Y, Monroe ME, Lopez-Ferrer D, Reno T, et al.: **Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells**. *Proteomics* 2011, **11**:2019–2026.

FIVE



CHAPTER SIX

Early induced resistance against targeted therapy shows extensive homogeneity on (phospho) proteome level

Celine Mulder[1], Simone Lemeer[1]

[1] *Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands*

Introduction

Numerous targeted therapies for the treatment of cancer have been FDA approved and many more are currently in clinical trials [1]. Despite the initial clinical successes, the development of resistance remains one of the biggest challenges in cancer therapy [2]. Extensive research on resistance has been carried out [3–6], leading to the implementation of additional and/or combinatorial treatments and the development of new drugs, targeting the resistance driver [7–9]. For long, resistance has been described as intrinsic or acquired, describing pre-existing or newly acquired resistance mechanisms, with the main focus on genomic alterations [10]. However, the continues research in this field has led to a better understanding of this process and it is now widely accepted that the development of resistance is a dynamic process with heterogeneous kinetics, in which it becomes more apparent that a various amount of other cellular processes beyond genomic alterations can have equally harmful effects [11]. Building on this knowledge, several studies describe a drug persisting state during early treatment, where the cells show a slow growing phenotype without any genetic cause, with a subsequent escape into a more proliferative state [10,12]. Some describe this as a form of intrinsic resistance, where slow growing cells escape the drug treatment, as those drugs are often mainly designed to target proliferating cells [13]. Others describe it as an acquired resistance mechanisms which is induced by drug treatment [14]. Terms including adaptive resistance, cellular plasticity or phenotypic changes are differentially used throughout literature to describe both the short-term persisting state as well as a long-term and increased proliferative state.

This great variety of resistance mechanisms that can evolve upon treatment, even apart from inter tumour and patient heterogeneity, hampers the success of cancer therapy [15]. In order to design the most optimal cancer therapy, it is necessary to study and understand not only all possible end points of resistance and possible oncogenic mutations, but also the entire process of cellular adaptation and development of resistance. An often used method of studying resistance *in vitro* is to expose a sensitive cell line to a drug of choice for an extensive amount of time, subsequently inducing drug resistance [16,17]. Unlike an artificially introduction of a genetic mutation *in vitro* in order to induce resistance [18], prolonged drug exposure pushes cells into resistance in a more physiological manner. Although this method is often used to find novel oncogenic mutations, it is additionally well suited to study both the time line of adaptive resistance as well as the amount of heterogeneity within the produced resistant cell lines.

In this study, the non-small cell lung cancer (NSCLC) cell line PC9 was used, which is driven by a constitutively active epidermal growth factor receptor (EGFR) due to the DelE746A750 deletion. This cell line is sensi-

tive for EGFR inhibition by small molecule treatment, including the first and second generation tyrosine kinase inhibitors (TKI) gefitinib and afatinib, respectively. Sensitive cells were treated with both drugs separately in order to induce resistance. After the desired decrease in drug sensitivity was reached, different clones were produced by clonal expansion in order to reduce heterogeneity within each cell line [19]. Doing so, if the drug exposure induced multiple resistance mechanisms, the different phenotypes would be separated enabling them to be studied individually. Furthermore, the parallel study of multiple clones makes it possible to determine drug and cell line specific resistance characteristics. In total, a panel of 14 resistant clones were produced and analysed by a shotgun proteomics workflow. Shotgun mass spectrometry provides a discovery based method to study the molecular characteristics of resistant cells, which makes it possible to delineate all adaptations and map the heterogeneity during adaptive resistance. Additionally, the use of an optimised sample preparation for the enrichment of phosphorylated peptides made it possible to map both the proteomic as well as the phosphoproteomic landscape. This comprehensive analysis permitted the portraying of different stages and outcomes of the development of drug induced resistance *in vitro*. Several aspects of the development of resistance were analysed in more detail, including heterogeneity between clones of similar or different treatment origin, (phospho)proteomic changes over time and drug specific resistance characteristics.

Results

Characterisation of afatinib and gefitinib resistant cell lines

Sensitive PC9 cells were exposed to 1 μM of either afatinib or gefitinib for several weeks to induce drug resistance (Figure 1A). After 16 weeks cells were considered resistant, based on proliferation speed in the presence of the drug (Figure 1B, C), and the heterogeneous pool of resistant cells were plated single cell for clonal outgrowth. In total, 14 clones were able to form (10 afatinib-resistant (AR) clones, 4 gefitinib-resistant (GR) clones). In addition to the 14 clones, the heterogeneous pools of gefitinib-resistant and afatinib-resistant cells of which the clones originated from were included in the analysis. All 16 cell lines were monitored for their speed of growth (Figure 1B, C). In addition, the morphology of all cell lines was profiled by light microscopy images of both low and high confluence plates (Supplementary Figure S1).

Prior to further analysis, the expression levels and activity of EGFR and ERK were probed by western blot in 48 hour treated parental cells and a selection of clones (Supplementary Figure S2A). As shown by the decrease of phosphorylation on activity inducing site Tyr1068, the EGFR was inhibited in both 48 hours treated cells and afatinib or gefitinib resistant clones. However, MAKP1/3-pY187/202 phosphorylation increased in

both afatinib and gefitinib clones, indicating the induction of alternative growth stimulating pathways, besides EGFR activation in these resistant clones. These results were in excellent correlation with the proteome and phosphoproteome data as described below (Supplementary Figure S2B). Indeed, EGFR protein levels were shown to be upregulated, whereas MAPK1/3 expression levels did not change. Additionally, the phosphoproteome data confirms the rapid and persistent phosphorylation of MAPK1-pY187 and MAPK3-pY202 (Supplementary Figure S2B).

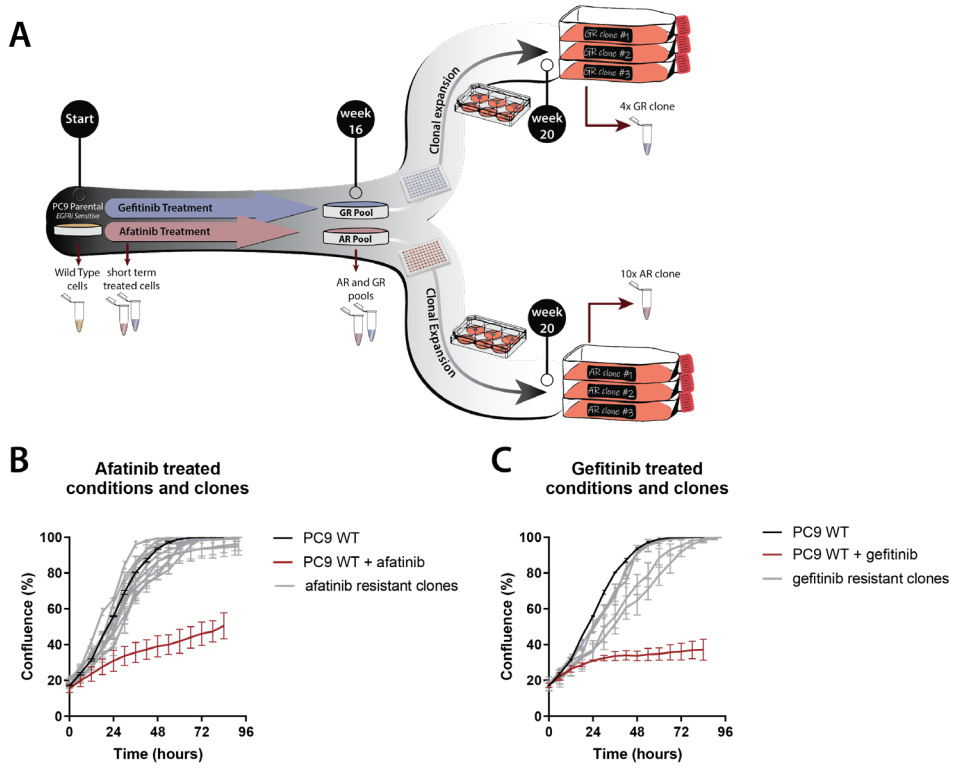


Figure 1 – Prolonged drug exposure induces resistance against EGFR inhibition
A: In order to induce drug resistance, PC9 parental cells, sensitive to EGFR inhibition, were treated with either afatinib or gefitinib. After 16 weeks, cells were found to be resistant (afatinib resistant pool (AR pool), gefitinib resistant pool (GR pool)). Both pools were plated out at the single cell level, and resistant clones were allowed to grow out in the presence of the drug. After 4 additional weeks, 4 gefitinib resistant and 10 afatinib resistant clones had grown out. In addition to the WT, AR/GR pools and the 14 clones, WT cells treated for 2 days with either drug were taken along in the analysis, to discover early drug responses and map TKI resistance development in time. **B-C**: Of all clones, the speed of proliferation was monitored in the IncuCyte to demonstrate resistance. The confluence was plotted over time as a measure of proliferation.

(Phospho)proteome analysis of resistant cell lines

Both afatinib and gefitinib target mutated Epidermal Growth Factor Receptor (EGFR), which is an important kinase upstream of several signalling pathways involved in growth and survival. In order to study altered signalling events, upon EGFR inhibition, a phosphopeptide enrichment [20,21] was performed prior to MS analysis enabling the study of both proteome as well as phosphoproteome during the course of resistance development. Besides the resistant clones, the analysis included three biological replicates of sensitive PC9 wild type cells as a control (from here on: WT), and WT PC9 cells treated for a limited time period (2 days) with afatinib or gefitinib to study short-term drug effects.

The cumulative dataset comprised quantification data for >4,100 proteins and >12,000 class I (>0.75 localisation probability) phosphosites, belonging to >3,100 phosphoproteins (Supplementary Tables 1, 2). As a start, for both proteome as well as phosphoproteome the Pearson correlations between each replicate were mapped (Figure 2A). In general, the patterns of proteome and phosphoproteome data were comparable. Furthermore, although all samples showed large correlations, with the smallest correlation being 0.9 in the proteome dataset and 0.7 in the phosphoproteome dataset, clear differences could be observed between clones. First, the lowest correlation could be observed between the two heterogeneous population of cells originating from either afatinib or gefitinib drug treatment. In contrast, both afatinib or gefitinib short-term treated cells showed a relatively high correlation to each other, suggesting similar short-term effects of both drugs on the cells, but a segregation of resistance mechanisms on the longer term. Second, both heterogeneous AR and GR pools did not show a high correlation to their respective clones, despite the fact that the pools are the clone's ancestors. Lastly, within the group of GR clones the correlation is high, indicating only minor differences between the clones. In contrast, within the group of AR clones more heterogeneity exists, which separates the clones in three distinct groups comprised of the clones 1C5, 2F1, 3D5, 3E7, 1B7, 1F9 and 1H2 as the first group, clones 2B7, 3E2 as the second group, and 2F6 as a third group.

Next, principle component analysis (PCA) showed the behaviour of the different clones and conditions as compared to one another (Figure 2B). Interestingly, the separation in the first component, indicating the main variation, is caused by differences in sensitive to resistant cell lines, whereas the separation on the second component describes the distinction between the drug treatments. This suggest that, even though treatment type can be distinguished, a certain homogeneity exist between all resistant clones, since they are more different from the untreated parental cells than they are from each other. Second, although according to the Pearson correlation the short-term treated cells showed great similarity, they were separated on the second component in the PCA, indicating some drug specific effects which were already visible after only

48 hours of treatment. Furthermore, the afatinib and gefitinib induced clones were closest to the short-term treated cells of the same drug. This suggests that the differences that exist between afatinib and gefitinib induced clones indeed already exist after short-term drug treatment, and can be measured after only 48 hours of treatment. Lastly, within the group of afatinib induced clones the PCA analysis distinguished 3 distinct groups which are the same as derived from the Pearson correlation plot (Figure 2A). For the group of gefitinib induced clones, PCA distinguished 3 clusters of 1G2 and 4E8 together, the GR pool and 2E8 together, and 4E7 which lie in between the two other clusters (Figure 2B). This is in contrast to the Pearson correlation plot, in which all clones except from the GR pool, showed a high correlation.

Resistance mechanisms could be detected at the phosphoproteome level after 2 days of treatment

The method of creating resistant clones with the here described approach is based on the assumption that the cells need several months of culturing in the presence of the drug to be able to develop resistance. Most likely, this is true for cells that acquire novel oncogenic mutations, since several cell divisions are necessary for the acquisition and positive selection of a new mutation. However, protein turnover, and even more so changes on the phosphoproteome level, can occur at a very short timescale. Therefore, induced resistance due to genomic alterations needs to be distinguished from adaptive resistance due to (phospho)proteomic alterations. In order to dissect the changes that could explain the differences in drug susceptibility, proteome and phosphoproteome data of all clones, AR and GR pools and afatinib and gefitinib short-term treated cells were compared to the untreated parental cells and statistical relevance was tested using a student's t-test. Cut-offs for determination of significance was set by a S0 curve with an FDR of 5%. After only 2 days of treatment, a minority of proteins was significantly changed (Figure 3A), however around 20% of all identified phosphosites were significantly regulated (Figure 3B), demonstrating a minimal change in protein turnover and/or expression, but a quick alteration at phosphoproteome level. After long term treatment and clonal expansion on average 20% of the proteins (Figure 3C) and 26% of all phosphosites (Figure 3D) identified and quantified in each clone was significantly different from untreated WT cells. Furthermore, proteins were in general found to be more upregulated than downregulated, whereas the phosphoproteome was equally up and down regulated. This can most likely be explained by the fact that a (de)phosphorylation could mean both activation as well as deactivation.

