

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER
FAKULTÄT FÜR CHEMIE UND PHARMAZIE DER
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN



BIOLOGICAL AND TRANSLATIONAL CANCER PROTEOMICS

SOPHIA SUSANNA DOLL

AUS
MÜNCHEN, DEUTSCHLAND

2018

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Professor Dr. Matthias Mann betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 20.12.2017

Sophia Susanna Doll

Dissertation eingereicht am _____16.01.2018_____

1. Gutachter: Prof. Dr. Matthias Mann _____

2. Gutachter: Prof. Henrik Daub _____

Mündliche Prüfung am _____26.02.2018_____

Abstract

Cancer is the second leading cause of death worldwide and many cancer subtypes remain poorly understood. Most conventional chemotherapeutic treatments are still associated with life-threatening toxic side effects that primarily result from a lack of specificity directed towards cancer cells. Recent breakthroughs in genomic and transcriptomic sequencing technologies have allowed the molecular profiling of thousands of tumors in different cancer types. It has become evident that cancer cannot be considered a singular disease and that its manifestations cannot exclusively be explained by the accumulation of genetic mutations. Instead, epigenetic and proteomic changes as well as posttranslational modifications (PTMs) of proteins are crucial drivers of oncogenesis. In this thesis, I investigated system-wide alterations in cancer at several biological and cellular levels using mass spectrometry (MS). Starting from the nucleus of the cell, I explored the epigenetic changes in lymphoma at the biotechnology company Genentech Inc. We found that the methyltransferase EZH2 is the most significantly over-expressed epigenetic regulator in cancer, and is co-regulated with a cell cycle network. Zooming out from the nuclear level, I analyzed phosphorylation-signaling alterations in primary and secondary glioblastoma cell line models at the University of California, San Francisco (UCSF). Here, I focused on the interplay between the MAPK and PI3K signaling cascades. At the Max Planck Institute (MPI) of Biochemistry, I moved on to translational proteomics, working with human cancer tissues. I optimized an MS-based proteomic workflow for the rapid screening of clinical tissue samples and showed that MS-based proteomics can be used for novel therapeutic target identification in end-stage chemorefractory cancer patients.

Through its industrial, academic and clinical perspective on a variety of proteomic methods, this PhD thesis demonstrates that MS-based proteomics is applicable to personalized oncology. My hope is that this unique resource of the identity, quantity, and alterations of proteins, phosphosites and histone PTMs, may reveal new insights into the field of oncology.

Table of contents

I) Introduction	1
1. The proteome and its large scale investigation by mass spectrometry ...	1
1.1 The human proteome	1
1.2. Mass spectrometry-based proteomics	3
1.2.1. Mass spectrometry history.....	3
1.2.2. Top-down vs. bottom-up proteomics	3
1.2.3. From cells or tissues to proteomes: Bottom-up proteomic workflows	4
1.2.4. Sample preparation: From 'in-gel' to 'in-solution', and 'in-StageTip'	6
1.2.5. PTM-enrichment strategies	7
1.2.6. Peptide fractionation for deep quantitative proteomes	10
1.2.7. Liquid chromatography-mass spectrometry.....	11
1.2.8. Fragmentation strategies.....	13
1.2.9. Mass spectrometer types	15
1.2.10. Acquisition methods	16
1.2.11. Protein identification and quantification	19
1.2.12. Label- and label free-based quantification.....	21
1.2.12. Nearly complete proteomes.....	24
2. The cancer proteome at the nuclear, cellular and tissue level	26
2.1. Cancer	26
2.2. Personalized cancer medicine	27
2.3. Nuclear level: The epigenetic landscape in oncology	29
2.3.1. Epigenetics.....	29

2.3.2.	Contributions of MS-based proteomics to epigenetics	30
2.3.3.	The histone methyltransferase enhancer of zeste homolog 2	31
2.3.4.	Histone demethylases	33
2.3.4.1.	JmjC domain-containing histone demethylases	33
2.3.4.2.	Lysine-specific histone demethylases	34
2.4.	Cellular level: The phosphorylation-based signaling landscape in oncology	36
2.4.1.	Phosphorylation.....	36
2.4.2.	The MAPK and PI3K pathways	37
2.4.3.	Development of kinase inhibitors.....	39
2.4.4.	MS-based phosphoproteomics of the MAPK and PI3K pathways.....	40
2.5.	Tissue level: ‘from bench to bedside’	42
2.5.1.	Current cancer diagnosis in clinical practice	42
2.5.2.	Cancer FFPE tissues proteomics	44
2.5.3.	In the quest of more reliable and early cancer biomarkers using MS-based proteomics.....	46
2.5.4.	Clinical cancer proteomics consortia	47
II)	Aims of the thesis	48
III)	Publications	50
Article 1:	Mass spectrometry-based detection and assignment of protein posttranslational modifications	51
Article 2:	Bioinformatics analysis of thousands of TCGA tumors to determine the involvement of epigenetic regulators in human cancer.....	61

Article 3: Quantitative proteomics reveals fundamental regulatory differences in oncogenic HRAS and Isocitrate Dehydrogenase (IDH1) driven astrocytoma.....	77
Article 4: Phosphoproteome analysis of the MAPK pathway reveals previously undetected feedback mechanisms	97
Article 5: Quantitative phosphoproteomic analysis of the PI3K-regulated signaling network.....	105
Article 6: Region and cell-type resolved quantitative proteomic map of the human heart and its application to atrial fibrillation.....	112
Article 7: Rapid proteomic workflow for solid tumors reveals LSD1 as a drug target in an end stage cancer patient	127
IV) Discussion	152
V) References.....	156
VI) Acknowledgments	172

To my Dad.

I) Introduction

1. The proteome and its large scale investigation by mass spectrometry

Each cell is a genuine masterpiece. Biological information is transcribed from DNA to RNA and finally translated into proteins. These proteins may be further modified post translationally, leading to intricate signaling patterns within and between different cell types. Recent breakthroughs in DNA and RNA sequencing technologies allow the molecular profiling of essentially complete genomes and transcriptomes¹. While these methodologies have revolutionized our understanding of a vast array of human diseases, including cancer, fundamental mechanisms are not only driven by genetic or transcriptomic alterations. Instead, they only manifest clinically if they involve changes at the protein level as well. Proteins are the paramount active biological entities in cells and work in concert with each other and other biomolecules as molecular machines, ensuring that each cell generates energy, communicates with its environment, divides, moves, performs its specific biological functions, or commits apoptosis. The full complement of proteins in a biological system is termed the proteome, while proteomics refers to the large-scale investigation of the proteome using a variety of technologies. The unceasing development of ever more powerful proteomic methods over the last decades now enables the analysis of proteomes in great depth. As a result, the investigation of complex biological functions and promising clinical applications are becoming realistic².

1.1 The human proteome

With approximately 20,000 protein-coding genes, the human genome is five times smaller than that of an onion in terms of genome size. Thus, the number of genes

alone does not determine the complexity of an organism. Additional biological differences originate from the regulation of the genes, alternative splicing, protein localization and interactions. The human proteome, in particular, is diversified by many protein variations (isoforms), and posttranslational modifications (PTMs), which alone add at least another order of magnitude of complexity. More than 90% of the human genes are spliced³ and altogether, more than 300 types of PTMs have been described⁴. Combined, these modifications yield millions of different possible 'proteoforms'⁵, resulting in a dramatic increase in the complexity of the human proteome. Among these PTMs, many have been shown to regulate normal and pathogenic cell biological functions. The most common covalent modifications are phosphorylation, ubiquitination, acetylation, glycosylation and methylation and together they vastly increase the functional diversity of proteins. Among the some 300,000 human PTM sites that have so far been recorded in the PhosphositePlus database⁶, only a small percentage have been assigned to regulatory or biological functions. Additionally, proteins are often modified at multiple sites, either through a combinatorial or sequential addition of functional groups.

Mass spectrometry (MS)-based proteomics has evolved into the method of choice for the large-scale identification and quantitation of nearly all expressed proteins as well as their site-specific PTMs^{2,7}. Phosphorylation is the most extensively studied PTM because it is a key modulator of cellular signal transduction. MS-based quantitative phosphoproteomics has already revealed site-specific phosphorylation dynamics after EGF stimulation⁸, identified molecular switches underlying FGFR cellular responses⁹, oncogenic signaling in the MAPK and PI3K pathways (Article 3), and 'druggable' kinases¹⁰. These examples among many highlight the potential of MS-based phosphoproteomics to improve our understanding of molecular mechanisms, identify clinically relevant biomarkers and uncover potential therapeutic targets.

1.2. Mass spectrometry-based proteomics

1.2.1. Mass spectrometry history

Mass spectrometry (MS) is a technology that determines the mass/charge (m/z) ratios of ions for which Thompson (Th) is the proper unit. Based on the charge state of an ion, the m/z ratios can be converted to its molecular mass with a standard unit of Dalton (Da). The analyte of interest (peptides in this thesis) must first be ionized in its intact form, as this is essential for its identification and quantification. This is difficult for labile biomolecules but this obstacle was overcome with the development of two soft ionization techniques (electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI)) in 1988. In ESI, for which John Fenn received a share of the Nobel Prize in Chemistry in 2002, analytes are directly vaporized and ionized from a liquid phase using a high voltage via rapid solvent evaporation^{11,12}. As the solvent of the droplet evaporates, the charge density increases, resulting in a stream of charged ions that are transferred into the vacuum of the mass spectrometer. ESI has become particularly popular since it can directly be coupled to a liquid chromatography (LC) system, which is ideal suited to the analysis of complex protein and peptide mixtures. Subsequent technological advances included the miniaturization of ESI in the form of the particularly sensitive, low-flow nano-electrospray¹³.