Next, it was reviewed if the changes on the phosphoproteome level that occurred after 2 days of treatment were related to regulated phosphorylation events that discriminate sensitive from resistant cell lines at a later stage. Therefore, we selected all phosphosites that were affected after

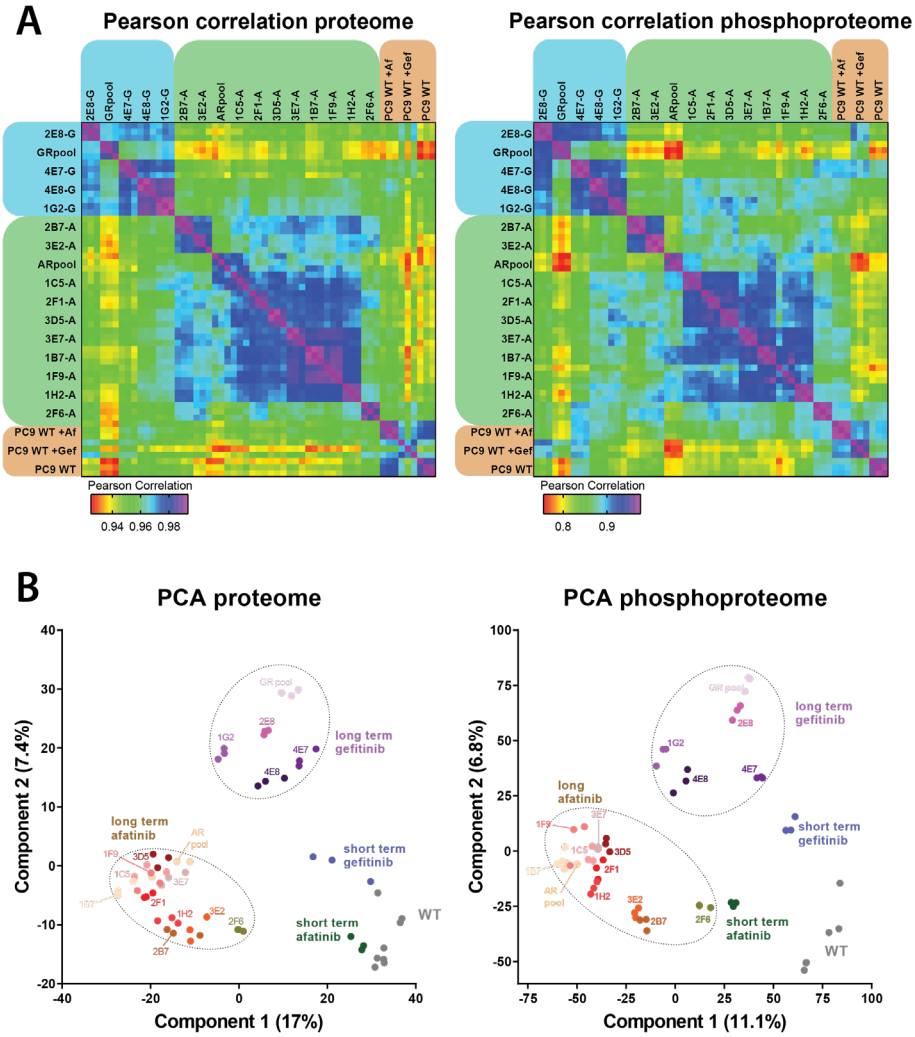


Figure 2 – Correlation and PCA of (phospho)proteome analysis of drug resistant clones
A: After MS analysis, the different clones and conditions were clustered and separated based on their Pearson correlations. The colour scale of the correlation is different in both plots (proteome: left, phosphoproteome: right) due to differences in the amount of variation between the samples. The colours beneath the column or row titles indicate the respective drug treatment or time scale (blue: gefitinib, long term; green: afatinib, long term; orange: Wild type, parental or short term treatment). **B**: An alternative approach of visualising variation or correlation between conditions/clones is a PCA plot. The X-axis represents the first component, describing the separation between sensitive and resistant cells. The Y-axis represents the second component, describing the different drug types (AR or GR clones are indicated with dotted circles). Colours depict different types of cells: untreated parental cells (grey), short term gefitinib treatment (blue), short term afatinib treatment (green), GR clones (different shades of purple), AR clones (different shades of red).

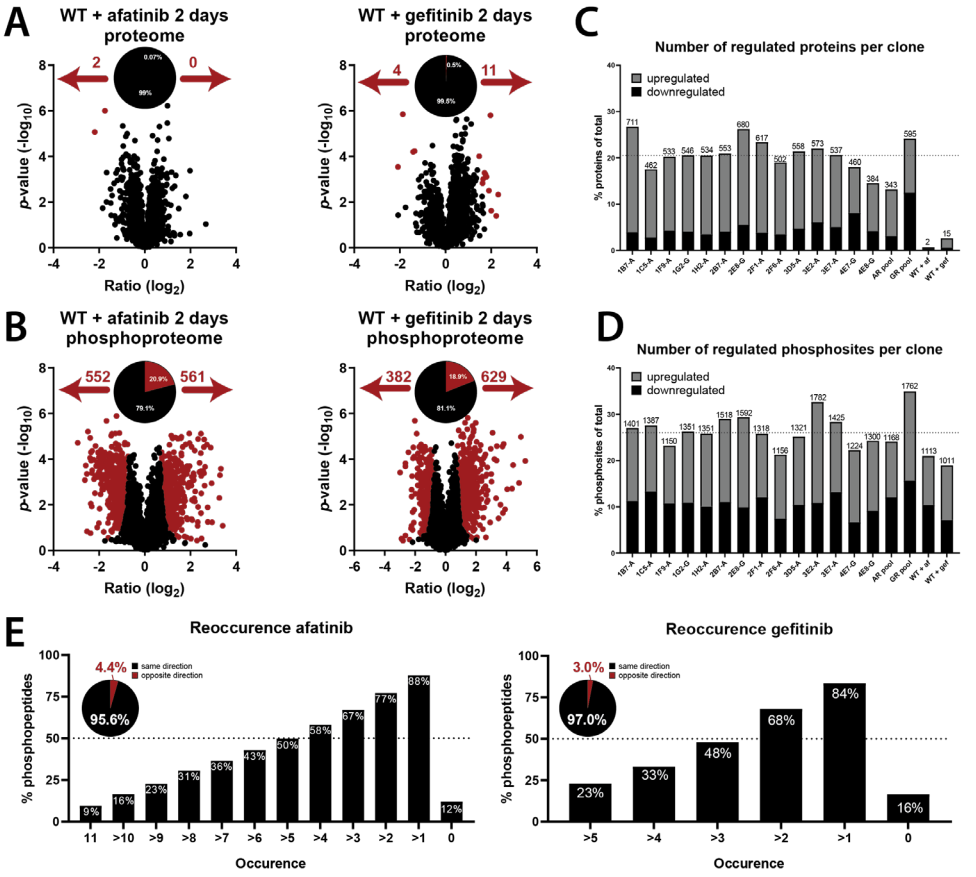


Figure 3 – Quantification of proteome and phosphoproteome
Each condition or cell line was compared to untreated WT cells. **A, B**: Volcano plots after short term treatment with afatinib (left) or gefitinib (right) showed very little changes on the proteome level (**A**) but large changes on the phosphoproteome level (**B**). On the X-axis the ratio between the treated cells and the untreated WT cells is shown, on the Y-axis the p-value ($-\log_{10}$) of the t-test between the samples is shown. Indicated in red are the significantly regulated proteins or phosphosites, according to a cut-off based on an FDR of 5%. The pie chart shows the distribution between regulated (red) and not regulated (black) proteins or phosphosites; arrows and numbers indicate the direction of the regulated proteins or sites. **C, D**: The amount of changed proteins (**C**) or phosphosites (**D**) are shown as bar plots. On the Y-axis the percentage of changed proteins/sites of all quantified proteins/sites per condition is shown. The bars are coloured for the direction of regulation: upregulated (grey) or downregulated (black). The number on top of each bar represents the number of changed proteins or sites. The dotted lines represent the average percentage of regulated proteins or phosphosites of all long term treated clones. **E**: All phosphosites that were affected after two days of treatment were analysed for reoccurrence as significantly regulated in a long term treated clone (afatinib: left, gefitinib: right). On the X-axis the number of reoccurrences of a phosphosite is shown, on the Y-axis the percentage of phosphosites of the total number of sites regulated after 2 days is shown. The dotted line indicates 50% of all sites. The inserted pie plots show the direction of regulation of sites that were regulated after two days and regulated in at least one long term treated clone. The percentage of sites regulated in the same direction (i.e. up- or downregulated in both short and long term) is shown in black, in the opposite direction in red (i.e. up after short term, downregulated after long term treatment or vice versa).

2 days of treatment and evaluated if these sites were still significantly affected in the resistant cell lines. Figure 3E shows the cumulative reoccurrence of a phosphosite which was affected after short-term treatment in one or more long term treated clones. In both treatments the overlap is extensive, since only a minority of phosphosites were no longer found in any of the clones (12% for afatinib and 16% for gefitinib treatment). Furthermore, more than 50% of all phosphosites were recurrently affected in >5 out of 11 (afatinib) or >2 out of 5 (gefitinib) clones. Moreover, if the phosphosites which were significantly regulated in both the short term treated cells and in at least one of the long term treated clones were inspected for the direction in which they were regulated, it appeared that 95,6% in afatinib treated clones, and 97% in gefitinib treated clones were changed in the same direction as after 2 days of treatment (Figure 3E, inserts). This indicates a large homogeneity between the clones and a large overlap between the long term treated clones and the short term treated cells. Therefore, we hypothesize that the proteomic changes which the long term treated clones adopted in order to become resistant were already present and could be measured after only 48 hours of drug treatment.

Large homogeneity exists between clones on both proteome and phosphoproteome level

In both the Pearson correlation and PCA plots, AR and GR clones could clearly be distinguished, indicating drug specific effects. Subsequently, calculated ratios of all conditions versus untreated WT cells were mapped by hierarchical clustering to identify differential regulated proteins and phosphosites between clones (Fig 4A). Similar to the PCA and correlation plots, the main visual clusters indicate the two groups of clones of similar origin (AR or GR clones). However, although the separate clones could be distinguished in the PCA and correlation plots, the hierarchical clustering showed negligible differences between the clones of similar drug treatment. The phosphosites inside the cluster which distinguished the two groups the most was investigated for enriched pathways to identify drug specific altered pathways. Pathway enrichment analysis showed the main differentiating factor between AR and GR clones to be RNA splicing (Fig 4A, bar plot).

In order to validate if the detected differences between the two drug treatments were also determining specific drug resistance, clones were treated with the opposite drug as to which they initially developed resistance to (i.e. AR clones with gefitinib, and GR clones with afatinib). All clones showed to be resistant for both afatinib and gefitinib, unrelated to the drug used for resistance induction (Fig 4B). This indicates that, apart

from minor differences between clones and a clear separation of origin from either AR pool or GR pool in clustering analyses, all clones adapted to either drug treatment with highly similar, or maybe even the same resistance mechanisms.

Generic affected pathways upon resistance development

In order to discover the most generic regulated pathways, shared by most clones, all proteins that were regulated in >13 clones were plotted in a heat map, to show their regulation in each clone compared to the untreated wild type cells (Supplementary Figure S3A). This analysis showed that these most robustly changing proteins across all clones are all regulated in the same direction (i.e. either up or down), with the exception of 4 proteins. All four, including two HLA genes and two metabolic enzymes, were upregulated in afatinib clones, but downregulated in gefitinib clones (Supplementary Figure S3B).

For the phosphoproteome data, similar to proteome data, all detected sites in at least 13 out of 16 clones were regulated in the same direction (i.e. up or down), showing a certain robustness in their (phospho)proteomic transformation. In this case, all sites were taken along in further analysis, for no distinction in up or down regulation was made (Supplementary Figure S3A).

In Supplementary Figure S4 STRING networks are shown of the commonly regulated (phospho)proteins indicated in supplementary figure S3A. Regarding the proteome (Supplementary figure S4A) 4 clusters of proteins could be identified, containing proteins related to adhesion and morphology (blue), metabolism (green) and DNA (orange) and RNA processing (yellow). Clusters of regulated phosphoproteins (Supplementary Figure S4B) showed great similarity with the proteome data, including regulation of proteins associated with transcription, DNA and RNA processing, protein folding and degradation, cell cycle, adhesion, metabolism and EGFR signalling. Considering that both afatinib as well as gefitinib inhibit the tyrosine kinase EGFR, it is not surprising to see large changes amongst phosphoproteins associated with this last pathway. The regulation at the cell cycle level are more difficult to explain, since the AR and GR clones are proliferating at a similar speed as the untreated parental cells (Figure 1B). This indicates, that despite EGFR inhibition, the clones have found a way to continue proliferation, but in order to do so, they might use different pathways compared to WT cells.

PHOTON analysis shows splicing to be the main discriminating factor between clones

Due to the diverse biological effects of phosphorylation on the function of proteins, interpretation of phosphorylation data often needs additional processing, for example by using the tool PHOTON [22]. This tool converts ratios of phosphorylation events on a peptide level into a functionality

score on a protein level. Additionally, this tool includes all phosphorylation events, in contrast to only the significantly changing events, thereby increasing the scope of the analysis. Figure 5 shows the hierarchical clustering of the data after PHOTON analysis. In contrast to the clustering based on significant ratios (Figure 4A), clones from different drug treatments did not all cluster together in this analysis. The main cause for this discrepancy is the cluster at the bottom (grey), consisting of proteins that are part of the spliceosome. This is in agreement with the hierarchical clustering of significantly regulated phosphosites, where the major differentiating proteins were associated with the spliceosome as well (Figure 4A). It has already been shown by others that regulation of the spliceosome and the following expression of splice variants can cause resistance or increased oncogenic circumstances [23]. What is interesting however is that the spliceosome is differentially regulated between different clones,

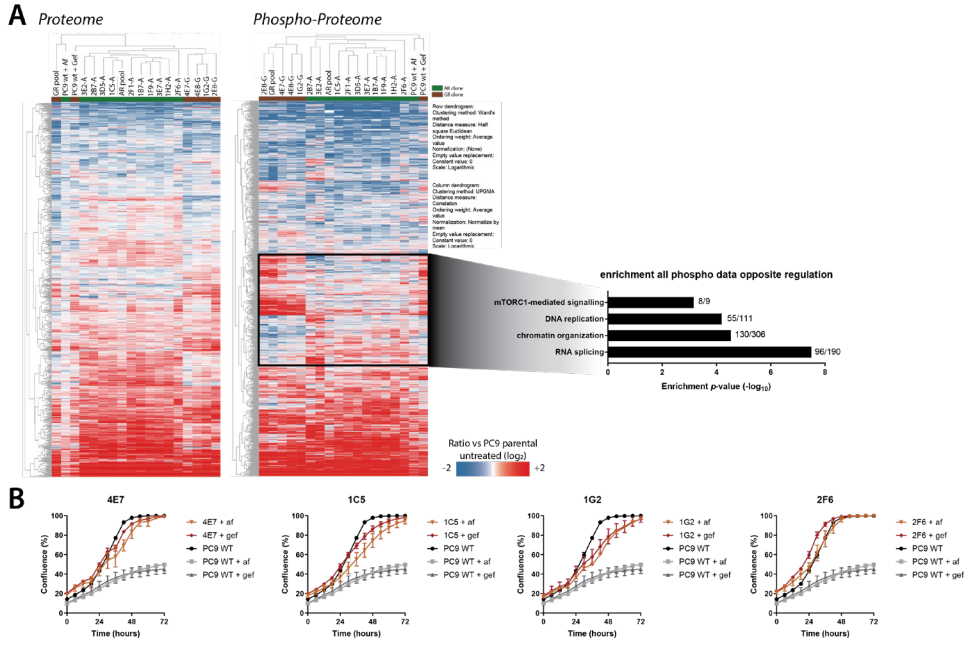


Figure 4 – A (phospho)proteomic signatures of resistant clones
A: Ratios of each clone or condition as compared to the untreated parental cells of both proteome (left) and phosphoproteome (right) data was clustered by hierarchical clustering. Dendograms on the top show a clear clustering of AR (indicated in green) of GR (indicated in brown) clones. The middle cluster in the phosphoproteomic data (indicated in the black box) was examined for GO term enrichments (bar plot). On the X-axis is the -$\log_{10}$ *p*-value of the enrichment, the numbers next to the bars are the amount of proteins found/the amount of proteins documented to belong to the specified pathway. **B**: Although some differences can be seen in the clustering, the general (phospho)proteomic signature was highly similar. This observation was confirmed by the cross-reactivity of selected clones against the alternative drug (i.e. AR clones for gefitinib, GR clones for afatinib resistance). On the Y-axis the percentage of confluence is shown as a measure for growth, on the X-axis the time of treatment is shown.

as shown by the PHOTON functionality score. In our data, the clones 2E8-GR, 4E7-GR, GRpool, 2F6-AR, 1F9-AR, 1B7-AR, 3E7-AR and 3E2-AR show an increased functionality of the spliceosome, where the clones 1C5-AR, ARpool, 2B7-AR, 1H2-AR, 2F1-AR and 3D5-AR show a decrease in functionality. The clones 4E8-GR and 1G2-GR show very little regulation. In addition, spliceosomal proteins are the most activated proteins after short term treatment for both drug conditions, indicating that the spliceosome is already affected after only a very short period of time. In addition, the cluster which consists of proteins associated with endocytosis (indicated in green), are differentially regulated between the different drug types. Mainly 1H2-AR, 2F1-AR, 3D5-AR, 4E7-GR, GRpool, 3E7-AR and 3E2-AR show increased functionality.