1.2.2. Top-down vs. bottom-up proteomics

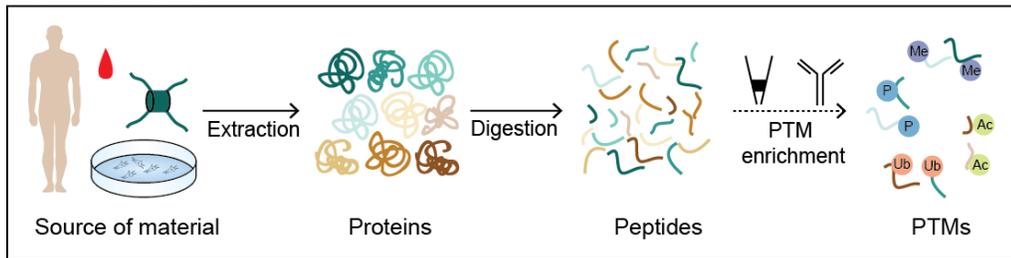
Conceptually, there are two MS-based proteomic strategies, termed 'top-down' and 'bottom-up'. In top-down proteomics, intact proteins are analyzed, typically in purified form¹⁴. This can be beneficial for the comprehensive analysis of protein isoforms, complex PTM conformations, and especially therapeutic antibodies. While theoretically appealing, top-down measurements, remain experimentally and computationally challenging because high molecular weight compounds are not

very sensitively ionized and their complex charge patterns render the MS and MS/MS spectra difficult to acquire and interpret. In contrast, bottom-up proteomics entails the digestion of proteins using sequence specific proteases into peptides, leading to much simpler MS analysis. Bottom-up or 'shotgun' proteomics has become the standard method for large-scale proteome analysis and has been broadly applied to the analysis of in-depth and cell-type-resolved proteomes. In this thesis, this advance has allowed me to map the human heart to unprecedented depth (Article 6).

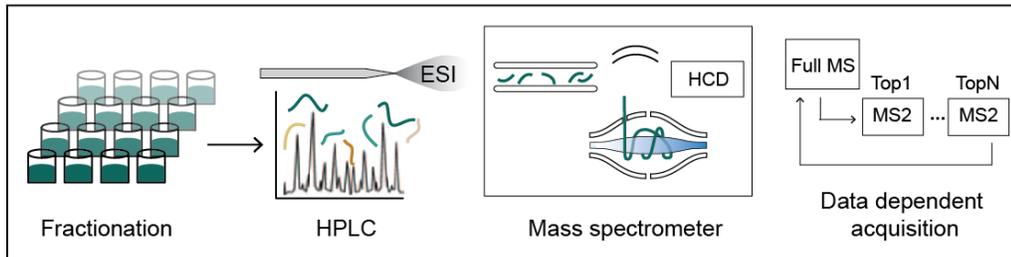
1.2.3. From cells or tissues to proteomes: Bottom-up proteomic workflows

A typical bottom-up MS-based proteomics workflow consists of three main steps: sample preparation, including protein digestion, the LC-MS measurement itself, and subsequent data analysis (Figure 1). Apart from providing a general overview, I here particularly focus on developments from our laboratory and those applied to the projects in this thesis.

A) Sample preparation



B) HPLC and mass spectrometer



C) Data analysis

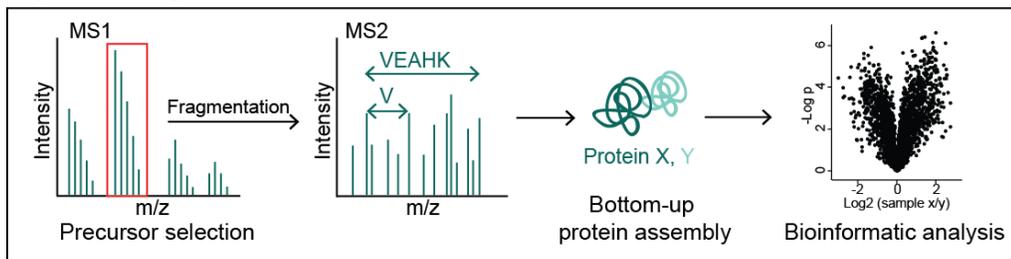


Figure 1 The bottom-up MS-based proteomic workflow A) Sample preparation consists of protein extraction, digestion and optional PTM enrichment. B) HPLC and online MS measurement. C) Data analysis includes peptide identification and quantification, assembly into proteins and subsequent bioinformatic analysis. Adapted from¹⁵.

Sample preparation includes the extraction of proteins from biological material and their digestion into peptides. These peptides are then subjected to separation by reversed phase high pressure LC (HPLC) based on their different strengths of hydrophobic interaction with a stationary phase, typically C18 modified silica. As the peptides elute from the chromatographic column, they are ionized via ESI and transferred into the mass spectrometer. The generated mass spectra provide information about the abundance (intensity) and the identity (amino acid sequence and PTMs) of the peptide.

1.2.4. Sample preparation: From 'in-gel' to 'in-solution', and 'in-StageTip'

The first step of the proteomic workflow, sample preparation, is a crucial procedure of MS-based proteomics. Sample preparation protocols have to be adapted to the source and the amount of material. In biological research, proteins have been analyzed mainly by 1D SDS polyacrylamide gel electrophoresis, a procedure that employs detergent-mediated (e.g. sodium dodecyl sulfate (SDS)) solubilization of the sample followed by SDS polyacrylamide gel electrophoresis. The development of in-gel digestion, in which proteins that are still present in the gel, are directly degraded by trypsin, paved the way for MS sample preparation that was immediately useful to biologists^{16,17}. By combining in-gel digestion with nano-electrospray, MS became applicable for the first time to proteomics^{18,19}. This procedure also permits fractionation of the proteome through the excision of the entire 1D gel in a chosen number of ranges. The development of Stop And Go Extraction tips (StageTips) allowed convenient sample handling, even of minimal sample amounts and optional peptide fractionation²⁰. Subsequently, improved LC-MS performance made it possible to move from time-consuming in-gel digestion to in-solution digestion^{21,22}. In-solution digestion employs chaotropic agents, such as urea, for protein extraction and digestion under denaturing conditions. This has the advantage of directly extracting, denaturing, and digesting the proteins in the lysis buffer. High concentrations of urea, however, are associated with decreased digestion efficiency and such weak agents do not solubilize membrane proteins, for instance. The 'Filter-Aided Sample Preparation' (FASP) allows removal of the detergents or chaotropic agents by trapping the denatured proteins on a spin-filter matrix, enabling efficient enzymatic digestion²³. The next step was the 'in-StageTip' protocol, as it permitted the robust preparation of samples in high-throughput, using robotic assistance. In-StageTip digestion employs somewhat milder detergents than SDS, such as sodium deoxycholate (SDC), which is particularly suitable for efficient cell lysis, reduction, alkylation and protein digestion in a single device. These

sample preparation developments resulted in a considerably reduced sample preparation time, contamination, and loss²⁴.

All sample preparation protocols first require the lysis of the biological source material in order to efficiently extract proteins. This step can also require mechanical breakdown, such as sonication, bead-milling or heating to increase lysis efficiency. The cysteines of the extracted proteins are then reduced and alkylated to disrupt disulfide bridges prior to digestion. The alkylation step prevents the reduced reactive thiol groups from reforming disulfide bridges. Typical reducing agents include tris(2-carboxyethyl)phosphine (TCEP) or dithiothreitol (DTT), while the most commonly used alkylating agents are iodoacetamide (IAM) or chloroacetic acid (CAA). For the proteolytic digestion of proteins, trypsin is the enzyme of choice due to its high cleavage specificity C-terminal to lysines and arginines, generating an average peptide length of 14 amino acids²⁵. The advantages of sequence specific digestion enzymes include the generation of a limited and defined set of peptides, placing a positive charge at the C-terminus in case of trypsin, and providing constraints in the bioinformatic identification of peptides²⁶. In some cases, however, the resulting tryptic peptides might be too long or too short for effective MS analysis. Other enzymes can then be employed, such as chymotrypsin, AspN, Lys-N, Lys-C, Arg-C, or Glu-C to increase the diversity of generated peptides, and to boost overall protein sequence coverage²⁷. Other methods preventing the generation of too short or hydrophilic peptides include chemical modification of lysines by propionic anhydride to neutralize charges and block lysine residues^{28,29}. This labeling approach is particularly suitable for bottom-up analysis of histone tails using trypsin as it improves sequence coverage across the lysine- and arginine-rich tails that harbor most modifications, as applied in article 2.

1.2.5. PTM-enrichment strategies

The analysis of the entire proteomes of cells or tissues is already challenging. Conceptually and practically, biological processes involving regulatory PTMs that

feature substoichiometric modifications are even more difficult and in practice always require additional enrichment steps during sample preparation. Common PTM enrichment strategies use affinity purification based on charge properties or antibody recognition. These are usually applied at the peptide level, ensuring higher accessibility of PTMs – allowing specific binding. The importance of phosphorylation has engendered highly effective protocols and approximately 240,000 human phosphorylation sites have been reported so far⁶. During the lysis of the samples, additional phosphatase inhibitors are generally added to prevent the dephosphorylation of the proteins during sample handling. Global analysis of serine- and threonine-phosphorylation (pS and pT) is commonly achieved by metal ion-based enrichment methods such as immobilized metal affinity chromatography (IMAC) or titanium dioxide (TiO₂). They rely on the interaction between the negatively charged phosphate groups and the positively charged iron (Fe) or titanium (Ti) ions, respectively (Figure 2). Non-phosphorylated peptides are washed away in the presence of salt to reduce non-specific binding of highly acidic peptides. Phosphopeptides are subsequently eluted with potassium phosphate to disrupt the phospho-Fe or -TiO₂ interactions. While initially phosphorylation enrichment required large starting material in the mg-range, workflow optimizations now allow the analysis of more than 10,000 phosphosites from minimal starting material in a 96-well format³⁰. Multiphosphorylated peptides, however, remain challenging to assign and quantify unambiguously. To this end, the combination of IMAC with TiO₂ (termed SIMAC) helps by efficiently separating of mono-phosphorylated from multiply phosphorylated peptides³¹. Alternative methods that also increase the identification of multiply phosphorylated peptides include preferential binding to graphite powder or Ti(4+)^{32,33}.