The other three highlighted clusters show a more distinct pattern between the two drug types. The top cluster (blue) describes the RNA polymerase, transcription related proteins, DNA repair and immune response related proteins. These proteins are decreased in functionality in the gefitinib treated clones. In contrast, the second cluster from the top (red) displays proteins that are increased in functionality in the gefitinib clones, describing vesicle transport, microtubule associated proteins and the ribosome. The last cluster (yellow) display proteins increased in AR clones and decreased in GR clones, containing proteins related to nucleic acid metabolic pathways, telomeres and chromatin organization.

Based on these results, the here presented data suggests that GR clones showed increased translation, whereas AR clones showed increased transcription and DNA repair. Furthermore, the regulation of the spliceosome was quick and abundant, and seemed to be independent of drug type.

Isoform specific peptides show differentially regulation due to alternative splicing

Alternative splicing of mRNAs before translation increases diversity of the proteome, and is used by the cell as an additional layer of complexity during transformation at the time of disease or resistance development [23]. Several mechanisms of splicing contribute to the production of different isoforms of the same gene transcript, retaining different cellular functions [23]. Shotgun proteomics is an excellent tool to study these regulations. The detection of opposing regulation of isoform specific peptides provides detailed insight in the regulation of different isoforms, in both full proteome as well as in phosphoproteome data.

Since the data presented here indicated regulation of the spliceosome, the data was inspected to find isoform specific peptides. In total in the proteome data, 215 isoform specific peptides were found originating from 51 proteins (counting different isoforms as one), for which at least 2 different isoforms were detected (Supplementary Table 1.3). Under the same criteria, in the phosphoproteome data, 106 isoform specific peptides, originating from 22 proteins were identified (Supplementary Table

2.3). For most peptides, no significant regulation could be detected (145 out of 215 on the proteome level, 60 out of 106 on the phosphopeptide level), or peptides from different isoforms showed identical trends indicating regulation of the protein or phosphosite independent of alternative splicing.
Figure 6 shows the alternative splicing, sequence and regulation of deoxyuridine 5-triphosphate nucleotidehydrolase (DUT). This protein is a UTPase, depleting the UTP pools in the cells by enzymatic conversion subsequently inhibiting DNA damage. Already in 1996 it was shown that this protein has two isoforms which have distinct localisations to either the mitochondria (DUT-M) or the nucleus (DUT-N) [24]. Furthermore, only the nuclear isoform has been shown to be phosphorylated, despite the presence of the homologue serine in both isoforms [25]. Although the function of this phosphosite is not known yet, it is considered a CDK1 target based on the consensus sequence and kinase directed phosphoproteomics, indicating a role of the DUT-N in the cell cycle [25,26]. Figure 6A and 6B show the alternative splicing of this protein and the subsequent amino acid alignment. Indicated in red and in blue (Figure 6B) are the peptides that were identified in this study. The abundance of DUT-M did not change significantly in any of the clones, whereas DUT-N, although not identified in all clones, showed increased abundance in 10 out of the 19 clones, including both short term treated conditions (Figure 6C). Furthermore, Ser11 on DUT-N was increasingly phosphorylated in 10 and downregulated in 2 out of 19 conditions (Figure 6D, top).
DUT-N has been suggested to be phosphorylated by CDK1 [25,26]. CDK1 expression did not change in any of the clones (data not shown). However, the known activating phosphorylation site of the kinase, Tyr15, was shown to be increased in all but 3 of the conditions (Figure 6D, bottom). Furthermore, the absence of increasing phosphorylation of CDK1 in 1F9 and the AR pool was accompanied with absence of regulation of phosphorylation of DUT-N (Figure 6D, dotted lined boxes). Also, the significant decrease of phosphorylation of CDK1 in the short term afatinib treated cells was accompanied by a decrease of phosphorylation of DUT-N in the same condition, despite increased abundance of this isoform.

Specific regulation of protein expression and phosphorylation status are indicative of individual clones

Although quite some homogeneity exists between the changes in proteome and phosphoproteome, some differences in regulation between the distinct clones could still be observed. In addition, several proteins showed quite an extreme regulation at the phosphorylation level. Below, 3 of these proteins were discussed in more detail.
High expression levels of the heat shock protein HSP27 (HSPB1) have been implicated in poor clinical outcome for cancer patients, increased metastatic potential and resistance to cancer therapy for a variety of can-

cers [27]. In our data, the expression levels of HSP27 seemed to vary over the different clones (Figure 7A). The phosphorylation status of HSP27 however showed some interesting trends. HSP27 Ser82 phosphorylation could only moderately be detected in WT cells (only a single detection in 9 measurements). After 2 days treatment however, Ser82 phosphorylation is well detected in both the afatinib as well as gefitinib condition (Figure 7A). Over time, phosphorylation levels continued to increase, except the

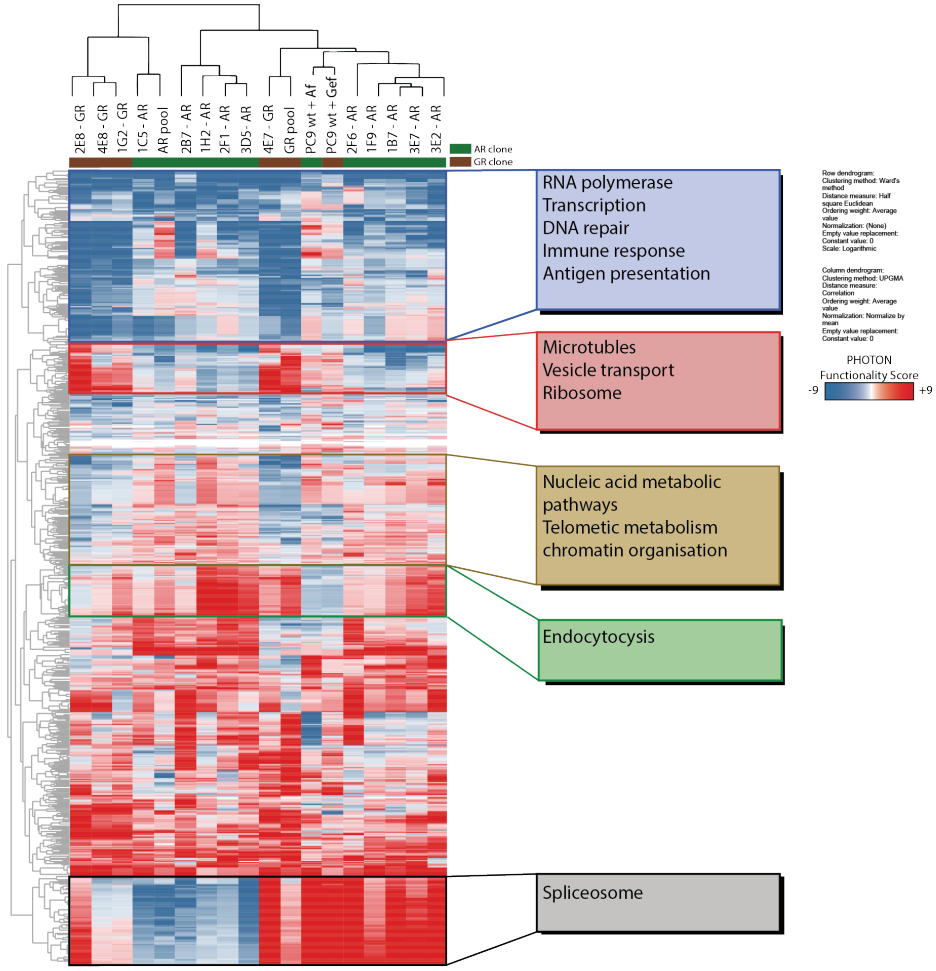


Figure 5 – PHOTON analysis of phosphoproteome data
Phosphoproteome data was processed with the PHOTON algorithm [22]. This tool transforms ratios per phosphosite into activity scores per phosphoprotein on a scale from -9 to 9, which can be clustered by hierarchical clustering. Several, with boxes indicated clusters of phosphoproteins were selected and manually annotated for associated pathways, based on information of GO terms, STIND-db, Uniprot and published literature. Interestingly, cell lines with similar drug origin did not cluster together. The main driver of the clustering was by differences in the activity of the spliceosome, indicated at the bottom cluster.

2F6 afatinib clone and the 4E7 gefitinib clone (Figure 7A). HSP27 Ser82 phosphorylation has been implicated in regulation of HSP27s oligomerization state, in which phosphorylation promotes oligomer dissociation, enhancing chaperone activity and anti-apoptotic signalling [27]. Another site on HSP27, Ser15, showed an opposite trend, as it was downregulated in most afatinib and gefitinib clones. Despite an important role for HSP27 in cancer, only limited studies have actually tried to target HSP27. Two inhibitors for HSP27 exist, one being the oligonucleotide Apatorsen (OGX-427) [28] the other being the Herpes Simplex drug Brivudine [29]. Lelj-Garolla et al. [30] observed erlotinib-resistance related to increased HSP27 levels in NSLC cell lines (HCC827 and A549) and presented preclinical evidence that inhibition of HSP27 using OGX-427 in combination with erlotinib treatment enhanced antitumor activity and delayed A549 xenograft growth in mice. Brivudine directly binds HSP27 and restores apoptosis by inhibiting the interaction of HSP27 with pro-apoptotic proteins. In future experiments it would be interesting to test the effect of HSP27 inhibition on the here presented clones and in the process of resistance development.

EPS8 has been implicated in cancer cell migration and epithelial to mesenchymal transition (EMT) before, and its expression levels has been correlated with chemotherapeutic response [31,32]. However, little is known about the role of EPS8 phosphorylation in these processes. In our data, EPS8 Ser685 and Ser476 phosphorylation decreased in all afatinib clones, whereas Ser659 phosphorylation increased in all clones compared to WT cells. Furthermore, EPS8 protein expression seemed to increase in all gefitinib clones. This indicates a differential response of EPS8 phosphorylation which correlates with the drug used. EPS8 is an adaptor protein which is directly involved in EGFR signalling through Rac, and trafficking through Rab, depending in its association partner [33]. As both drugs inhibit EGFR, but in a reversible (gefitinib) or irreversible (afatinib) manner, differences in direct downstream signalling, especially on adaptor proteins might be a reflection of drug action.

TRIM33 is an ubiquitin E3 ligase and has recently been identified as a tumour suppressor that can abolish tumour cell proliferation and tumorigenesis by degrading nuclear β-catenin [34]. In addition, loss of TRIM33 has been linked to resistance to BET inhibitors in cancer [35]. In our data, TRIM33 expression did not vary significantly in the clones compared to the WT. However, TRIM33 Ser1119 phosphorylation is upregulated in all gefitinib and afatinib clones, an effect that additionally already could be observed after only 48 hours of gefitinib treatment. Although no functional implications have yet been correlated with Ser1119 phosphorylation, the site represents a classical AKT1 phosphorylation site, bearing the RXRXXS motif. Future research should reveal whether TRIM33 phosphorylation status influences its role as a tumour suppressor.

Discussion

Several studies in the field of cancer research show that the development of drug resistance is a fluent and dynamic process, in which a variety of pathways are involved [11,12]. Unfortunately, the discovery of more and more resistance causing pathways and oncogenes has only further raised the awareness of the extremely heterogeneous nature of cancer. Furthermore, since resistance can arise from a multitude of origins, which all take a different amount of time to emerge, it is important to consider the timeframe of resistance development. An increasing amount of cancer research focusses on the initial adaptive, non-mutational response of cells upon drug treatment addressing both the varying time scale and the wide range of resistance mechanisms [11,12,36]. In this study, the adaptive response was studied on a proteome level, trying to map the early response of NSCLC cells upon EGFR inhibition, thereby revealing potential therapeutic targets and the level of heterogeneity during early adaptation.

The main finding of this study indicates that large homogeneity exists between treatment conditions at different time points and between the produced resistant clones. We showed that already after 48 hours of drug treatment a large set of phosphosites were differently regulated compared to untreated parental cells. Following weeks of treatment and further induction of resistance the same phosphosites remained affected. Therefore, we hypothesize that the development of resistance we observed after prolonged treatment could potentially already have been observed after only 2 days of treatment. Although further expansion of these cells might lead to additional mutations which cannot yet be predicted, this data indicates that the early, initial drug response is extremely quick and lasting.

Furthermore, the emergence of resistance in the 2 pools and subsequent 14 clones developed extremely homogenous. Despite the fact that the used drug, the exact speed of proliferation and morphology differed slightly between the clones, the proteome and phosphoproteome signatures were extremely similar. Even though this study was performed *in vitro*, this might be an indication that when the development of resistance is considered as a therapeutic target, and not the end result (i.e. an acquired mutation), a common treatment strategy could be beneficial; the earlier a resistance process can be targeted, the more homogenous the cell population will be. Further underlying the observation of early homogeneity is the lack of reactivity of GR clones to afatinib treatment. As a second generation EGFR inhibitor, afatinib was initially designed to overcome resistance against the first generation inhibitor gefitinib [37]. However, this is mainly specified to the acquisition of the T790M mutation of the EGFR. The results shown here indicate that gefitinib resistance due to other factors than the T790M mutation fail to be inhibited by afatinib

treatment.

Cancer has mainly been described in literature as a genetic disease, with the main cause of oncogenic phenotypes being chromosomal instability and subsequent mutation of key oncogenes [38–40]. However, an increasing amount of literature has become available describing oncogenic features of non-mutational cellular processes [11]. In the data described in this study, the analysis of the set of clones showed significant changes in protein abundances. In addition, a large portion of these regulated (phospho)proteins were involved in processes associated with transcription, translation and protein folding, underlying the observation of pro-

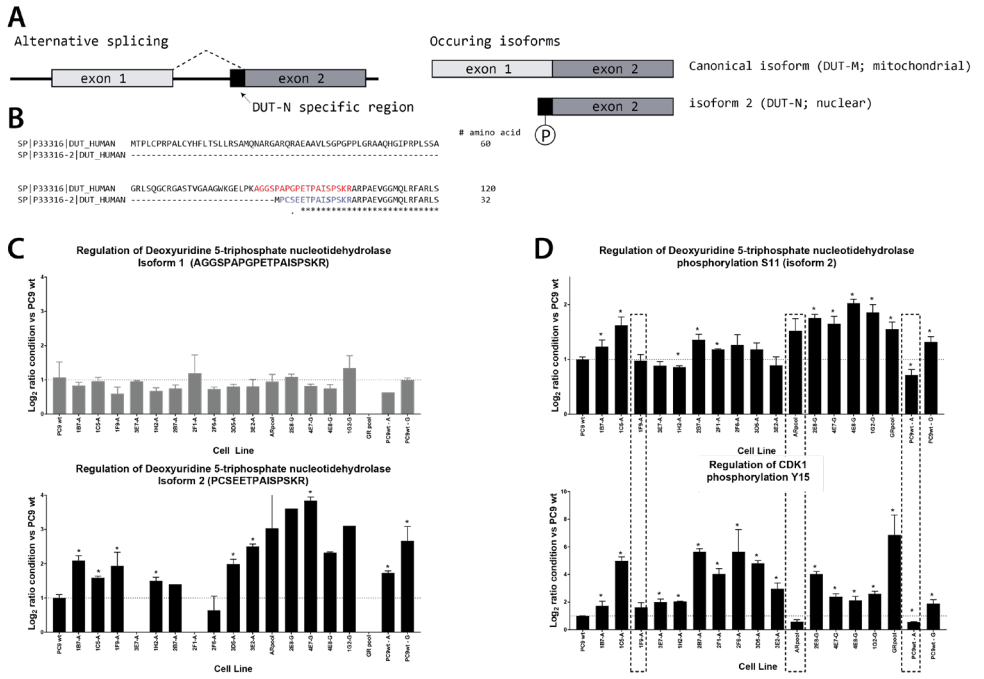


Figure 6 – Alternative splicing of DUT in a subset of resistant clones

A: Alternative splicing of the DUT gene result in the expression of two isoforms. The main biological difference between the different isoforms is the localisation, where DUT-M resides in the mitochondria and DUT-N in the nucleus. Furthermore, only DUT-N can be phosphorylated, as indicated with a -P. **B**: both isoforms were identified by MS analysis by the identification of an isoform specific peptide, indicated in the aligned sequence in red (DUT-M) or blue (DUT-N). Only the first part of the protein is shown. Annotations: "-" no amino acid; "." Different amino acid; "*" the same amino acid. **C**: Both isoform specific peptides were identified and quantified in the majority of the clones in the proteomic data. The isoform specific peptide of isoform 1 (DUT-M) showed no regulation in either of the clones (top). In contrast, the isoform specific peptide for isoform 2 (DUT-N) showed significant regulations in 9 conditions (bottom). **D**: It has been shown that only DUT-N can be phosphorylated on Ser11. Indeed, only a phosphopeptide of isoform 2 could be identified, showing significant regulation in 12 clones (top). This phosphosite is hypothesized to be phosphorylated by the kinase CDK1, of which the activating site (pTyr15) showed significantly upregulation in 15 conditions (bottom). Dotted boxes indicate three interesting clones, where the lack of DUT-N regulation matches the lack or downregulation of CDK1 activity. Of all bar graphs: on the Y-axis is shown the $\log_2$ ratio of the specified condition on the X-axis against the WT cells. Error bars indicate triplicates. *: p-value < 0.05.

teomic adaptation upon oncogenic transformation.
Proteomic changes that were detected to be altered in all resistant clones were found to be involved in metabolic pathways, endocytosis and cytoskeleton reorganisation. Despite the fact that cancer research is mainly focussed on genomic alterations to explain drug resistance, the processes identified in this study have indeed been shown before to enhance drug resistance and oncogenic properties. Firstly, a large group of metabolic enzymes were differentially regulated in all clones. Indeed, involvement of metabolic alterations enabling cell to resist drug treatment has been shown repeatedly [41–44]. Furthermore, the analysis of the phosphoproteome data revealed differentially regulation of endocytosis between the different clones. Endocytosis of transmembrane growth factor receptors is a mechanism of regulation inducing either amplification of the growth signal by altered localisation into endosomes or degradation of an activated receptor into the lysosomes via a negative feedback loop [45,46]. Lastly, a clear regulation was observed of adhesive structures and morphology. Additionally, the microscopic pictures of the changes in morphology also confirm large changes in cell polarity, adhesive structures and cell-matrix connections. We and others have shown before that early altered adhesion can enhance anti-apoptotic processes enabling drug resistance [47,48]. Moreover, further development of the metastatic properties of cells can lead to the extremely oncogenic process of epithelial-to-mesenchymal transition (EMT) [49]. Further validation experiments would be necessary to delineate the influence of these pathways in this particular disease model.
Despite the large homogeneity between clones, they could still be distinguished based on their proteomic signature. The major discriminating factor between clones was shown to be the regulation of splicing. Splicing is one of the mechanisms the cell uses to increase proteomic diversity by the production of protein isoforms executing different cellular functions [23]. Splicing has been shown to be heavily involved in oncogenic transformation, and several therapeutic strategies have been designed to target the spliceosome machinery in cancer [50]. Splicing in cancer can be studied by shotgun proteomics through the identification of isoform specific peptides and their changes in abundance or level of modification. The major advantage of the use of proteomics data over genomic or transcriptomics data is the evidence of the actual presence of the proteomic isoform, since it has been shown that the majority of alternative transcripts never produce a working protein [51]. However, the disadvantage is the necessity of large sequence coverage of the different isoforms in question and confident fragmentation of their peptides in order to identify and quantify multiple isoform by their specific peptides.
A protein found to be differentially expressed and phosphorylated as two isoforms was deoxyuridine 5-triphosphate nucleotidehydrolase (DUT). This protein is known to be expressed in the form of two isoforms, DUT-N

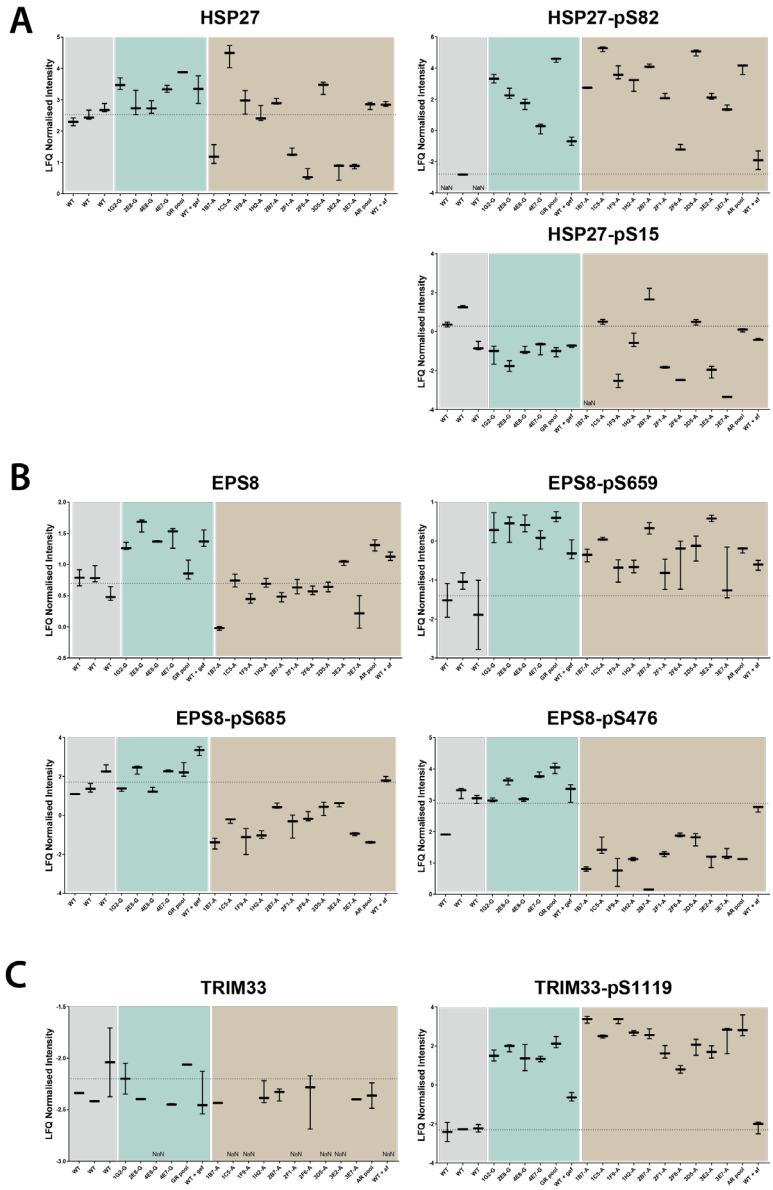


Figure 7 – Individual regulation of phosphorylation of HSP17, EPS8 and TRIM33
A few individual proteins were shown for detailed discussion. **A:** HSP27 protein expression, HSP17-pSer82 and HSP17-pSer15. **B**: EPS8 protein expression, EPS8-pSer659, EPS8-pSer685 and EPS8-pSer476. **C:** TRIM33 protein expression and TRIM33-pSer1119. Background of graphs are indicative of drug treatment: grey: wild type; green: gefitinib (both long and short term); brown: afatinib (both long and short term). Error bars represent replicates. Dotted horizontal line indicates the average LFQ intensity of the WT measurement. NaN: not a number.

and DUT-M, which are localised to either the nucleus or the mitochondria [24]. DUT-N has been assumed to be a target of the kinase CDK1 and presumably plays a role in the cell cycle [25,52]. In the presented data, cell cycle related proteins, among others CDK1, were demonstrated to be regulated, despite the similar speed of proliferation between the resistant clones and the parental cells. Both the proteome as well as phosphoproteome data showed regulation of DUT-N but not of DUT-M, combined with the significant regulation of the DUT-N specific phosphorylation, indicating a possible role of DUT-N in the regulation of the cell cycle in the resistant clones.

In conclusion, prolonged drug exposure *in vitro* can be used to study the initial and adaptive response in the development of drug resistance. In this study, we mapped the (phospho)proteome regulations upon EGFR inhibition, showing a large proteome transformation, independent of genomic alterations. Furthermore, a large homogeneity was observed in time and between clones, suggesting common mechanisms of drug resistance during the initial response. Whereas genomic heterogeneity is one of the most difficult obstacles to overcome in cancer therapy, detailed study of these homogeneous, non-mutational processes might lead to additional therapeutic windows and further developments in the field of cancer therapy.

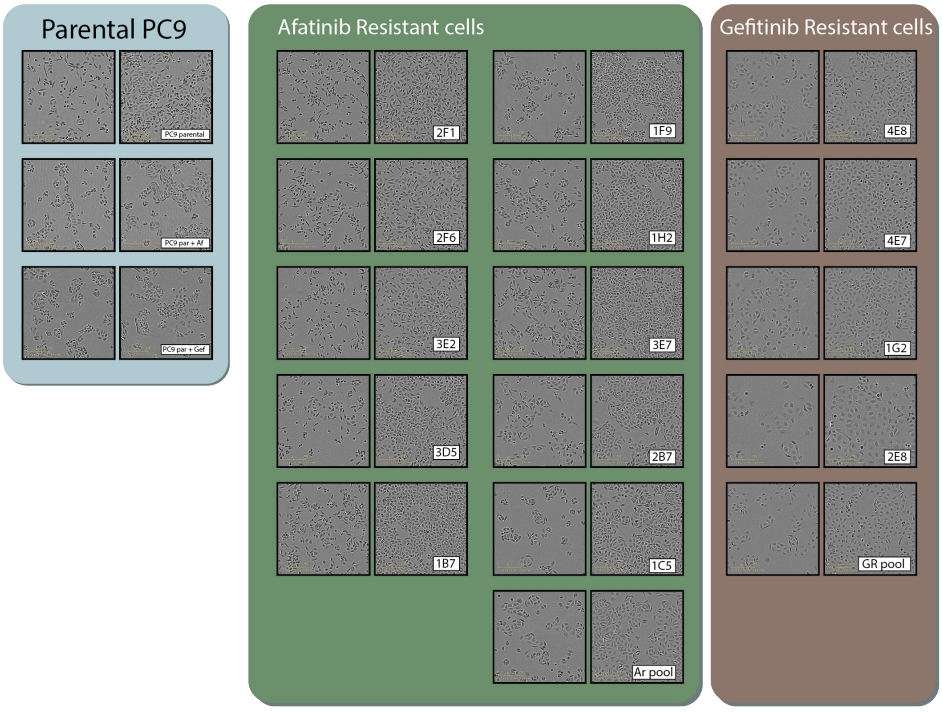


Figure S1 – Morphology of resistant clones
In order to monitor the confluence of all conditions and resistant clones, all cell lines were grown from low to high confluence and light microscopic pictures were taken to show the cell shape in either low or high confluence. Background colour indicates drug treatment: blue: parental, with either afatinib or gefitinib short treatment; green: afatinib resistant clones; brown: gefitinib resistant clones. Bars within the pictures indicate the scale.

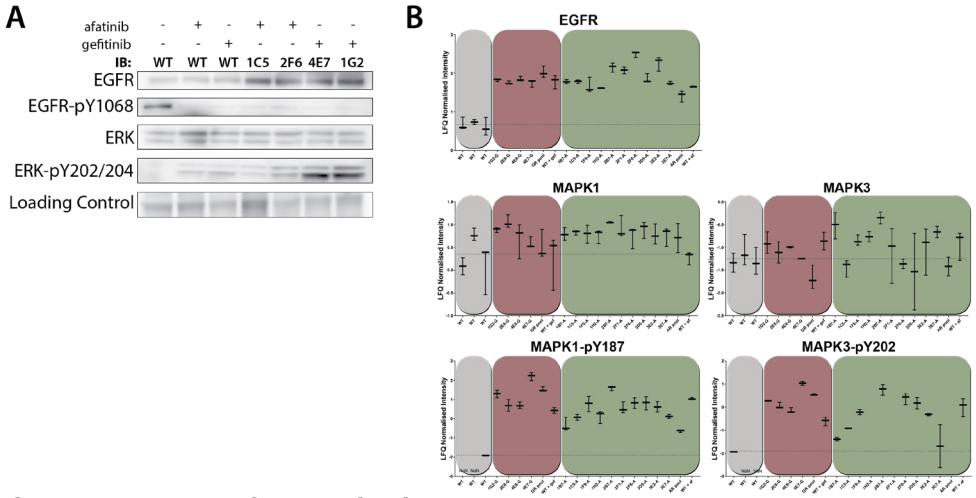


Figure S2 – EGFR and ERK activation
A: EGFR and ERK protein expression and activity was probed by western blot analysis. Wild type (WT) cells were treated for 2 days with either afatinib or gefitinib to show short-term drug response. Four clones were chosen to represent the entire clone collection: 1C5 and 2F6 (AR), 4E7 and 1G2 (GR). Used antibodies: EGFR, EGFR-pY1068, ERK1/4, ERK1/3-pY202/204. Full protein staining was used as a loading control. **B**: LFQ intensity of EGFR expression, ERK expression and ERK phosphorylation as measured by MS is shown for all conditions and clones. In grey background (red lines): 3x WT cells, in blue background (black lines): gefitinib samples (long+short term), orange background (blue lines): afatinib samples (long+short term). Error bars represent replicates. Dotted horizontal line indicates the average LFQ intensity of the WT measurement. NaN: not a number.