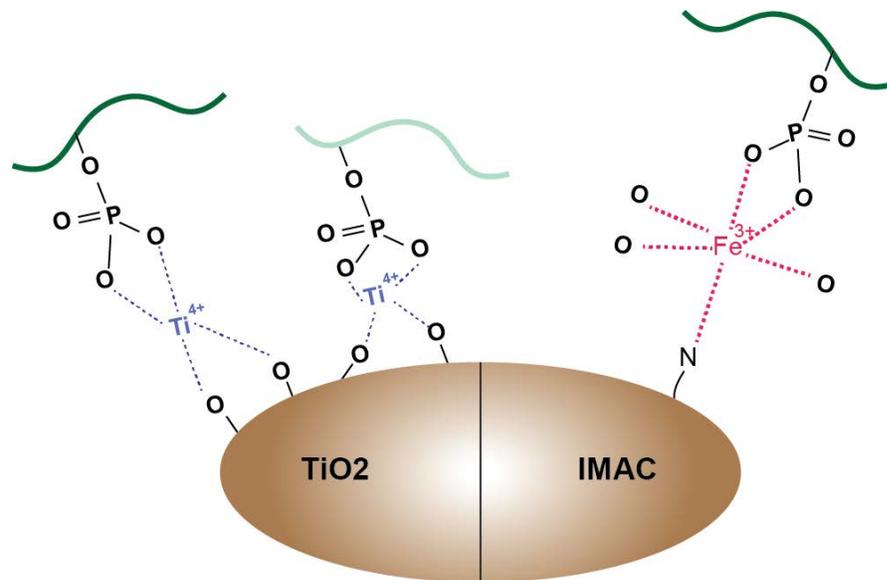


Figure 2 TiO₂ and IMAC phosphopeptide enrichment strategies.

Global phosphorylation distributions revealed that about 93%, 6%, and 1% of phosphorylated sites occur on serine, threonine, and tyrosine, respectively (Figure 3) consistent with previous observations⁸. Although phosphotyrosines (pY) are present at even lower site occupancies than pS and pT overall, they activate receptor tyrosine kinases (RTKs) and play pivotal role in multiple diseases, including cancer. For optimal identification of pY-containing peptides, they are purified from the mixture of digested peptides by specific immuno-capture. For instance, the combination of the recently developed P-Tyr-1000 antibody (Cell Signaling Technology) with LC-MS/MS analysis, has recently resulted in the identification of several hundreds of pY sites in colorectal cancer cells with a very high enrichment specificity^{34,35}.

Antibody-based enrichment strategies have been extended to the quantitative analysis of ubiquitinations, methylations, and acetylations. In particular, the development of anti-di-glycine remnant antibodies led to the identification of more than 10,000 ubiquitination sites^{36,37}. Furthermore, antibodies separately targeting mono-, di-, and tri- methylated lysines or mono- and di-methylated arginines peptides have been applied to map the human methylome in great depth³⁸. Proteome-wide acetylation analyses have likewise been made possible by antibody-

based enrichment and have uncovered pivotal cellular processes that are regulated by acetylation³⁹.

In addition to charge- or antibody-based enrichment strategies, enzyme-based enrichment can uncover proteolytic sites of biologicals of interest. For example, the subtiligase approach enabled the identification of more than 8,000 proteolytic sites, including 1,700 caspase cleavage sites in human cells⁴⁰.

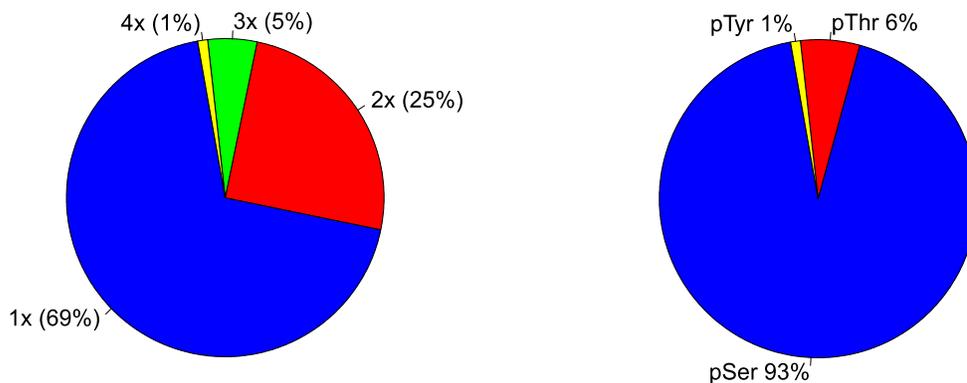


Figure 3 distribution of phosphorylation events in HCT116 cells, A) Distribution of single, doubly, triply and quadruply phosphorylated peptides B) Distribution of phosphorylated sites per amino acid. (Data generated by the author).

1.2.6. Peptide fractionation for deep quantitative proteomes

In principle, protein level fractionation would be attractive but in practice it is hampered by solubility issues and low resolution⁴¹. To reach deep proteome coverage in complex biological samples, an additional step of peptide fractionation is applied prior to LC-MS measurements. Here, tryptic peptides are separated into different fractions based on principles such as high pH reversed-phase fractionation or strong cation exchange (SCX). Since the peptides are separated from each other – decreasing complexity in LC-MS and more material can be injected onto the analytical column in total, pre-fractionation will increase overall detectability of low

abundance peptides, leading to increased proteome depth and sequence coverage. Off-line high pH reversed-phase fractionation combined with the low pH of the online LC-MS has become popular because it tends to yield overall better peptide identifications than using SCX⁴². This is mainly due to the fact that individual first dimension fractions are pooled from different parts of the gradient ('concatenation'), effectively providing orthogonal separation (meaning that peptide retention times are not correlating)⁴³⁻⁴⁵. However, because of the large diameter of the C18 columns used for fractionation, such approaches required starting material in the mg-range and the concatenation of the different fractions was generally done manually (Article 3). To make the fractionation more streamlined, our group has developed a 'loss-less nano-spider' fractionator, which enables the fractionation of very low- μ g starting material and automatically concatenates the collected fractions via a rotating valve⁴⁶. In cell lines, this approach resulted in the quantification of almost 12,000 proteins using 24 fractions. In article 6, I applied spider fractionation for the first time to tissues and quantified over 10,000 proteins in a very challenging sample.

1.2.7. Liquid chromatography-mass spectrometry

The (fractionated) peptide mixtures are then subjected to HPLC separation, which is based on the different hydrophobic interaction with a stationary phase, typically C18-silica. Peptides elute in a time dependent manner by a linear increase of an organic solvent such as acetonitrile. As the peptides elute from the chromatographic column, they are ionized via ESI. The better the chromatographic resolution, the lower the number of co-eluting peptides and the higher their concentration. This makes very long columns and very small particle sizes attractive, albeit at the cost of extremely high pressures. For instance, our group uses 75 μ m inner diameter columns with 50 cm lengths, filled with sub-2 μ m particles and requiring a pump pressure of more than 1,000 bar. To improve the ionization efficiency, formic acid is added to the solvent to provide a source of protons. The addition of the polar aprotic dimethylsulfoxide (DMSO) solvent further enhances the ionization and has been

reported to increase peptide signals⁴⁷. We also observed this increase in sensitivity in the heart atlas (Article 6), where I used 5% DMSO containing solvents for the LC-MS/MS measurements. After the ionization step, the resulting charged ions are transferred via an ion transfer tube to the vacuum region of the instrument. The mass analyzer of the mass spectrometer then assigns m/z and intensity values to the eluting peptides. They include quadrupole, time-of-flight (TOF), ion trap and Orbitrap analyzers. Quadrupoles are characterized by high reproducibility and high sensitivity but poor resolution and speed. TOF instruments have the highest scanning speed but until recently suffered from comparatively low mass resolution. Orbitraps are the most commonly used mass analyzer in proteomics today due to their high resolution, sensitivity and accurate mass capabilities⁴⁸. The Orbitrap is composed of one central spindle and two outer electrodes. It captures ions by 'electrodynamic squeezing' resulting from rapidly dropping the voltage on the central electrode. The ions subsequently oscillate around the central electrode and the frequency of oscillation is proportional to the square root of the mass of the ions. The time varying signal (the 'transient') is subsequently Fourier transformed (FT), converting the waveform of the ions into mass spectra. Coupling FT to a phased spectrum deconvolution method (Φ SDM) has been recently shown to result in doubling of the mass resolution, enabling the use of shorter transients and consequently faster analysis cycles. However, this exciting method requires extremely high computational power and is therefore not implemented on a broad scale yet⁴⁹.

As the peptide mass alone does not permit its complete characterization, a second step of mass spectrometry, termed tandem MS, MS2 or MS/MS, is needed. While, the MS1 scan yields the m/z values of the precursor ions (intact peptides), the MS2 scans result in the m/z values and intensities of their fragments. The analysis cycles in shotgun proteomics consists of selecting the TopN most abundant peptides from each MS1 scan (also termed survey of full scan) and subsequently fragmenting them to generate the MS2 scans. Cycle times can be selected by choosing the number of peaks to fragment (N) and – in Orbitrap analysis – the transient times for

MS1 and MS2 scans. To ensure fragmentation of peaks with very short elution times, they should not exceed one or a few seconds.

1.2.8. Fragmentation strategies

The selection of appropriate fragmentation methods, such as collision-induced dissociation (CID), higher energy collisional dissociation (HCD), or electron-transfer dissociation (ETD) is important for the optimal generation of sequence ion series required for peptide identification and unambiguous PTM site assignments. The peptide or protein precursor ions are positively charged, with protonation sites usually at the amino-terminus and the basic amino-acid residues and in CID, they undergo collisions by interactions with inert gas molecules, such as helium. This induces vibronic activation, leading to peptide bond dissociation and generating primarily N-terminal b- and C-terminal y-type ions^{50,51} (Figure 4). The CID process in ion trap is generally more effective for small and low-charge state peptides but is strongly influenced by the amino acid sequence and the distribution of the positive charges along the peptide backbone. HCD is a similar strategy of fragmentation as CID but is characterized by higher activation energy compared to CID. The higher fragmentation efficiency for HCD predominantly results in y-type fragment ions⁵².

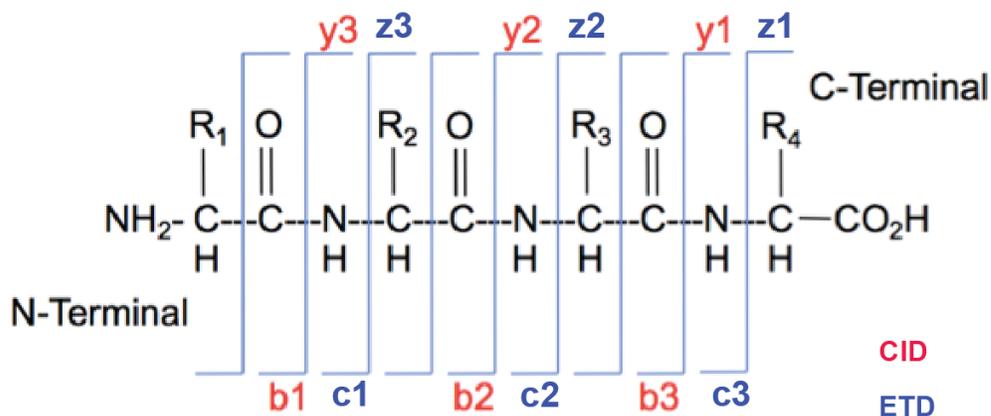


Figure 4 Different fragmentation strategies lead to formation of different ion species. While CID and HCD based fragmentation generate b- and y-type ions, ETD leads to the formation of c- and z-type ions (adapted from⁵³).