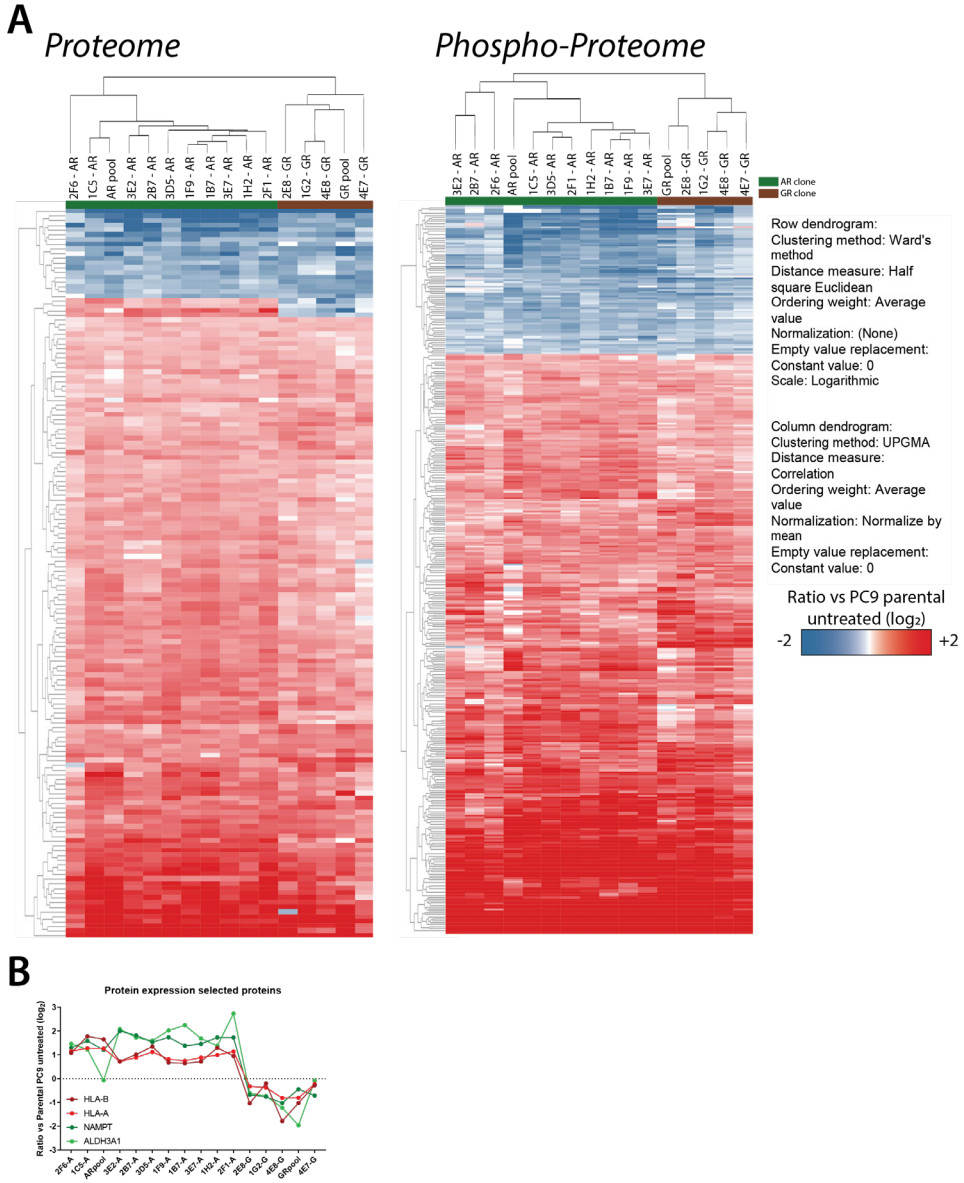


Figure S3 – Top regulated (phospho)proteins across all clones
A: In heatmaps all proteins (left) or phosphosites (right) are shown that were significantly regulated in a minimum of 13 clones. The clustering on the top is, even though the similar regulation of al proteins/sites, still according to drug type (indicated on the top: green: AR, brown: GR). B: In the proteome dataset, 4 proteins were oppositely regulated between afatinib and gefitinib. These four proteins (HLA-B, HLA-A, NAMPT and ALDH3A1) are shown in more detail in a line plot, with the log2 ratio vs the untreated WT cells on the Y-axis and the different clones on the X-axis, in the same order as in the heatmap.

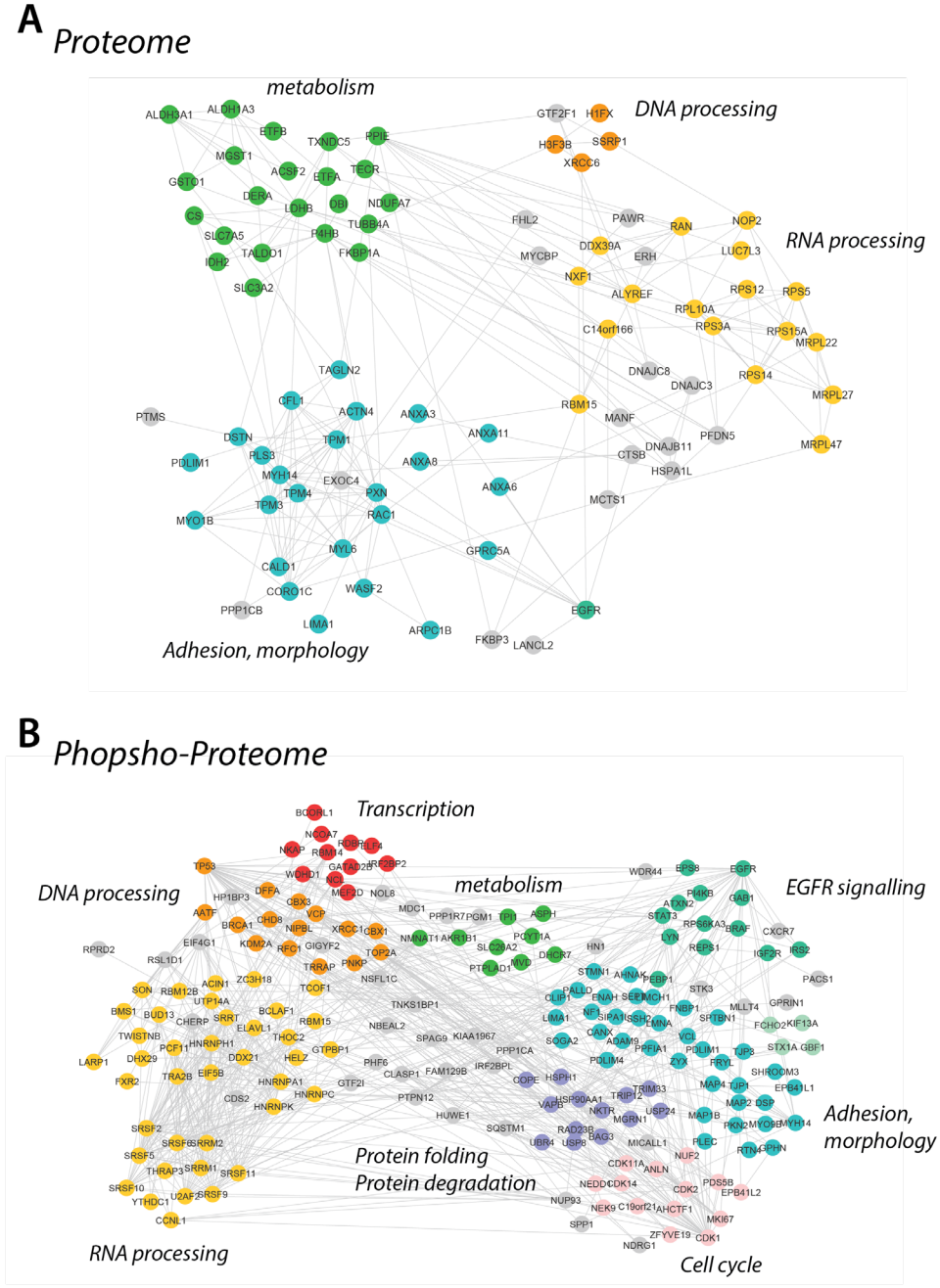


Figure S4 – General resistance pathways
A: Proteins and **B**: Phosphoproteins that were commonly regulated in all clones indicate general regulated pathways across EGFRi resistant cells. These proteins were visualised in clusters, manually annotated within their respective pathways. Different colours represent the different pathways, as described.

Material and Methods

Cell culture and IncuCyte
PC9 cells, formerly known as PC-14 (Sigma 90071810-1 VL), were cultured in a humidified atmosphere and in 5% CO_2. Growth medium was RMPI-1640 (Lonza), supplemented with 10% fetal bovin serum (Thermo), 1% Pen/Strep (Lonza) and 5 mM L-glutamine (Lonza). Cells were tested negative for mycoplasma infections on a regular basis.
Cells were monitored in the IncuCyte (Essen Bioscience) to determine growth speed. Cells were plated in 96-wells plates in medium containing appropriate drug concentrations (no drug, 1 µM afatinib (Sellechchem, BIBW2992) or 1 µM gefitinib (Selleckchem, ZD1839)), calculated to contain ~20% confluence. Cells were grown and monitored for up to 4 days or until full confluence was reached.
If no drug was in the standard medium and drug treatment was desired, cells were allowed to attach after plating for 24 hours, after which the medium was replaced with the appropriate drug containing medium.

Western Blot
Protein lysates were labeled with the Amersham QuickStrain Protein Labeling Kit, according to manufactures protocol. Read out of cys5 fluorescence of total protein labeling after transfer was used as a loading control. After labeling, cell lysates were prepared with 100mM DTT (Sigma) in XT sample buffer (Biorad), and heated at 95°C for 10 minutes for protein denaturation. 10-20 ug of sample was loaded on a precast gel (Biorad, 12% SDS) and run at 40 mAmp for 2 hours in XT mops (Biorad) running buffer. Proteins were transferred to a methanol activated PVDF membrane in transfer buffer (1x TG buffer (Bio-Rad), 20% methanol), for 90 minutes at 90 volts.
The membrane was blocked using 5% milk in TBS-Tween (Tris-buffered saline (pH7.5) with 0.1% Tween (Biorad)) for 1 hour. Primary antibodies were diluted in 5% milk and the membrane was incubated over night at 4°C. Membranes were washed in TBS-Tween and incubated with secondary antibodies (Dako, anti-rabbit P0448, anti-mouse P0447) for 2 hours. Membranes were developed using Pierce ECL Plus Substrate (Thermo). Primary antibodies used: EGFR (Cell Signaling, C74B9), EGFR-1068 (Abcam, #ab5644), ERK1/2 (Cell Signaling, CST9102s), ERK1/2 p-208/p206 (Cell Signaling, #4370).

Resistance development clones
Drug sensitive, parental PC9 cells (hence wild type (WT)) were treated with either 1 µM afatinib or 1 µM gefitinib. Drug containing medium was replaced at least once in 4 days, and cells were split when confluence approached 100%. Cells were cultured for 16 weeks, after which the drug sensitivity was determined to be comparable to WT cells. Cells were now called Afatinib Resistant (AR) or Gefitinib Resistant (GR) pools. Both pools were used for further clonal expansion of resistant clones. Cells were plated in 96-wells plate with an assumed confluence of 0.5 cell per well in drug containing medium. Medium was refreshed every 4 days. Plates were manually checked by light microscopy for the developing of a clone. After several weeks, multiple wells contained clones. Cells were transferred to 6-wells plates for further growth and maintenance.

Sample Preparation before MS
Cells were grown to near confluence before harvesting. Plates were washed with PBS (Lonza) to remove residual FBS and detached from the plate with 1% trypsin (Lonza) in PBS. Suspended cells were washed with PBS 3x by cycles of centrifugation and resuspension in PBS. Washed cell pellets were stored in -80 for further use.
Cell pellets were lysed by resuspension in lysis buffer (1% sodium deoxycholate (SDC), 10 mM TCEP, 100 mM TRIS, 40 mM chloroacetamide, protease inhibitors (Complete mini tablet, EDTA-free, Roche), phosphotase inhibitors (PhosSTOP tablet, Roche), at pH 8.5) and subsequently sonicated for 15 min (30 sec on, 30 sec off). Lysis buffer without SDC was added to decrease the SDC percentage to 0.4%. Next, 2.5 mM MgCl and 0.5% Benzonaze (Millipore) was added for nuclease treatment, which was incubated for 2 hours at room temperature, while shaking. Lysate was spin down (20 min, 45000 rpm) to clear cellular debris. With a Pierce™ BCA Protein Assay Kit (Thermo) protein concentration was determined. Per condition or clone 3 replicates of 200 µg of protein was digested by Trypsin (1:100), at 37°C over night.
After digestion, SDC was precipitated by the addition of 1% formic acid (FA) and lowering of the pH to 2 and centrifugation. Samples were desalted on a C18 column (AssayMAP 5uL C18) with the use of the AssayMAP Bravo Platform (Agilent Technologies). The columns were washed with 10% acetonitrile (ACN) and 1% FA, after samples were loaded onto the columns with a speed of 5 µl/min. Peptides were eluted by 80% ACN and 1% FA and dried in a speedvac.

Automated phospho enrichment Fe-IMAC
Digested peptide mixtures were enriched for phosphorylated peptides by Fe-IMAC, using the AssayMAP Bravo Platform (Agilent Technologies). Columns were primed and washed with 0.125% trifluoroacetic acid (TFA) in ACN, and washed with loading buffer (80% ACN, 0.1% TFA). 200 µg of desalted peptides were loaded on the columns (Fe(III)-NTA, 5µl, Agilent technologies). Non-phosphorylated peptides were collected as flow through, after which bound phosphorylated peptides were eluted with 1% NH_3 directly in 10% FA. Peptides were dried in a speedvac and stored in -80°C until LC-MS/MS analysis.

MS analysis
Peptide mixtures of either complete lysate (full proteome) or after phospho enrichment (phospho proteome) were analysed on an Agilent 1290 system coupled to a Orbitrap Q-Exactive (Thermo). Dried peptides were solubilised in 0.1% FA and loaded on a trap column (10 mm x 100 µm inner diameter, Poroshell C18 material, Agilent Technologies). Subsequently, peptides were separated via elution over an analytical column (55 cm x 75µm inner diameter, Poroshell C18 material (Agilent Technologies)) by a gradient of buffer B (0.1% FA, 80% ACN) of 2 hours from 8 to 32% B (phospho) or of 3 hours from 10-36% B (proteome) with a flow rate of ~350 µL/min.
Peptides were brought into the MS by electro spray, applying a voltage of 1.9 kV. Transfer capillary temperature was set to 320°C. S-lens RF level was set at 60. Full scan mass spectra (ranging from 375-1600 m/z) were acquired using a resolution of 60K and a AGC target of 3e6 with a maximum IT of 20 ms. For every top15 (proteome) or top12 (phopsho) peaks fragmentation energy of 27 NCE was applied to produce HCD fragment spectra, using a resolution of 30K, and an AGC target of 1e5, with a maximum IT of 50 ms. Dynamic exclusion was 16 (2 hours gradient) or 24 (3 hours gradient) seconds.

Data analysis and statistics
RAW data was analysed by MaxQuant (v 1.6.2.3), using a Andromeda search engine. For protein identification a uniprot fasta file was used. 2 missed cleavages were allowed, with a minimum peptide length of 7 amino acids. Fixed modification was Carbamidomethyl. Variable modifications allowed were methionine oxidation and serine, threonine and tyrosine phosphorylation. A revert decoy database was used with a protein FDR of 1%. The minimum score of modified peptides was set to be 40 and all contaminants were included. Mass tolerance was 20 ppm. Match between runs option was checked. For protein and peptide quantification label free quantification was used.
Further data processing was done in Perseus (v1.5.6.0). MQ output was filtered for contaminants, only identified by site and identified as reverse sequence. Proteins or peptides were included if identified at least 2 times in 1 condition. No imputation of missing values was allowed. LFQ intensities were normalised by subtracting the column mean. A student t-test was used to determine fold change and p-value. A S0 curve [53] was calculated with a FDR of 5% to determine statistical relevance per protein or phospho site.
Hierarchical clustering was done in TIBCO Spotfire (v7.6.0). Phospho peptides were analysed by PHOTON [22]. GO enrichments were done in the online environment of Metascape [54] (http://metascape.org), using the measured (phospho)proteome as a background. All graphs were made in GraphPad Prism 7. Networks were annotated manually based on Uniprot, STRING-db, PhosphoSitePlus and publically available literature, and visualised in Cytoscape (v3.7.0).