When comparing HCD coupled to an Orbitrap analyzer to ion trap fragmentation and detection, HCD produces higher quality spectra because of the superior resolution and mass accuracy. Spectral acquisition times, however, are longer compared to CID, because more ions need to accumulate to generate a signal by image current detection in the Orbitrap. For phosphoproteomic analysis, both CID and HCD induce so called ‘neutral losses’, meaning that uncharged phosphorylation moieties are cleaved from their precursor peptides, creating a -98 Da (H₃PO₄) mass shift. It is debated which of CID or HCD is more appropriate for phosphorylation analysis, but clearly HCD improves the formation of rich fragment ion spectra for phosphopeptides⁵⁴. ETD achieves fragmentation through neutralization of backbone protonation sites with radical anions, used as the electron transfer species, but generally at lower efficiency than CID or HCD. The resulting random nonergodic N–C α backbone bonds breaks generate c- and z-type fragment ions⁵⁵. ETD is more effective for large, multi-charge state peptides and is particularly suitable for detecting labile PTMs because peptide backbone fragmentation is virtually independent of the amino acid sequence. For example, O-GlcNAc elimination does not occur using ETD⁵⁶.

1.2.9. Mass spectrometer types

Major developments during the last decade have led to new high performance instrumentation that provide both high resolution and high mass measurement accuracies for MS1 and MS2 levels. Resolution is the ability to distinguish two peaks of different m/z ratio (modern instruments can resolve peaks down to 1 mDa at mass 1,000), whereas mass accuracy describes the difference between measured and theoretical mass. The high quality mass spectra that are typical today have increased the reliability and efficiency of protein identification at the peptide level.

In this thesis, all measurements were performed on Thermo Fisher Scientific Orbitrap mass spectrometers, which currently comprise six different instrument types including LTQ-Orbitrap, Orbitrap Velos and Elite, Exactive, Q Exactive, and Orbitrap Fusion.

The LTQ-Orbitrap, Orbitrap Velos, and Orbitrap Elite are hybrid configurations where low-resolution linear ion-traps are combined with high-resolution Orbitraps analyzers. The Orbitrap Velos was equipped with a novel ion source that replaced the previous tube lens with a radiofrequency (RF) driven S-lens, enabling 10-fold better ion transmission. Further improvements of the Orbitrap itself were implemented in the third generation of hybrid mass spectrometers, the Orbitrap Elite. It was equipped with a compact high-field Orbitrap analyzer, where the inner diameter of the outer electrode was reduced from 30 to 20 mm, yielding twice the resolving power. The Exactive consists only of an Orbitrap analyzer and is mainly used for small molecule analysis. In contrast, Q Exactive type instruments are additionally equipped with a quadrupole enabling ion selection, isolation and fragmentation upstream of the Orbitrap. In this type of instrument, the Orbitrap is the only mass analyzer, where MS1 and MS2 scans are always measured with high resolution. Due to its simple design and excellent performance, the Q Exactive has become an instrument of choice for proteomics in general (Figure 5). It is a benchtop mass spectrometer, which is an important step to move MS towards clinical applications. Finally, the Orbitrap Fusion (Lumos) is a high end instrument that can

perform CID/HCD/ETD, detect intact and fragmented peptides in the ion trap and perform multi-stage fragmentation (MS3 and higher), Orbitrap or both.

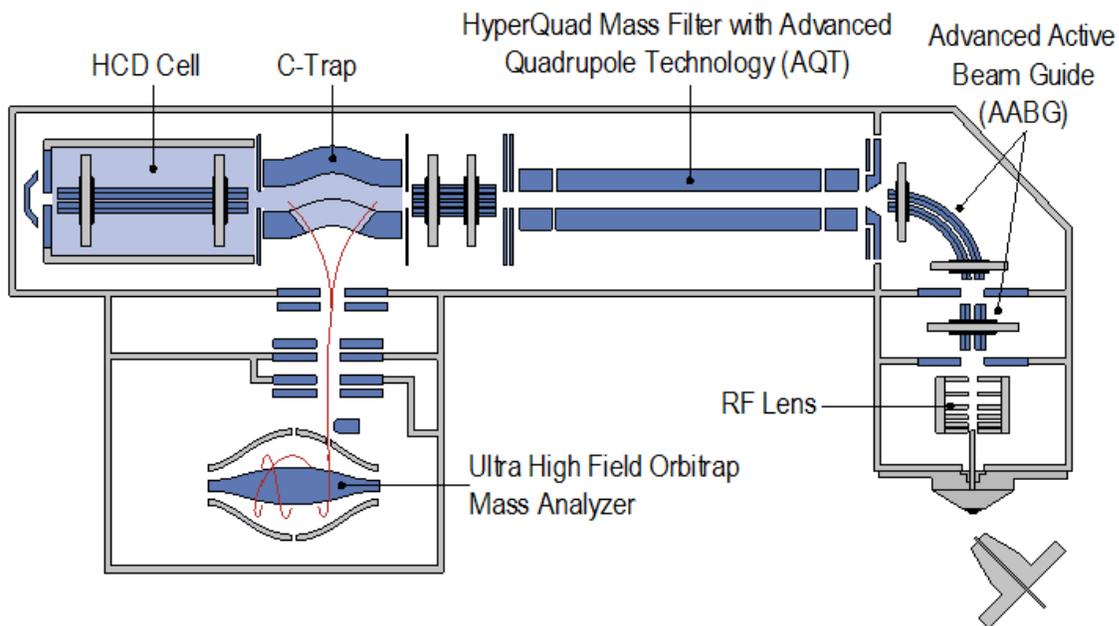


Figure 5 Functional elements in the Q Exactive HF mass spectrometer⁵⁷.

1.2.10. Acquisition methods

In bottom-up proteomics, three main acquisition strategies are used². The topN method described above is a data dependent acquisition (DDA) strategy that has been the mainstay of hypothesis-free shotgun (discovery) proteomics. In contrast, targeted proteomics methods is used for acquiring a predefined set of peptides. They monitor specific precursor-fragment transitions and come in flavors such as single or multiple/parallel reaction monitoring SRM, MRM and PRM. Finally, data independent acquisition (DIA) acquisition cycles through relatively large mass windows to generate complex MS2 scans that cover all peptide precursors. It is an emerging technology that has the advantage of generating comprehensive fragment-ion maps.

In more detail, a typical, top10 acquisition cycle on the Q Exactive HF instrument consists of one MS1 scan followed by 10 MS2 scans and takes roughly one second. Spectra are collected in the Orbitrap mass analyzer and the top 10 most intense ions of the full scan are selected by the quadrupole with an isolation width chosen to balance ion transmission and specific fragmentation of the intended precursor (typically 1.4 m/z on our current instruments) for subsequent fragmentation in the HCD collision cell. Likewise, an optimal fragmentation energy is chosen to yield good fragment coverage and high MS2 identification rate ('normalized collision energy' of 27% on a Q Exactive HF instrument)⁵⁸. To prevent the re-fragmentation of peptides, precursors with the same mass are excluded from resequencing for about 30 sec, more than the time taken for a typical peptide to elute from the HPLC column. The resulting spectra contain information about the m/z values, retention times and ion intensities for all the detected fragment ions. They are submitted to software packages like MaxQuant⁵⁹ to extract peptide information for identification and quantification (see below). Due to the semi-stochastic selection of precursors at the MS1 level in DDA, some precursors are not fragmented in every LC-MS/MS runs, leading to missing values, which can lead to problems in downstream analysis. This occurs mainly for low abundance peptides. To overcome this challenge, MaxQuant has a 'match between runs' feature which transfers identifications from runs where a peptide was sequenced to another where it was not, based on the m/z ratios and retention times of the MS1 features^{60,61}. This way, if in a given LC-MS run the required MS2 scan is not present or not interpretable, it can be transferred from another LC-MS run. This matching strategy is particularly powerful in challenging, high-dynamic range proteomics such as plasma⁶² and the heart muscle (Article 6).

A more recent acquisition method, termed 'BoxCar' further boosts the depth of primarily high dynamic range proteomes⁶³. The capacity of the C-trap is limited to about one million charges^{54,64}, therefore high abundant ions often fill the C-trap in a very short time (<1 ms), effectively displacing low abundance peptides. This 1 ms corresponds to less than 1% of the transient time for a high resolution mass spectrum (128 ms for 60,000 resolution). These observations imply that 99% of the generated ions are not used for mass analysis at the MS1 level. BoxCar increases

the dynamic range at the MS1 level by dividing the mass range into many segments, which are sequentially filled (typically 30 segments in three separate MS1 scans). This maximizes the usage of the incoming ion current by giving low abundant ions longer injection times compared to high abundant ions. Consequently, fewer missing values are observed using BoxCar, which makes optimal use of matching identifications from a library at the MS1 level. In the heart atlas (Article 6) we applied BoxCar for the first time in the context of a translational study.

There are several implementations of DIA⁶⁵, the most well-known of which is 'sequential window acquisition of all theoretical mass spectra' (SWATH), in which ranges of precursors (typically 25 Da windows) are selected and fragmented together⁶⁶. This results in complex fragment ion mixtures deriving from different precursors. Until recently, SWATH was limited to the detection of only the most abundant part of the proteome, for instance 2,000 proteins in cancer tissues⁶⁷. Recent developments have made DIA acquisitions much more competitive. They have recently been shown to largely eliminate the missing value problem and reproducibly quantify protein abundances, reaching CVs down to 5% in technical triplicates⁶⁸. However, DIA generally relies on information from pre-existing high-quality spectral libraries. Both DDA and DIA are discovery-oriented and unbiased.