References

1. Bhullar KS, Lagarón NO, McGowan EM, Parmar I, Jha A, Hubbard BP, Rupasinghe HPV: **Kinase-targeted cancer therapies: Progress, challenges and future directions**. *Mol Cancer* 2018, **17**:1–20.
2. Ramirez M, Rajaram S, Steininger RJ, Osipchuk D, Roth MA, Morinishi LS, Evans L, Ji W, Hsu C, Thurley K, et al.: **Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells**. *Nat Commun* 2016, doi:10.1038/ncomms10690.
3. Hata AN, Niederst MJ, Archibald HL, Gomez-Caraballo M, Siddiqui FM, Mulvey HE, Maruvka YE, Ji F, Bhang HEC, Radhakrishna VK, et al.: **Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition**. *Nat Med* 2016, **22**:262–269.
4. Song JY, Kim CS, Lee JH, Jang SJ, Lee SW, Hwang JJ, Lim C, Lee G, Seo J, Cho SY, et al.: **Dual inhibition of MEK1/2 and EGFR synergistically induces caspase-3-dependent apoptosis in EGFR inhibitor-resistant lung cancer cells via BIM upregulation**. *Invest New Drugs* 2013, **31**:1458–1465.
5. Stuhlmiller TJ, Miller SM, Zawistowski JS, Nakamura K, Beltran AS, Duncan JS, Angus SP, Collins KAL, Granger DA, Reuther RA, et al.: **Inhibition of lapatinib-induced kinome reprogramming in ERBB2-positive breast cancer by targeting BET family bromodomains**. *Cell Rep* 2015, **11**:390–404.
6. Zhang Z, Lee JC, Lin L, Olivas V, Au V, Laframboise T, Abdel-rahman M, Wang X, Levine AD, Rho JK, et al.: **Articles Activation of the AXL kinase causes resistance to EGFR- targeted therapy in lung cancer**. *Nat Publ Gr* 2012, **44**:852–860.
7. Engelman JA, Jänne PA: **Mechanisms of acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non-small cell lung cancer**. *Clin Cancer Res* 2008, **14**:2895–2899.
8. Joshi M, Rizvi SM, Belani CP: **Afatinib for the treatment of metastatic non-small cell lung cancer**. *Cancer Manag Res* 2015, **7**:75–82.
9. Day D, Siu LL: **Approaches to modernize the combination drug development paradigm**. *Genome Med* 2016, **8**:1–14.
10. Hammerlindl H, Schaider H: **Tumor cell-intrinsic phenotypic plasticity facilitates adaptive cellular reprogramming driving acquired drug resistance**. *J Cell Commun Signal* 2018, **12**:133–141.
11. Rosell R, Karachaliou N, Morales-espinosa D, Costa C, Molina MA, Sansano I, Gasco A, Viteri S, Massuti B, Wei J, et al.: **Adaptive resistance to targeted therapies in cancer**. 2013, **2**:152–159.
12. Sharma S V., Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, McDermott U, Azizian N, Zou L, Fischbach MA, et al.: **A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations**. *Cell* 2010, **141**:69–80.
13. Ahn A, Chatterjee A, Eccles MR: **The Slow Cycling Phenotype: A Growing Problem for Treatment Resistance in Melanoma**. *Mol Cancer Ther* 2017, **16**:1002–1009.
14. Kim TK, Herbst RS, Chen L: **Defining and Understanding Adaptive Resistance in Cancer Immunotherapy**. *Trends Immunol* 2018, **39**:624–631.
15. Mroz EA, Tward AD, Pickering CR, Myers JN, Ferris RL, Rocco JW: **High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma**. *Cancer* 2013, **119**:3034–3042.
16. Jensen NF, Stenvang J, Beck MK, Han B, Gupta R, Rasmussen MH, Tarpgaard LS, Pfeiffer P, Bosman F, Tejpar S, et al.: **Establishment and characterization of models of chemotherapy resistance in colorectal cancer : Towards a predictive signature of chemoresistance**. *Mol Oncol* 2015, **9**:1169–1185.
17. Sagawa Y, Fujitoh A, Nishi H: **Establishment of three cisplatin-resistant endometrial cancer cell lines using two methods of cisplatin exposure**. *Tumor Biol* 2011, **32**:399–408.
18. Knott GJ, Doudna JA: **CRISPR-Cas guides the future of genetic engineering**. *Science (80-)* 2018, **362**:866–869.
19. Terai H, Soejima K, Yasuda H, Nakayama S, Hamamoto J, Arai D: **Activation of the FGF2-FGFR1 Autocrine Pathway : A Novel Mechanism of Acquired Resistance to Gefitinib in NSCLC**. *Mol Cancer Res* 2013, **11**:759–768.
20. Ruprecht B, Koch H, Medard G, Mundt M, Kuster B, Lemeer S: **Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns**. *Mol Cell Proteomics* 2015, **14**:205–215.
21. Potel CM, Lin M-H, Heck AJR, Lemeer S: **Defeating major contaminants in Fe 3+-IMAC phosphopeptide enrichment**. *Mol Cell Proteomics* 2018, **17**:1028–1034.
22. Rudolph JD, Graauw M De, Water B Van De, Geiger T: **Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Report Elucidation of Signaling Pathways from Large-Scale Phosphoproteomic Data Using Protein Interaction Networks**. *Cell Syst* 2016, **3**:585–593.
23. Wang B-D, Lee N: **Aberrant RNA Splicing in Cancer and Drug Resistance**. *Cancers (Basel)*

2018, **10**:458.
24. Ladner RD, Caradonna SJ: **The Human dUTPase Gene Encodes both Nuclear and Mitochondrial**. *J Biol Chem* 1997, **30**:19072–19080.
25. Ladner RD, Carr SA, Huddleston MJ, McNulty DE, Caradonna SJ: **Identification of a consensus cyclin-dependent kinase phosphorylation site unique to the nuclear form of human deoxyuridine triphosphate nucleotidohydrolase**. *J Biol Chem* 1996, **271**:7752–7757.
26. Blethrow JD, Glavy JS, Morgan DO, Shokat KM: **Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates**. *Proc Natl Acad Sci* 2008, **105**:1442–1447.
27. Katsogiannou M, Andrieu C, Rocchi P: **Heat shock protein 27 phosphorylation state is associated with cancer progression**. *Front Genet* 2014, **5**:346.
28. Chi KN, Yu EY, Jacobs C, Bazov J, Kollmannsberger C, Higano CS, Mukherjee SD, Gleave ME, Stewart PS, Hotte SJ: **A phase I dose-escalation study of apatorsen (OGX-427), an antisense inhibitor targeting heat shock protein 27 (Hsp27), in patients with castration-resistant prostate cancer and other advanced cancers**. *Ann Oncol* 2016, **27**:1116–1122.

29. Heinrich J-C, Tuukkanen A, Schroeder M, Fahrig T, Fahrig R: **RP101 (brivudine) binds to heat shock protein HSP27 (HSPB1) and enhances survival in animals and pancreatic cancer patients**. *J Cancer Res Clin Oncol* 2011, **137**:1349–1361.
30. Lelj-Garolla B, Kumano M, Beraldi E, Nappi L, Rocchi P, Ionescu DN, Fazli L, Zoubeidi A, Gleave ME: **Hsp27 Inhibition with OGX-427 Sensitizes Non–Small Cell Lung Cancer Cells to Erlotinib and Chemotherapy**. *Mol Cancer Ther* 2015, **14**:1107–1116.
31. Li Q, Bao W, Fan Q, Shi W-J, Li Z-N, Xu Y, Wu D: **Epidermal growth factor receptor kinase substrate 8 promotes the metastasis of cervical cancer via the epithelial-mesenchymal transition**. *Mol Med Rep* 2016, **14**:3220–3228.
32. Gorsic LK, Stark AL, Wheeler HE, Wong SS, Im HK, Dolan ME: **EPS8 Inhibition Increases Cisplatin Sensitivity in Lung Cancer Cells**. *PLoS One* 2013, **8**:e82220.
33. Lanzetti L, Rybin V, Malabarba MG, Christoforidis S, Scita G, Zerial M, Di Fiore PP: **The Eps8 protein coordinates EGF receptor signalling through Rac and trafficking through Rab5**. *Nature* 2000, **408**:374–377.
34. Xue J, Chen Y, Wu Y, Wang Z, Zhou A, Zhang S, Lin K, Aldape K, Majumder S, Lu Z, et al.: **Tumour suppressor TRIM33 targets nuclear β-catenin degradation**. *Nat Commun* 2015, **6**:1–16.
35. Shi X, Mihaylova VT, Kuruvilla L, Chen F, Viviano S, Baldassarre M, Sperandio D, Martinez R, Yue P, Bates JG, et al.: **Loss of TRIM33 causes resistance to BET bromodomain inhibitors through MYC- and TGF-β–dependent mechanisms**. *Proc Natl Acad Sci* 2016, **113**:E4558–E4566.
36. Kim TK, Herbst RS, Chen L: **Defining and Understanding Adaptive Resistance in Cancer Immunotherapy**. *Trends Immunol* 2018, **39**:624–631.
37. Keating GM: **Afatinib: A review of its use in the treatment of advanced non-small cell lung cancer**. *Drugs* 2014, **74**:207–221.
38. Shi H, Hugo W, Kong X, Hong A, Koya RC, Moriceau G, Chodon T, Guo R, Johnson DB, Dahlman KB, et al.: **Acquired resistance and clonal evolution in melanoma during BRAF inhibitor therapy**. *Cancer Discov* 2014, **4**.
39. Baeissa HM, Benstead-Hume G, Richardson CJ, Pearl FMG: **Mutational patterns in oncogenes and tumour suppressors**. *Biochem Soc Trans* 2016, **44**:925–931.
40. Hanahan D, Weinberg RA, Francisco S: **The Hallmarks of Cancer**. *Cell* 2000, **100**:57–70.
41. Ruprecht B, Zaal EA, Zecha J, Wu W, Berkers CR, Kuster B, Lemeer S: **Lapatinib resistance in breast cancer cells is accompanied by phosphorylation-mediated reprogramming of glycolysis**. *Cancer Res* 2017, **77**:1842–1853.
42. Zaal EA, Wu W, Jansen G, Zweegman S, Cloos J, Berkers CR: **Bortezomib resistance in multiple myeloma is associated with increased serine synthesis**. *Cancer Metab* 2017, **5**:7.
43. Grasso C, Jansen G, Giovannetti E: **Drug resistance in pancreatic cancer: Impact of altered energy metabolism**. *Crit Rev Oncol Hematol* 2017, **114**:139–152.
44. Rahman M, Hasan MR: **Cancer metabolism and drug resistance**. *Metabolites* 2015, **5**:571–600.
45. Ménard L, Floc'h N, Martin MJ, Cross DAE: **Reactivation of mutant-EGFR degradation through clathrin inhibition overcomes resistance to EGFR tyrosine kinase inhibitors**. *Cancer Res* 2018, **78**:3267–3279.
46. Sadowski L, Pilecka I, Miaczynska M: **Signaling from endosomes: Location makes a difference**. *Exp Cell Res* 2009, **315**:1601–1609.
47. Mulder C, Prust N, van Doorn S, Reinecke M, Kuster B, van Bergen en Henegouwen P, Lemeer S: **Adaptive Resistance to EGFR Targeted Therapy by Calcium Signaling in NSCLC Cells**. *Mol Cancer Res* 2018,
48. Hsu M, Tseng C, Chen F, Hsu JT: **Non-small-cell lung cancer cells combat epidermal growth factor receptor tyrosine kinase inhibition through immediate adhesion-related responses**. *Onco Targets Ther* 2016, **9**:2961–2973.

49. Suda K, Tomizawa K, Dds MF, Murakami H, Osada H, Maehara Y, Yatabe Y, Sekido Y, Mitsudomi T: **Epithelial to Mesenchymal Transition in an Epidermal Growth Factor Receptor -Mutant Lung Cancer Cell Line with Acquired Resistance to Erlotinib**. *J Thorac Oncol* 2011, **6**:1152–1161.
50. Lee SC-W, Abdel-Wahab O: **Therapeutic targeting of splicing in cancer.** *Nat Med* 2016, **22**:976–86.
51. Tress ML, Abascal F, Valencia A: **Alternative Splicing May Not Be the Key to Proteome Complexity**. *Trends Biochem Sci* 2017, **42**:98–110.
52. Blethrow JD, Glavy JS, Morgan DO, Shokat KM: **Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates**. *Proc Natl Acad Sci* 2008, **105**:1442–1447.
53. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response**. *Proc Natl Acad Sci* 2001, **98**:5116–5121.
54. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, Moulton HM, Dejesus P, Che J, Mulder LCF, et al.: **Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding**. *Cell Host Microbe* 2015, **18**:723–735.


MASS
SPECTROMETER
PhD

Part 4

Synopsis

CHAPTER SEVEN

Outlook
Lay summary
Nederlandse Samenvatting
Curriculum Vitae
List of Publications
Acknowledgements

Outlook

SEVEN

When I started my PhD, I started with the idea of using proteomic techniques and mass spectrometry to answer a biological question related to drug resistance in cancer. However, the major challenge I found along the way was not so much understanding the underlying biology, as well as finding it in the complex and elaborate data produced by shotgun mass spectrometry. Where others work on the optimisation of the production of high quality proteomics data, I focussed on it's interpretation in order to truly value the significance of any of the found result. In this thesis, I discussed ways of improving the biological implication of a proteomics experiment by either reducing proteomic complexity (Chapter 2) or enhancing proteomic coverage (Chapters 3 and 4), while using a model of drug resistance against tyrosine kinase inhibitors in epithelial carcinomas (Chapters 5 and 6).

Show what we already know: from repetition to discovery

Shotgun proteomics, similar to other omics techniques, is hypothesis generating, producing more questions than answers and should therefore be used as such. Consequently, the main objective of a shotgun proteomics experiment is not to arrive at an absolute conclusion, but merely to generate new hypotheses to base further research on. The main hurdle in this process is finding the balance between describing observations that already have been made by others versus delineating new observations, which are unique to the subject studied. Both have added value, since reproducing other people's results will strengthen both yours as well as their data, while novel findings will push the research field forward. However, more than once, presented research data *only* shows those observations that already have been discussed in previously published literature. On the other hand, acquired data can remain unpublished due to little success in validation experiments or lack of novelty. The step from showing over and over again what is already known, to really discovering something new is incredibly difficult to make in this type of research, and, unfortunately, sometimes new discoveries are lacking in the average shotgun proteomics study.

Several technical obstacles specific to MS based proteomic research play a role in the preservation of this issue. For example, the impossibility of signal amplification creates missing values and lower confidence in identification and quantification. Furthermore, the proteome itself is extremely dynamic, due to an extremely fast protein turnover and a multitude of ways to regulate protein functionality (i.e. abundance, localisation, post-translational modifications, protein-protein interactions etc.), which cannot be measured simultaneously. These factors dilute the signal and will create noise in which important details might be lost. However, despite these technical difficulties, one of the main restraints on arriving

at a beneficial result when undertaking a discovery based experiment, is the fact that all findings are mapped onto existing knowledge. By using clustering algorithms like STRING-db, Panther, DAVID or PHOTON merely regulations, of which we already know they exist, are being visualised. The "new" aspect of the project is limited to the description of the regulation of a known pathway in a specific biological model system. Although interesting, there is much more to be discovered, which is currently completely gone unnoticed because of this knowledge-based barrier. Due to the current gap between data production and data interpretation, the power of a proteomic shotgun experiment remains mostly unused, like playing a game of patience on a super computer.

Proteomic success is a two-way street

The transformation of the proteomic field from characterisation [1] towards functional [2] proteomics has been a major topic over the last years [3,4]. Mainly, an increasing amount of statistical and computational tools have become available in order to interpret and visualise large scale proteomic data [5]. Although great advances have been made, the virtue of data processing tools available alone, does not always determine the success of a functional proteomic experiment. A large factor in the possibility of proper data interpretation is the initial design of the experiment. Although shotgun proteomics is discovery-based opposed to hypothesis-driven, proper preparation and design of the experiment before executing it will aid the data interpretation and will increase the success rate of the study. In other words: the better the question, the better the answer.