The third bottom-up acquisition mode is targeted proteomics (SRM/MRM/PRM^{69,70}) whose goal is to detect a limited set of peptides with high reproducibility and specificity. Targeted measurements require the creation of an inclusion list of predetermined transitions (precursor/product ion pairs) and peptides. SRM and MRM measurements are performed on triple quadrupole instruments (QQQ), where the first selects the ions, the second contains the ions during fragmentation, and the third quadrupole detects one (SRM) or more (MRM) product ions. MRM can yield very reproducible and sensitive but suffers from limited specificity as a result of the poor resolution and in practice is restricted to monitoring a small number of peptides. In contrast, PRM is performed on a quadrupole-Orbitrap mass spectrometer offering a clear advantage in terms of the high resolution, trapping, and high-throughput

capabilities of this instrument. In article 3, we describe a histone PTM inclusion list for targeted PRM measurements using the Skyline software for quantitation.

1.2.11. Protein identification and quantification

A key step in the shotgun proteomic workflow is the identification of proteins, which relies on the interpretation of MS2 product ion spectra. Each peptide will generate specific peptide fragment ions forming N-terminal (b-ions) or C-terminal (y-ions) sequence ladders. With a complete series, the entire peptide sequence can be assigned ('de novo' sequencing) (Figure 6). More commonly, the series are incomplete and identification of spectra is carried out by database searches using peptide search engines, such as Mascot or Andromeda^{71,72}. Both are based on matching experimental to theoretical MS2 spectra that have been obtained through *in silico* digestion of all proteins of a given organism⁷³. To control for false positive hits (i.e. spectra that are assigned the wrong peptide), spectra are matched to a database that includes each peptide in the *in silico* digest in both the true amino acid order and reverse order. The resulting numbers of hits to the reversed database can be used to define a false discovery rate (FDR), which is usually set to less than 1%.

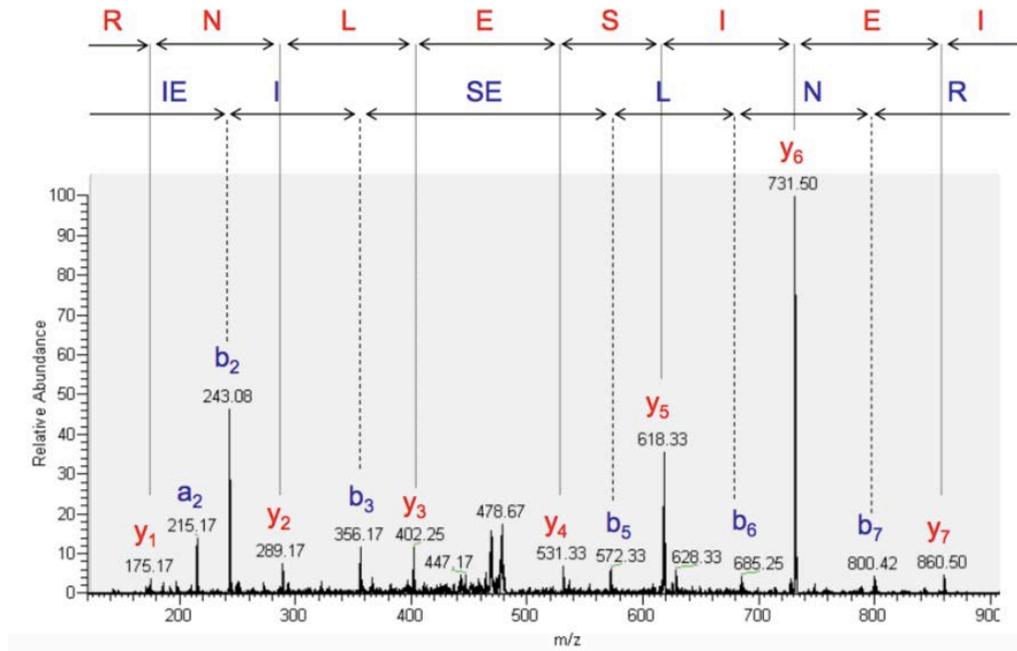


Figure 6 De novo interpretation of an MS/MS spectrum acquired in an ion trap, assigning the sequence IEISELNR⁷⁴.

Next, peptide identifications are assembled into protein identifications, which is also FDR controlled. Some peptide sequences are not unique to a specific protein, such as those that match several protein isoforms. Following Occam’s razor principle – that the simplest explanation is the most likely – they can be assigned to the protein sequence that already contains the most unique peptides and are then called ‘razor peptides’.

Protein identification is only a first step in gaining insight into the complexity of a biological system. Proteins span an abundance range of more than six orders of magnitude (or even more than ten orders in plasma and muscle), which already makes it clear that quantitative information about protein abundance changes is indispensable in the study of biological systems. Quantitative information can be obtained in a ‘relative’ or ‘absolute’ fashion, where protein abundances are either compared to each other or their concentration or copy number is determined. Absolute quantities can be estimated indirectly or – potentially most accurately – measured by comparison to a spiked-in absolute reference. Many quantification strategies have been developed (Figure 7) and at the highest level they can be

categorized in those using stable isotopes and those relying on 'label-free' approaches.

1.2.12. Label- and label free-based quantification

Label-based quantification methods involve the incorporation of stable isotopes either metabolically, by supplying labeled compounds, or chemically via conjugation of stable isotope tags to peptides or proteins. Introduction of these stable isotopes generally do not affect the physical behavior of a given (tagged) peptide in terms of elution profiles or physical properties. In-vivo metabolic labeling, most commonly by stable isotope labeling by amino acids in cell culture (SILAC), introduces a defined mass shift between the samples that can be distinguished at the MS1 level⁷⁵. For example, labeling can be achieved by providing heavy $^{13}\text{C}_6$ $^{15}\text{N}_2$ -lysine and $^{13}\text{C}_6$ $^{15}\text{N}_4$ -arginine (K8R10), which introduces an 8 or 10 Da mass increase for a labeled lysine- or arginine-containing tryptic peptide, respectively. SILAC experiments can be extended to a triplex format, using 'medium' $^2\text{H}_4$ -lysine and 'heavy' $^{13}\text{C}_6$ - arginine (R6K4). The abundance differences of the proteins are determined from the relative intensity of the corresponding heavy and light labeled peptides. In metabolic labeling, samples are combined up-front, thus it has the advantage to directly correcting for any sample preparation biases at the LC-MS level. However, this method suffers from reduced peptide identification because of increased spectral complexity at the MS1 level, and it is only applicable to cellular and certain mammalian systems, such as the SILAC mouse or fly or plants⁷⁶⁻⁷⁹. Much smaller changes (in the mDa range) can be introduced in a SILAC variant called Neutron-encoded (NeuCode) SILAC labeled samples⁸⁰. Other extensions of the classical SILAC approach, include spiking in entire labeled proteomes (mix of cell lines (super-SILAC)⁸¹, or SILAC-labeled protein epitope signature tags (PrESTs)⁸². For absolute quantification, protein standard absolute quantification (PSAQ)⁸³, or peptides (AQUA peptides)) can also be used as internal standards, but this is limited to a moderate number of proteins of interest. In the SILAC-PrEST approach, a

known quantity of recombinant heavy-labeled standards is added into each of the experimental samples, which then are processed and analyzed together. Peptide ratios between samples and the heavy-spike-in standard are readily calculated and absolute quantification of up to 40 proteins of interest has been reported⁸². Other methods for absolute quantification include the estimation of copy numbers per cell using the 'proteomic ruler' approach, which relies upon the fixed relationship between histones and DNA allowing MS-signals to be placed on an absolute scale⁸⁴.

Chemical labeling is a strategy to incorporate stable isotopes, and can be used for any sample, including primary tissues and body fluids. Here, labeling is generally performed by derivatization at the peptide level. One such method is dimethyl labeling that has three available channels due to conversion of each primary amine to a secondary amine with isotope labeled formaldehyde⁸⁵. More popular strategies for chemical labeling, simultaneously overcome the spectral complexity problem, and include isobaric tags for relative and absolute quantification (iTRAQ, 8 channels)⁸⁶ and tandem mass tags (TMT, 11 channels)⁸⁷. These tags are composed of three main functional parts, an amine-reactive group for attaching the label to the N-terminal amine groups of lysine residues of the amino terminus of the peptides, a reporter ion group for relative quantification, and a mass balancer group so that peptides share the same MS1 mass. These tags offer greater multiplexing without increasing spectral complexity and can decrease measurement time by the degree of multiplexing. The MS2 spectra obtained from fragmentation of isobarically labeled peptides contain two types of product ion peaks: reporter ion peaks that reflect the abundance of the input material in each channel, and the peptide fragment ion peaks that permit identification. While TMT-based multiplexing greatly increases the overall throughput of sample measurements, it suffers from 'ratio compression', where the actual ratio between channels are underestimated due to reporter ions from co-isolated and co-fragmented 'contaminating peptides'^{88,89}. This phenomenon is a result of the relatively low resolution of quadrupole isolation of precursors. In the low-mass reporter ion region of the fragmentation spectra, the target peptide and the contaminating peptide give rise to identical reporter ions, 'compressing' the actual ratios resulting in an underestimation of peptide abundance differences. To

partially overcome this challenge, the isolation window can be narrowed, but requires high performance quadrupoles⁹⁰. Alternatively, the tag-containing fragments of the precursor can be further isolated and fragmented but this requires mass spectrometer capable of MS3, albeit at the expense of sequencing speed and proteome coverage. Isobaric tags generally also generate an ion species containing the intact peptides and the tag without the low mass reporter ion⁹¹. Quantification using these ‘complementary reporter ions’ does not suffer from ratio compression⁹². With future improvements in tag chemistry and instruments, this strategy holds great promise for accurate measurement of proteins changes at high throughput.

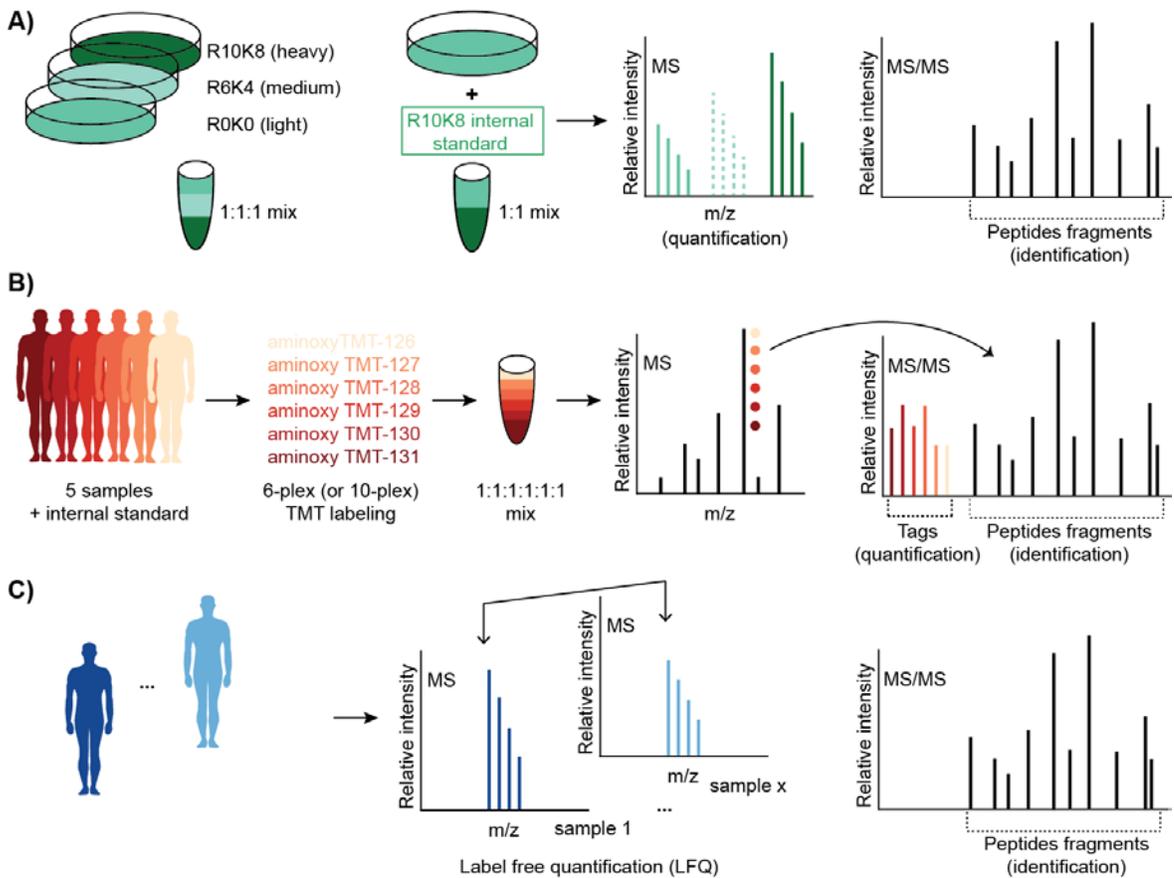


Figure 7 Different MS-based quantification strategies. A) SILAC based quantification is performed at the MS1 level. B) TMT-based quantification can currently be multiplexed to 11-fold and here quantification is performed at the MS2 level and based on reporter ions. C) Label free quantification can be done at the MS1 or MS2 level. In all strategies, the peptide identification is performed after peptide fragmentation at the MS2 level.

As label free quantification (LFQ) completely omits additional chemical reagents and procedures it is by its nature the simplest and most economical approach. Early label-free quantification methods rely the fact that the number of peptide spectrum matches tends to correlate with protein abundance, and included 'spectral counting' and the emPAI method⁹³. More recent methods are greatly superior as they directly quantify the intensity of the precursor ions by measuring the area underneath the extrapolated curve of the precursor intensities in MS1 scans over the LC peak. In combination with sophisticated algorithms, such as MaxLFQ this has been shown to yield very accurate quantification, especially on high-resolution instruments. Since sample preparation and measurement is performed for each sample individually, the reproducibility of LFQ is usually worse than labeled-based quantification. To overcome this, our group has developed automated sample preparation workflows^{24,62}, which together with MaxLFQ largely eliminate potential technical variations introduced during sample preparation and MS measurements. LFQ, together with robust and streamlined sample preparation workflows and the unlimited number of samples that can be compared, has become a popular approach.

1.2.12. Nearly complete proteomes

Significant improvements of all steps of today's MS workflow, encompassing sample preparation to measurement and subsequent bioinformatics analysis, have enabled the characterization of nearly complete proteomes^{7,46,94}. More than 10,000 proteins can be routinely quantified in cell lines, model organisms and even in very challenging tissues, such as heart muscle (Article 6). In a recent study⁷, Notably the recent report of a very deep HeLa proteome also showed that the depth of proteomics is now on par with RNA-based methods. The authors combined extensive high pH reversed-phase fractionation with short LC-MS/MS gradients, resulting in a sufficient number of peptide precursors to saturate the sequencing speed of modern MS instruments. Using the aforementioned method, more than

12,000 proteins (and >14,000 protein isoforms) were catalogued. Moreover, more than 7,000 acetylation sites and 10,000 phosphorylation sites were identified and quantified, even without any enrichment. Other proteomic studies have explored key regulatory mechanisms at a large scale. For example, temporal changes in phosphorylation have been investigated, on both long and short timescales^{30,95,96}. 'Organellar proteomics' can also be employed to determine subcellular localization on a system-wide scale^{97,98}, and more recently this was done in a dynamic fashion to monitor subcellular localization changes⁹⁹. At higher resolution, interaction partners of specific proteins can be uncovered through immunoprecipitation experiments followed by MS. Global application of interactomics has resulted in draft maps of the human interactome, an extensive network analysis of thousands of proteins^{100,101}. Finally, the integration of measurements from several types of human tissue proteomes combined with data generated by the community resulted in two 'drafts of the human proteome'^{102,103}. These drafts, however, were very incomplete; to illustrate this, our human heart atlas identified three times as many cardiac proteins – mostly of low abundance. Furthermore, the total number of proteins identified in these draft proteome studies is overestimated since their FDR is unusually high¹⁰⁴. Nevertheless, these human proteome drafts illustrate the desire of the community to determine the complete proteome to better understand the complex protein composition in the human body.

2. The cancer proteome at the nuclear, cellular and tissue level

'Cancer begins and ends with people' – June Goodfield

2.1. Cancer

The emperor of all maladies¹⁰⁵, cancer, accounts for approximately nine million deaths worldwide annually¹⁰⁶. About one in three women and one in two men will develop cancer during their lifetimes. In some regions, cancer is even predicted to surpass cardiovascular diseases. Cancer is the Latin word for Crab, this association was penned by Hippocrates around 400 B.C. most likely because of its finger-like projections that are reminiscent of the outline of a crab. Malignant cell growth, which is dictated by 'hallmarks of cancer', is the fundamental feature that is shared between these cells^{107,108}. Cancer incidence has risen significantly over the last decades. With longer life expectancies, it has become evident that the risk of developing cancer correlates with age. For instance, a 30-year-old woman has a one in 400 annual risk to develop breast cancer compared to one in nine for a seventy-year-old. Our capabilities to detect cancer at early stages have much increased in the last century. Introduction of mammography screening in the early 1980s, was followed by higher reported breast cancer incidences¹⁰⁹. Changes in modern life styles have also influenced cancer statistics. For example, lung cancer incidence has exploded in the 1950s, correlating with the increase of cigarette smoking and this cancer still remains the most common cause of cancer death. There is also a significant link between obesity and cancer, as approximately one in five cancer deaths are associated with obesity¹¹⁰.

Many cancer subtypes, in particular rare cancers, remain poorly understood and conventional chemotherapeutic treatments are still associated with life-threatening toxic side effects¹¹¹. This is primarily due to a lack of specificity towards cancer cells or the known or suspected molecular drivers. Only a small percentage of patients will benefit from the treatment, and the number of patients that need to be treated

before one of them actually benefits from the treatment (NNT) remains strikingly high (>40 in prostate cancer, for example¹¹²). Chemotherapy, specifically, has serious side effects, to which a substantial percentage will succumb. These numbers underscore the urgent need to develop more effective medicines, and uncover predictive biomarkers that will help to stratify patients and target those most likely to respond to a specific therapy.

2.2. Personalized cancer medicine

The concept of precision or personalized medicine was already born thousands of years ago when Hippocrates said, '*It's far more important to know what person the disease has than what disease the person has*'. Today, the goal of personalized medicine is to individualize clinical decisions, thus distinguishing patients that are most likely to respond and benefit from a given treatment from those who will only suffer from detrimental side effects without benefit while still incurring health care costs. The sequencing of the human genome and ongoing rapid technological developments have set in motion the transformation of personalized medicine from an idea to practice. For a small subset of patients, this has already become reality. A pioneering study showed that imatinib (Gleevec, Novartis)¹¹³ specifically inhibits the fusion protein BCR-ABL, which drives chronic myeloid leukemia. This drug led to patient survival in about 90% of patients¹¹⁴. Another example is the monoclonal antibody Trastuzumab (Herceptin, Genentech), prescribed for HER2-positive breast cancers where patients show improved survival¹¹⁵. Monoclonal antibodies targeting EGFR have also shown clinical efficacy, for instance in patients with EGFR-expressing metastatic colon cancer¹¹⁶.

While next generation sequencing has driven the field since the turn of the millennium, and allowed the molecular profiling of thousands of tumors in different cancer types¹, it has become evident that the development and complexity of cancer does not lie in genetic changes alone. Clearly, epigenetic changes, protein expression alterations, and aberrant PTMs play essential roles in the development

of cancer. Thus, the use of MS-based proteomics holds great promise to unmask the cause and highlight potential therapeutic targets in the next generation of treatment.

2.3. Nuclear level: The epigenetic landscape in oncology

2.3.1. Epigenetics

The nucleus is at the heart of each cell. Genetic information is stored in the form of chromatin, which consists of DNA that is tightly wrapped around octameric histone proteins (H3, H4, H2A, and H2B), forming nucleosomes. Modifications at the histone level can result in changes in gene expression even without alterations in the DNA sequence. These 'epigenetic changes' are frequently observed in tumors^{117–120}. DNA itself can also be modified, resulting in hyper- or hypomethylation at the global or local DNA level at certain CpG islands of promoter regions, and these are commonly altered epigenetic patterns of cancers.