A possible strategy to achieve this is, is by limiting the scope of the experiment, following the strategy: less is more. As discussed in chapter 2 and put into practice in chapter 5, existing biochemical techniques that enable the researcher to focus on a specific subset of proteins can be applied, combined with MS as a read-out. Examples are, pull downs (e.g. immunoprecipitation [6], PTM or kinase enrichments [7,8]) depletion (e.g. thermal profiling [9]) or elimination (e.g. limited proteolysis [10]) of specific proteins of interest. By carefully choosing a focus point, relatively more time can be spent on each observation within the data. This will decrease the coverage of the entire model proteome but increase the coverage of the partial proteome of interest, leading to a clearer picture and eventually better and easier data interpretation.

High quality data interpretation is a two-way street: both the rationale of the experiment as well as the quality of available tools contributes to the best possible result. In the end, it is the responsibility of the proteomic researcher to make the most out of the techniques that are at his/her disposal.

The sky is the limit

Opposed to decreasing the size of a mass spectrometric dataset to achieve more detail, many endeavours have already been described to increase the size of a dataset in order to get a better coverage. As discussed in chapter 3, increasing the numbers of samples, optimising lysis [11] and/or mass spectrometric settings and protocols [12] or additional fractionation [13] can be applied to uncover as much proteins and peptides within a single experiment as possible. However, as discussed in chapter 4, the same, and maybe even more, can be achieved without a minute in the lab, and without a single euro spend.

Due to the high-throughput nature of proteomic experiments, but the low-throughput nature of validation experiments, a large part of published mass spectrometric data remains unexplored. Instead of redoing similar experiments, reusing and combining publically available datasets will result in more biological insight to be uncovered per study and less data (and time) to be lost [14,15].

An interesting but extremely important consideration to be made in this regard is the perspective used during data (re)processing. Similar to defining of a clear research question before the start of an experiment, it is important to assess the purpose of the data analysis. As demonstrated in chapter 4, alternative conclusions will be obtained if data is used at the level of identification or quantification, or originates from a specific biological background. By changing the question (i.e. what is identified, what is regulated or what is regulated specifically in this model system?), the same set of data points will lead to very different results.

The fact that a change in viewpoint will lead to a different conclusion is also the main argument in favour of data reuse. There is an incredibly large number of ways to analyse and explore proteomics data, which are complementary to each other and can be applied in parallel. Therefore, each and every proteomics dataset will only benefit from data reuse, for an alternative analysis will aid further understanding of the model studied and decrease the waste of information, time and resources.

Drug resistance studied from the first moment on

Cancer is a disease comprised out of many indispositions and maladies. Besides a wide range of therapeutic challenges caused by different cancer origins and inter-tumour and patient heterogeneity, the often inevitable development of drug resistance poses a major challenge on cancer therapeutics. A still regularly used perspective on drug resistance is the study of genomic alterations after a cell (or patient) has been exposed to a drug for an extensive amount of time. A large variety of anti-drug mechanisms have been resolved through this approach, ranging from drug target mutations [16], upregulation of drug efflux pumps [17] or mutations of alternative driver oncogenes [18].

Despite these prominent examples of the importance of oncogenic mu-

tations, it becomes increasingly apparent that resistance is driven by a much larger variety of cellular processes. The main difference between the study of either genomic alterations or cellular adaptations is the time frame in which resistance is considered to develop. As shown and discussed in chapters 5 and 6 of this thesis, it takes only a matter of days to transform a drug sensitive cell into a morphologic and oncogenic different cell. Within this relative short time frame significant differences, at both proteome and phosphoproteome level, affect all major cellular processes, including transcription, translation, splicing, apoptosis and metabolic pathways.
Additionally, as shown in Chapter 6, a great homogeneity existed between different resistant clones. Probably the main reason for this observation was the relative short time frame in which resistance was allowed to develop. Most likely, if these cells would have been cultured in the presence of the drug longer and would have been allowed time to acquire additional adaptations or even mutations, heterogeneity would increase. This can be perceived as a technical artefact of cell culture, or a valuable lesson to learn: the longer the cells are given the chance to adjust to a treatment, to more heterogeneous a population will become. Therefore, the study of early drug responses could provide new therapeutically windows and decrease the issue of heterogeneity [19,20].

Intrinsic, acquired and adaptive: all part of the same story

The terminology used in literature to describe cancer resistance is dominated by the use of three terms: intrinsic, acquired and adaptive resistance. These terms are used in a mixed manner throughout literature, while actually describing similar processes. This results in the existence of different definitions of these terms, mainly determined by the context of the study it is used in. The majority of studies which describe intrinsic or acquired resistance pursue genomic alterations to explain resistant phenotypes. However, these terms do not necessarily imply the nature of the resistance mechanism, but merely its timing, i.e. it was already there before treatment or it was induced by the treatment. The term adaptive resistance on the other hand, is often used to describe short-term, dynamic or reversible effects. However, it is merely implying the nature but not the timing of the resistance mechanisms: i.e. it is a cellular adaptation, not a genomic alteration. Hence, adaptive resistance mechanisms can either be intrinsic or acquired, and intrinsic or acquired mechanisms can either be adaptive or genomic. Since these three terms describe quite different processes and stages in the development of resistance in cancer, they do not exclude each other. Most likely, they co-exist within in a tumour, separated by time and aiding each other into existence.
With the increasing knowledge we have on the development and manifestation of drug resistance in cancer, the immense complexity of this disease becomes more and more apparent. Consequently, the awareness

of the difficulty of therapeutic success increases. However simultaneously, the more knowledge we have, the better and more targeted treatment can become as well. The immense leaps ahead in cancer therapy of the last decades are a living proof of this.

Final considerations

In this thesis I have ventured to extract biological knowledge out of shotgun proteomics data and I have contemplated about the complex problems around drug resistance in cancer. For either case it is true that the existence of multiple research strategies, methods of data analysis or various cellular mechanisms complicate matter, although simultaneously add more insight than a single point of view could ever provide. For they do not exclude each other, they are intrinsically linked and cannot work without each other. By combining multiple views, individual values and contributions are not diluted, but will augment to the complete picture, for the whole is greater than the sum of its parts.

Just as in the world of Jack, Algernon and Earnest, it turned out that the truth is more complicated than anyone could have guessed. Different perspectives on the same matter, based on different experiences, priorities and available knowledge will lead to different conclusions. Only, when all these bits and pieces are brought together, the simple truth will come forth.

SEVEN

Introduction to proteomics and lay summary of this thesis

SEVEN

Every cell in our body contains an identical piece of DNA that we got from our parents, bearing the information that we need to be who we are. Still, every cell is different, ranging from round blood cells, square skin cells to moving muscle cells. The differences appear because every cell contains a unique set of proteins, read from the static piece of DNA. This collection of proteins, or ***proteome***, is highly dynamic and will change according to the needs of every cell. If the surroundings of a cell change, the proteins do too. This dynamic characteristic of these molecules makes the study of proteins, or ***proteomics***, an extremely rich source of information about the ongoing processes inside a cell, in sickness and in health. Changes in the proteome will result in direct knowledge of any event, and contains valuable information on how to act next.

However, this rich source of information is at the same time the biggest hindrance in the field of proteomics: the complexity of the proteome increases the level of difficulty of studying it. Proteins are made and broken down extremely quickly, they are folded into 3D structures, decorated with extra chemical groups, they sometimes work alone, sometimes bound to each other and their masses can vary greatly. Additionally, as in contrast with the study of the DNA, it is not possible to copy the molecules in the laboratory: you have to deal with what you've got. A series of chemical steps and an enormous sensitivity of the measurement instruments, the ***mass spectrometer*** (short mass spec or MS), is needed to make it even possible to study proteins with high enough confidence.

Because of the large developments of instrumentation over the last decades, it is now possible to measure several different characteristics of a protein. Besides finding your protein in cells (*identification*) it is also possible to see how much of it is there (*quantification*) or how it is decorated (*modifications*). These modifications represent specific chemical groups that can be put on or off a protein and will change its features, like its activity, location or stability (the time it takes to break down the protein). Placing or removing such modifications can be done unlimited by enzymes, throughout the entire life cycle of a protein. One of these modifications is a phosphate (or ***phosphorylation***), and is put on a protein by an enzyme called ***kinase***. Phosphorylation plays an important role in cellular biology and is a focus in this thesis. Several techniques exist to study phosphorylation of proteins on a large scale with MS. The specific research field of protein phosphorylation is called ***phosphoproteomics***.

Due to the important role phosphorylation has in the function of proteins, a large amount of drugs have been developed specifically targeting kinases: ***kinase inhibitors***. The inhibitors are mainly used in the treatment of cancer, in which too high activity of a kinase often causes tumour growth. Even though this is a very specific and effective treatment, it usually

doesn't last long. Almost every single patient develops immunity against these drugs, after which the tumour can grow back. The development of immunity is induced in several ways, in different layers of the processes within the cell. One route that, so far, has been understudied is ***adaptive immunity***, in which the cells are able to adapt to a treatment after only a very short amount of time. For this, the cell uses quick, dynamic and reversible processes, like protein phosphorylation. Because only limited knowledge exist on this topic, mass spectrometry proves to be an excellent tool to map the entire process on a large scale.

In this thesis, new ideas are expressed about increasing the yield of a proteomic experiment by either optimising the experiment or the data processing. Next, the study of the development of adaptive immunity in breast- and lung cancer cells is shown.

In **Part 1** some relevant literature is discussed about the different topics that this thesis beholds. **Chapter One** describes the entire protocol that is necessary in order to study proteins. This ranges from getting the proteins out of the cells, processing the proteins so they can be measured, and finally measuring them using a MS. Furthermore, some words are spent on kinase inhibitors, the rise and development of immunity in cancer and ways to study these proteins in the laboratory. **Chapter Two** provides increasing detail about all possible ways in the lab to study the functionality of a kinase inhibitor within a cell. Several chemical approaches are discussed that provide information about to which protein an inhibitor binds and the subsequent effects on the cell.
Part 2 describes the processes of making and processing proteomics data. One of the first steps in a proteomic experiment is breaking down a protein in smaller pieces (*peptides*), by using specific enzymes (*proteoases*). **Chapter Three** describes the parallel use of several different proteases to increase the diverse collection of peptides, in order to increase the amount of information per protein. **Chapter Four** addresses the reuse of proteomic data for new analyses. Because of the increasable large amount of data produced by a single MS experiment, but the limited time a researcher can spend on its analysis, a large part of each dataset stays undiscovered. Recycling of data is an easy and quick method to get new conclusions without doing a day of lab work, and at the same time decrease the amount of data that is lost. Due to large differences between datasets and experiments originating from different laboratories, it is necessary to re-analyse these datasets. In this study, we chose the focus on the phosphoproteome. However, by not looking at individual phosphorylation sites, but group them in several ways, we made it possible to combine a large amount of different studies. The combined studies made it possible to arrive at a new conclusion, as an addition on previously published work.

In **Part 3** we made the step from using the MS to study the biological process of adaptive immunity. **Chapter Five** describes a short treatment of lung cancer cells with kinase inhibitors, in order to study the direct effect on the phosphoproteome. One of these effects, we found, was increased calcium consumption from the environment. This dependency could be used as a weak spot for the cells that gained immunity: by combining a kinase inhibitor with calcium deprivation, the development of immunity was prevented. In **Chapter Six** cells were made immune against several kinase inhibitors in the laboratory. After a while, several clones evolved, of which we studied the phosphoproteome. The main conclusion of this study was that, even after quite a while, the cells showed remarkable similarities (homogeneity). This is in contrast with the large differences (*heterogeneity*) that can develop upon immunity on the level of the DNA (*mutations*). This homogeneity on the level of the proteome could possibly lead to new insight of where these cells are most vulnerable, leading to new methods of treatment.

Part 4 closes this thesis with some of my ideas about the future of the development of the field of proteomics, kinase inhibitors, immunity and cancer research.

Inleiding in proteomics en leken samenvatting proefschrift

SEVEN

Elke cel in ons lichaam bevat een zo goed als identieke streng DNA, verkregen via onze ouders, en drager van alle informatie nodig om te zijn wie we zijn. Toch is elke cel anders, van ronde bloedcellen, vierkante huidcellen tot bewegelijke spiercellen. Deze verschillen worden veroorzaakt doordat elke cel een andere set van eiwitten bevat, afgelezen van het statische DNA. Deze set van eiwitten, of ***proteoom***, is dynamisch, en verandert naargelang de behoeftes van elke cel. Verandert de omgeving, dan veranderen de eiwitten. Dit dynamische karakter van deze klasse moleculen maakt de studie naar eiwitten, of ***proteomics***, een belangrijke bron van informatie over het reilen en zeilen van cellen in ons lichaam, zowel tijdens ziekte als gezondheid. Veranderingen in het proteoom geven een direct beeld van wat er gebeurt, en bevatten belangrijke informatie over hoe te handelen.

Deze rijke bron aan informatie is tegelijkertijd de grootste hindernis in het veld van proteomics: de complexiteit van het proteoom maakt het een hele uitdaging om het goed te kunnen bestuderen. Eiwitten worden snel gemaakt en afgebroken, ze zijn gevouwen tot 3D structuren, versiert met extra chemische groepen, ze werken soms alleen, soms gebonden aan elkaar en hun massa kan sterk variëren. Verder is het, in tegenstelling tot het DNA, niet mogelijk de moleculen te kopiëren in een laboratorium: je moet het doen met wat je krijgt uit de cellen die je tot je beschikking hebt. Een reeks van chemische handelingen en een enorme gevoeligheid van het meetinstrument, de ***massa spectrometer*** (afgekort mass spec of MS), is daarom nodig om het mogelijk te maken eiwitten op een betrouwbare manier te kunnen onderzoeken.

Door enorme ontwikkelingen van de instrumentatie over de afgelopen decennia is het vandaag mogelijk veel verschillende aspecten van eiwitten te meten. Naast de mogelijkheid om de aanwezigheid van een specifiek eiwit in een cel te bepalen (*identificatie*) is het mogelijk de hoeveelheid (*kwantificatie*) en versieringen (*modificatie*) op eiwitten te meten. Deze modificaties bestaan uit verschillende chemische groepen die op en af eiwitten geplaatst kunnen worden en die de activiteit, plaats in de cel of stabiliteit (hoe snel het eiwit wordt afgebroken) van een eiwit bepalen. Het plaatsen en weghalen van modificaties kan ongelimiteerd worden gedaan door enzymen tijdens de levensduur van een eiwit. Een modificatie die een belangrijke rol speelt in de biologie van cellen, en die een focus is van het onderzoek in dit proefschrift, is een fosfaatgroep (een ***fosforylering***), op een eiwit geplaatst door een enzym genaamd ***kinase***. Er bestaan verschillende technieken die het mogelijk maken op grote schaal fosforylering op eiwitten te meten met een MS. Het veld die eiwitfosforylering beschrijft wordt ***fosfoproteomics*** genoemd.

Door de belangrijke rol die fosforylering speelt in de functionaliteit van ei-

witten is er een grote groep van moleculen ontwikkeld die de kinase kan remmen: ***kinase remmers***. Deze remmers worden vooral veel gebruikt in het behandelen van kanker, waar een te hoge activiteit van een kinase vaak de oorzaak van de tumor groei is. Al is dit een zeer specifieke en effectieve behandelmethode, hij is helaas vaak van korte duur. Zo goed als elke patiënt wordt na verloop van tijd immuun, waarna de tumor terug komt. De ontwikkeling van immuniteit gebeurt op vele manieren, in verschillende lagen en processen in de cel. Een nog weinig onderzochte weg is ***adaptieve immuniteit***, waarin de cel op zeer korte termijn zich aanpast tegen een behandeling. Hierin maakt de cel gebruik van snelle, dynamische en omkeerbare processen, zoals eiwit fosforylering. Omdat er nog relatief weinig over dit proces bekend is, is massa spectrometrie een zeer geschikte techniek om op grote schaal dit proces in kaart te brengen.

In dit proefschrift zijn nieuwe ideeën over het vergroten van de opbrengst van een proteomic experiment door middel van aangepaste onderzoeksmethoden en verbeterde dataverwerking beschreven. Vervolgens is MS gebruikt om het fenomeen adaptieve immuniteit in borst- en longkanker te bestuderen.

In **Deel 1** wordt relevante literatuur besproken over de verschillende onderwerpen in dit proefschrift. **Hoofdstuk Een** beschrijft het protocol om eiwitten te meten. Van het uit de cellen halen, het verwerken van de eiwitten tot meetbare stukjes, en uiteindelijk het meten van hun massa met een MS. Verder worden kinase remmers en het ontstaan van immuniteit in kanker besproken, en de verschillende manieren om dit proces in het laboratorium te kunnen bestuderen. **Hoofdstuk Twee** gaat in meer detail in op de mogelijkheden om in het laboratorium te kunnen onderzoeken hoe een kinase remmer precies werkt in de cel. Verschillende chemische methoden worden uitgelicht die gebruikt kunnen worden om te bepalen aan welk eiwit een remmer bindt en wat het effect ervan is op het proteoom en op de cel.
In **Deel 2** wordt ingegaan op het maken en het verwerken van data afkomstig van MS experimenten. Een van de eerste stappen in het mogelijk maken van het meten van eiwitten is het afbreken van eiwitten in kleinere stukjes (*peptiden*) door middel van specifieke enzymen (*proteasen*). **Hoofdstuk Drie** beschrijft het parallelle gebruik van verschillende proteasen om op deze manier een completere verzameling van peptiden te verzamelen, en daarmee een completer beeld van de gemeten eiwitten te krijgen. In **Hoofdstuk Vier** gaat het over het hergebruik van bestaande proteomic datasets voor nieuwe analyse. Door de grote hoeveelheid data die wordt geproduceerd in een enkel proteomic experiment, maar de gelimiteerde tijd die een onderzoeker heeft om de data te analyseren, blijft een groot deel van de data onbelicht. Het recyclen van gepubliceer-

de data is een manier om snel en zonder labwerk tot nieuwe conclusies te kunnen komen, en om tegelijkertijd het verloren gaan van data te verminderen. Doordat verschillende datasets en experimenten niet direct met elkaar te vergelijken zijn, omdat er teveel parameters verschillen, is het nodig om de data opnieuw te bewerken. Wij hebben er in deze studie voor gekozen om ons te focussen op fosfoproteomic data. Door niet naar individuele fosforylaties te kijken zoals gebruikelijk is, maar ze te groeperen, hebben wij het mogelijk gemaakt verschillende studies samen te voegen. De gecombineerde dataset maakt het mogelijk tot nieuwe inzichten te komen, als toevoeging op de al gepubliceerde analyses.

In **Deel 3** wordt de stap gemaakt naar het gebruik van MS om het proces van adaptieve immuniteit te bestuderen. In **Hoofdstuk Vijf** worden longkankercellen kort behandeld met kinase remmers om zo het directe effect op het (fosfo)proteoom van deze specifieke moleculen te kunnen besturen. Een van de effecten die is gevonden is de toename van calcium opname uit de omgeving. Deze afhankelijkheid kon worden gebruikt als zwakke plek voor de immuun wordende cellen: door de kinase remmer te gebruiken in combinatie met het remmen van calcium opname kon er worden gezorgd dat de cellen tot een sneller einde kwamen, en niet de kans kregen immuun te worden tegen de behandeling. In **Hoofdstuk Zes** zijn cellen immuun gemaakt tegen verschillende kinase remmers in het laboratorium. Na verloop van tijd ontstonden verschillende clones, waarvan het (fosfo)proteoom onderzocht is. De belangrijkste conclusie hieruit was dat, ondanks de verstreken tijd, de clones nog steeds een grote gelijkenis lieten zien (*homogeniteit)*. Dit is in contrast met de grote *heterogeniteit* die kan ontstaan naar aanleiding van veranderingen in het DNA (*mutaties*) die kunnen leiden tot immuniteit. De homogeniteit op het niveau van het proteoom zou kunnen leiden tot nieuwe inzichten over waar een cel het meest kwetsbaar is, wat kan leiden tot nieuwe behandelmethoden.

Deel 4 sluit het proefschrift af met mijn ideeën over de toekomst en het verdere verloop van proteomics, kinase remmers, immuniteit en kanker onderzoek.

Curriculum vitae

I studied Biomedical Sciences at the Utrecht University from 2008 until 2012. Next, I continued with the Life Science Master 'Biology of Disease', in Utrecht, from 2012 until 2014. Within this master programme, I performed two research internships. The first was in the lab of Prof. Paul Coffer, in the UMC Utrecht. My research was focussed on mapping differences during osteogenic differentiation upon shRNA knockout of a panel of histone methyltransferases. I worked with several biological and biochemical techniques, ranging from cell culture, PCR, shRNA transfections and several types of microscopy.

Next, I wrote a literature thesis in the lab of Prof. Albert Heck in Utrecht, under the supervision of dr. Arjan Scholten. The title of the thesis was "Overview of the secretome of human immune cells", with a focus on mass spectrometry based techniques.

As the last part of my master, I left Utrecht to join the lab of Prof. Geir Slupphaug, in Trondheim, Norway. Here, I worked in a proteomics lab, familiarizing myself fully with mass spectrometry. The project I worked on was focussed on mapping the proteomic differences between different stages of multiple myeloma, using patient's material.

After my studies, I remained in Utrecht, where I started my PhD in the lab of Prof. Heck, under the supervision of dr. Simone Lemeer. My project was based on Lemeer's acquired VIDI grant, redarding the interest to the development of adaptive resistance in non-small cell lung cancer. The resulting work of the last 4,5 years are combined in this thesis.

List of Publications

Mulder C, Leijten N, Lemeer S: Proteomic tools to study drug function. *Curr Opin Syst Biol* 2018, 10:9–18.

Mulder C, Prust N, van Doorn S, Reinecke M, Kuster B, van Bergen en Henegouwen P, Lemeer S: Adaptive Resistance to EGFR Targeted Therapy by Calcium Signaling in NSCLC Cells. *Mol Cancer Res* 2018 Nov;16(11):1773-1784.

Zatula A, Dikic A, **Mulder C**, Sharma A, Vågbø CB, Sousa MML, Waage A, Slupphaug G: Proteome alterations associated with transformation of multiple myeloma to secondary plasma cell leukemia. *Oncotarget* 2017, 8:19427–19442.

Schmidlin T, Garrigues L, Lane CS, **Mulder TC**, van Doorn S, Post H, de Graaf EL, Lemeer S, Heck AJR, Altelaar AFM: Assessment of SRM, MRM 3 , and DIA for the targeted analysis of phosphorylation dynamics in non-small cell lung cancer. *Proteomics* 2016, 16:2193–2205.

Acknowledgements

During these last years I have laughed, cried, celebrated and cherished. I have learned more than I could ever imagine, and I would do it all over again if I had the choice. But most importantly, I could not have done it without all the great people around me. There is absolutely no way to capture how thankful I truly am in only a few pages, there are so many people, so many years and so many things that happend. I will never have enough words to really thank you all. So to everyone: I could not have done it without you, and I really hope you all stay in my life for ever.

The first person I have to thank is **Simone**. You were truly the greatest, with you there was never a dull moment. Tosti's and burgers were never far away, and whenever the moment asked for it you would take a grasp out of your reservoir of one-lines. Thank you so much for your strength and your endless enthusiasm. You are a beautiful person, thank you for that. Never change, you're awesome!

Dear **Esther**, also without you I could not have survived these last years. It was really great to have the mental support whenever I needed it, right there in our safe, little office. All anger, frustrations, joy and celebrations we could share. Thanks for many beautiful years, and I am glad you stand beside me one more time at my defence. I wish you all the best in the future, you are a great scientist, and I believe you will get really far with whatever you set your mind to.

I really need to thank the entire **Heck Lab**. There are way too many people to thank individually, although I do need to call out a few people in particular.

I have to start with **Séverine**. Because of you I got my first real confidence in the lab. You taught me how to make a proper lysis buffer and how to deal with proteomics data. Thank you, and I wish you all the best. Dear **Thierry**, our coffee drinking activities started in Brixen, and never ended, which I am really happy about. Thank you for all the weird discussions and support no matter what. You are a great person, and I hope we will see each other again, either in the Netherlands or in Switzerland.

And let's not forget all the hours spend at fixing Orbi's and LCs. So, Alba, Matina, Nadine, Clement and Kelly S, thank you all so much for all the hours in the lab, the anger and love we shared over our little monsters. Thank you, **Matina**! We met at my job interview when you gave me a lab tour with such enthusiasm that made me remember you most of all. And luckily that enthusiasm never ended. Thanks for all the endless (but gezellige!) hours at all our different orbi's we had to take care of, you were

lovely to work with. **Nadine**, you are the best! You are fun and down-to-earth, lovely and sometimes completely crazy. You are a really good scientist and a wonderful person. Dear **Clément**, your knowledge about the mass spec and LC is amazing, thank you for all your help. But most of all, thank you for all the borrels, wine and beers. And thank your so much, **Kelly**, you are so much fun, and such an incredible positive person, so full of energy. It was lovely to work with you!

Thank you so much, dear **Kyle**, for all your great stories and the incredible power to make everyone laugh, always. There will never be a boring lunch, borrel or party with you around. I hope you will take care, I am really glad we met and I hope you will stay in my life for a long time. Thank you **Donna**, I loved it to talk to you and laugh with you, at the mass spec, in the cell culture or over a beer. Hang in there, you have so much determination in you that you will finish a beautiful PhD, and whatever comes next. If you put your mind to something, you can do anything you like. Thank you, **Kelly D**, for your unique personality and incredible confidence. Thank you **Anna Ressa** for your strength, and that you showed that you can do anything if you set your mind to it. Thank you so much, **Aneika**, for the joy you bring, your positivity and skills as a scientist and as a good friend. Thanks, **Niels** and **Sander**, It was really nice to have met you and worked with you. Thank you, and the rest of our little subgroup, for an amazing couple of retreats and gezellige Friday meetings! Thank you **Arjan**, without you, no heck lab (borrels..)! Always ready to get some beers at Friday afternoon, eating heart breakers will never be the same. Thank your **Mirjam**, without you no work will ever be done. You are absolutely essential for the lab, thank you for all your patience with us and your hard work, every single day. You were great to work with, you make all our lives so much easier.
Thank you, **Albert**, for your always honest opinion and great knowledge.
Thank you, **Corine,** for your patience and taking care of all of us so well.

Of course, also outside of the lab I have had the much needed support. To address the people in this part of my life, I will continue in Dutch.

Niets is zo belangrijk als een plek waar je even af kan koelen van die cellen, mass specs en experimenten. Wat heb ik veel gehad aan mijn drukke leven buiten het lab, om alle frustraties en energie even kwijt te raken, en de volgende dag weer door te kunnen gaan.

Als eerste natuurlijk het **FDA**! Ieder van jullie is fantastisch, ik beschouw het als een voorrecht om elke dinsdag even aan de dagelijkse sleur te kunnen ontsnappen, en gewoon lekker te dansen, te lachen en te genieten. Lieve **Wijnand**, ik wil jou in het bijzonder bedanken. Wat is het telkens weer een feestje om met jou te mogen dansen. Zowel de maandag als de dinsdag breng jij mij even weg van alle drukte, elke week is binnen een half uurtje warming-up mijn hoofd weer leeg. Jij brengt mij de ontspanning die ik zo hard nodig heb, en ik hoop dat je beseft hoe fantastisch je bent. Blijf alsjeblieft nog lang doen wat je doet! Lieve **Anne**, ik kan met niemand lachen zoals met jou. Het is zo fijn dat het nergens over hoeft te gaan, en dat dansen, samen eten en lachen zo gemakkelijk gaat met jou, je bent geweldig! Dankjewel lieve **Maartje**, voor je prachtige persoonlijkheid en je enorme power waarmee je zingt en danst, en hoe je dat zo mooi met ons kan delen. Dank jullie, **Elske, Roos** en **Amanda**, voor alle gezelligheid in de auto en tijdens het freubelen. En dankjewel lieve **Leo**, wat ben jij een mooi mens! Ik vind het fantastisch dat ik jou heb mogen leren kennen en met je heb mogen dansen. Alle geluk en liefde toegewenst!

En natuurlijk bedankt aan iedereen van **Utrecht for life**! Wat zijn jullie een heerlijke groep mensen, en wat fijn dat ik daar deel van uit mag maken. De loyaliteit, het gemak en de gezelligheid is zo bijzonder, ik zou dat voor geen goud willen missen. In het bijzonder, bedankt **Joppe** en **Lotte**, voor de gezellige bridge avondjes. Bedankt voor jullie geduld als ik eindeloos moest nadenken, en bedankt voor het heerlijke eten, wijntjes en gezelligheid elke keer weer.

Bedankt **Borinka** en **Maike**! Wat is het geweldig dat we na al die jaren nog steeds het DNA-lab zijn. Al die gezellige avondjes eten, kletsen en spelletjes doen, naar het theater en de dierentuin, van Noorwegen tot Zwitserland. Ik hoop dat we voor altijd vriendinnen blijven, jullie zijn geweldig. En bedankt, Bo, voor de heerlijke avondjes squash, ook al win ik nog steeds niet van je…

Bedankt lieve **Anna**, voor de persoon die je bent, nog steeds de liefste, na bijna 25 jaar. Jij leert mij jezelf te zijn en de wereld vanuit een compleet ander perspectief te bekijken. Wat ben je fantastisch, alle liefs toegewenst!

En natuurlijk ook heel erg veel liefs voor mijn fijne familie! Lieve **Papa** en **Mama**, bedankt voor de onvoorwaardelijke steun die mij zo ver heeft laten komen. Jullie zijn geweldig, al zeg ik dat veel te weinig. Bedankt lieve zusters, **Anne-Minke** en **Carola**. Jullie zijn uniek en fantastsich, ik reken mij rijk met jullie twee. Lieve **Hendrik-Jan**, bedankt voor alle gezelligheid, biertjes en spelletjes. Bedankt lieve **Frits, Martine, Jeremy, Joel**, en **Yvonne** in gedachten, ik heb me vanaf dag één enorm welkom gevoeld als ik bij jullie was.

Als aller laatste moet ik degene bedanken die mij het meest heeft bijgestaan, mijn eeuwige steun en toeverlaat. Lieve **Lars**, wat had ik moeten doen zonder jou. Jij was er in mijn moeilijkste uurtjes en zag mijn donkerste kanten. Jij veegde mij op als ik in stukjes thuis kwam, jij zette mij weer met mijn voeten op de grond en hielp mij elke dag weer verder te gaan, buitengewoon onvoorwaardelijk. Ik had dit nooit zonder jou kunnen doen, jij haalt de beste versie van mijzelf in me naar boven.
Je bent geweldig en ik hou van je.