Histones contain lysine-rich N-terminal tails whose positive charges interact with the negatively charged phosphate backbone of the DNA. The structure of this nucleosome complex can be modified by multiple PTMs, including acetylation and methylation of histone tails that influence gene expression. In general, tightly packed chromatin (heterochromatin) is associated with histone methylation and gene-silencing, whereas open chromatin (euchromatin) is associated with histone acetylation and gene-expression. These acetyl and methyl marks present another regulatory mechanism of the epigenetic machinery. Proteins that catalyze ('writers'), recognize ('readers') or reverse ('erasers') the transfer of histone marks (Figure 8) are frequently altered in tumors.

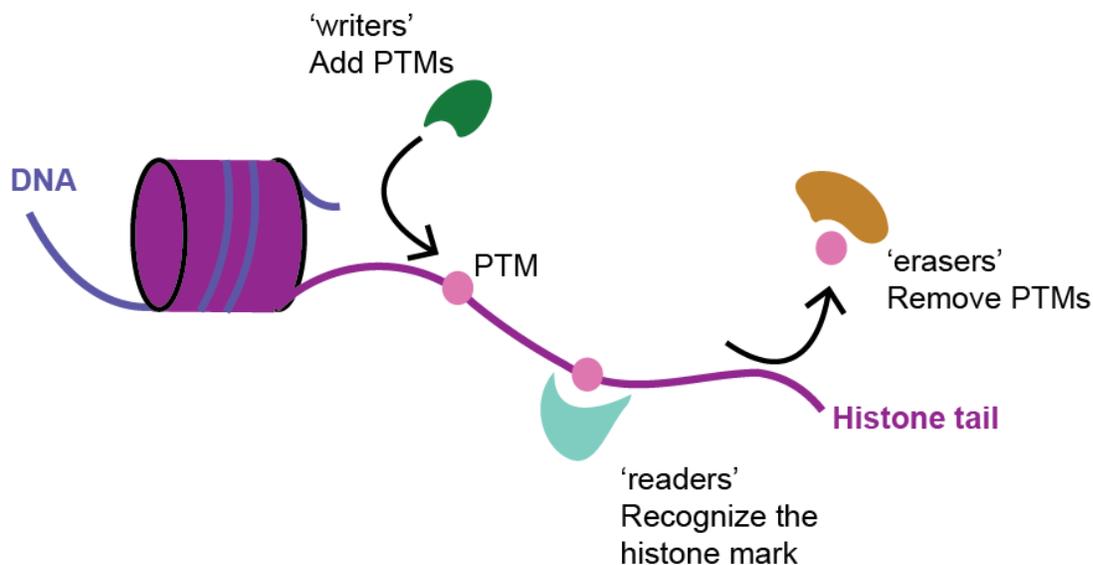


Figure 8 The main epigenetic regulators can be classified into three categories: epigenetic ‘writers’ (enzymes that deposit covalent modifications on histone tails), ‘readers’ (proteins that bind to histone modifications), and ‘erasers’ (enzymes that catalyze the removal of modifications on histone tails)

2.3.2. Contributions of MS-based proteomics to epigenetics

MS-based proteomics has become an invaluable method to study epigenetic regulation in an unbiased way. While conventional proteomic workflows are not suitable for lysine-rich and hydrophilic histone tails, specialized protocols have been developed and are in routine use^{28,29,121}. For bottom-up histone tail analysis, optimizations include the propionylation of histones to increase sequence coverage and peptide-level phenyl isocyanate labeling to improve HPLC retention and the detectability of hydrophilic peptides. For the more complete analysis of combinatorial histone PTMs, relatively long amino acid sequences (~50-mers) can be directly analyzed via ‘middle-down’ proteomics¹²². To this end, intact N-terminal tails are generated by Glu-C or Asp-N mediated digestion of histone H3 or H4, respectively. Middle-down approaches, however, still suffers from relatively poor sensitivity in comparison to bottom-up workflows and the interpretation of MS2

spectra remains computationally intensive and laborious. Adding ion mobility separation to MS-based strategy holds great promise for the analysis of challenging histone PTM combinations, such as H4 acetylations, which are chromatographically almost indistinguishable^{123,124}. Here, ions are separated by their mobility in a low pressure gas under the influence of an electric field. This mobility is determined by their size-to-charge ratio (collisional cross section). Together, these diverse MS-based workflows have enabled the accurate description of histone variants such as the nearly identical histone variants H3.1 and H3.3. Multiple novel histone marks, such as O-GlyNAc, butyrylation, crotonylation, citrullination, and formylation have been described on histones¹²⁵. The more well described marks, including histone acetylation and methylation can now be reproducibly and accurately quantified²⁹. In addition to these histone marks, we show in article 3 that histone butyrylation alterations can be quantified. Other examples of successful application of MS-based workflows to epigenetics include the discovery of the effects of the histone demethylase KDM5 inhibitors¹²¹, the role of macroH2A in melanoma¹²⁶, and quantification of histone PTM dynamics in cellular systems¹²⁷. In addition, cross-talks between different sites and types of histone PTMs have been described and histone marks triggering the recruitment of specific epigenetic ‘writer’^{128,129}.

2.3.3. The histone methyltransferase enhancer of zeste homolog 2

The epigenetic writer enhancer of zeste homolog 2 (EZH2) is the driving catalytic subunit of the polycomb repressive complex 2 (PRC2) (Figure 9). As a methyltransferase, it tri-methylates histone 3 on lysine 27 (H3K27me3), a transcriptionally repressive epigenetic mark that silences gene expression. It has been found to inactivate multiple tumor suppressor genes, such as CDKN1C^{130,131}. In a second mode of action, EZH2 promotes cell proliferation via STAT3 methylation¹³² or BRAC1 inactivation¹³³. EZH2 is overexpressed in multiple human cancers, including lymphoma and its inhibition has been shown to induce apoptosis

of cancer cells^{134–136}. Adding to these findings, we show in article 2 that cell cycle regulators are down-regulated at the proteomic level following H3K27me3 demethylation in lymphoma cells.

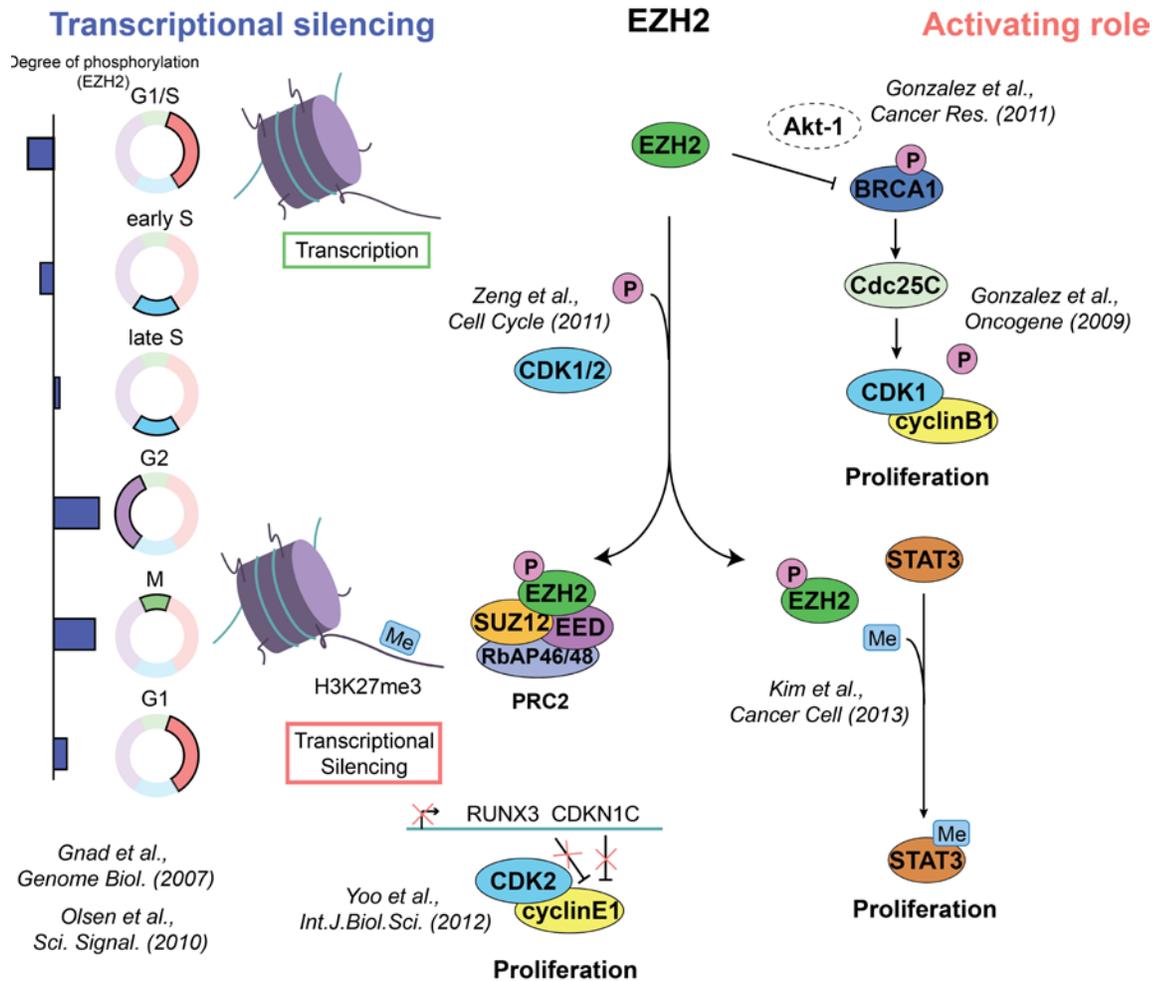


Figure 9 Model of the role of EZH2 in cell cycle regulation. By repressing transcription as a member of the PRC2 complex (left panel), EZH2 enhances the expression of cell cycle regulators indirectly by repressing associated tumor suppressors. EZH2 can also act as a direct activator of cell proliferation following phosphorylation (right panel). In its phosphorylated form, EZH2 activates STAT3 via methylation, which subsequently induces cell proliferation. EZH2 can also phosphorylate BRCA1 via Akt1 interaction, resulting in cell cycle activation as well (Figure from article 2).

2.3.4. Histone demethylases

Histone demethylases are epigenetic erasers and come in two main varieties, (i) jumonji (JmjC) domain-containing demethylases, and (ii) the lysine-specific demethylase (KDM1A/LSD1) family. Both classes play important roles in cancer¹³⁷.

2.3.4.1. JmjC domain-containing histone demethylases

JmjC domain-containing histone demethylases are the largest class of histone demethylases and employ an oxidative mechanism that is Fe²⁺ and alpha-ketoglutarate (α -KG) dependent. They can remove all three histone lysine methylation states, unlike the LSD1 family, which can only remove mono- and dimethyl lysine modifications. In the context of glioblastoma, which is among the deadliest human cancers, we and others have shown that α -KG-dependent demethylases are inhibited in glioblastoma cell line models that harbor somatic isocitrate dehydrogenase 1 (IDH1) mutations (Article 3)¹³⁸. Clinically, there are primary glioblastomas (which rapidly progress and develop de novo) and secondary glioblastoma (which progress slowly after initially presenting as low-grade gliomas). Interestingly, the majority of secondary (>70%) but rarely primary glioblastomas harbor IDH1 mutations, involving Arg132 in nearly all cases¹³⁹. IDH enzymes are key regulators of the TCA cycle by converting isocitrate into α -KG. The monoallelic IDH mutation, however, generates an enzyme with neomorphic ability to convert α -KG into 2-hydroxyglutarate (2-HG). As a result, the oncometabolite 2-HG accumulates to very high levels in IDH mutant tumors and inhibits α -KG dependent histone demethylases and DNA demethylases^{140,141} (Figure 10). This subsequently results in increased global DNA hyper methylation¹³⁸ and down-regulation part of the proteome as shown in article 3. These observations suggest that different therapeutic interventions should be used in primary versus secondary glioblastoma.

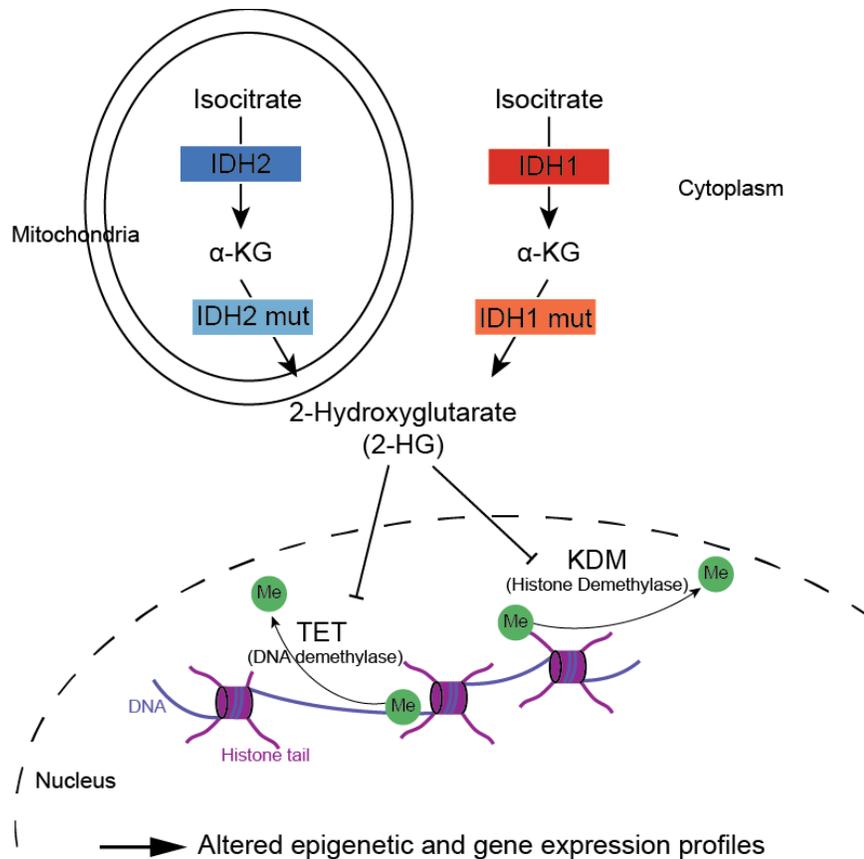


Figure 10 Mechanism by which IDH mutations lead to alterations at the epigenetic and gene expression levels (Figure from article 3).

2.3.4.2. Lysine-specific histone demethylases

The lysine-specific histone demethylase 1 (LSD1, also known as KDM1A) is a flavin adenine dinucleotide (FAD)-dependent amine oxidase. LSD1 was the first histone demethylase to be discovered¹⁴². Acting as an epigenetic eraser, it demethylates mono- and di-methylated lysines (H3K4 and H3K9), thereby acting as a coactivator or corepressor, depending on the context¹⁴³. LSD1 has emerged as an interesting therapeutic target because it is overexpressed in many cancer types, including lung cancer^{144,145}. We also found that LSD1 is upregulated in the extremely rare and little studied urachus carcinoma cancer (Article 7). Furthermore, inhibition of LSD1 inhibits tumors that gave rise to the development of multiple anti-LSD1 drugs that

are currently in clinical trials¹⁴⁵⁻¹⁴⁷. Tranylcypromine, an FDA-approved antidepressant¹⁴⁸, arose particular interest in the context of cancer treatment because it revealed a strong LSD1-specific inhibitor side effect^{149,150}. Tranylcypromine and derivatives of this drug showed clinical efficacy for several indications, including the treatment of AML and are currently in clinical trials¹⁵¹.

2.4. Cellular level: The phosphorylation-based signaling landscape in oncology

2.4.1. Phosphorylation

Phosphorylation is the replacement of the neutral OH of the side chains of amino acids serine, threonine, or tyrosine by the negatively charged phosphoryl group (PO_3^-) and is subsequently converted to phosphate PO_4^{2-} (Figure 11). The negative charge can induce conformational changes, modulate protein activity, and mediate or inhibit interaction with other proteins by providing a docking site. Kinases and phosphatases are enzymes that covalently modify proteins by either adding or removing phosphate groups. More than 500 putative protein kinase genes have been described, constituting about 2% of all human genes, that can be classified in a kinome tree¹⁵².

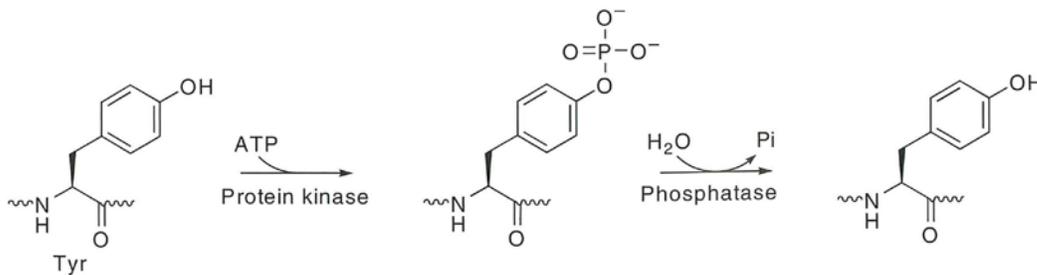


Figure 11 Phosphorylation and dephosphorylation mechanisms. The phosphorylation of protein residues (serine, threonine or tyrosine) is catalyzed by protein kinases. The reaction of dephosphorylation is mediated by protein phosphatases¹⁵³.

Cellular phosphorylation signaling networks are complex interaction systems that connect sequence-specific kinases and/or phosphatases to their respective target proteins. A typical phosphorylation-signaling cascade starts with the stimulation of RTKs via growth factors. This leads to their phosphorylation and activates a

phosphorylation cascade via adaptor proteins, for instance those containing a Src homology 2 (SH2) domain that specifically bind phosphorylated tyrosines.

2.4.2. The MAPK and PI3K pathways

The RAS-RAF-MEK-ERK (MAPK) and PI3K-AKT-mTOR (PI3K) pathways (Figure 12) were discovered about 30 years ago and are evolutionary conserved kinase families that control key cellular mechanisms, such as cell proliferation, survival, metabolism, and motility upon extracellular stimuli^{154–158}.

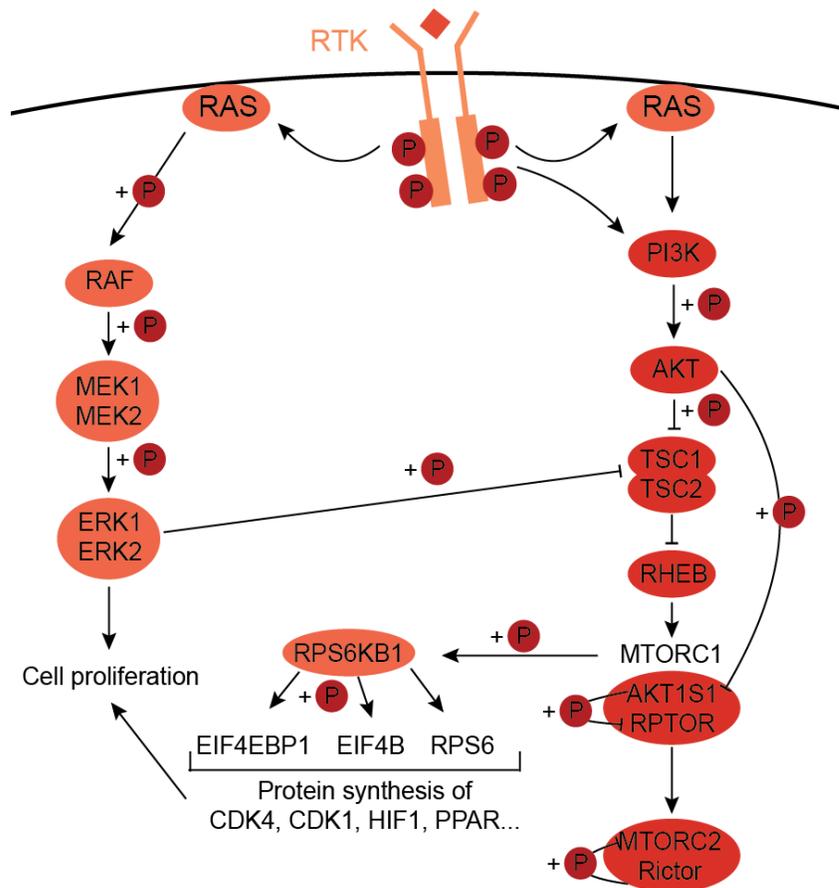


Figure 12 The MAPK and PI3K pathways are activated in multiple cancers driving cell proliferation, adapted from article 3.

When the master regulator RAS is activated it mediates a phosphorylation cascade including key kinases, such as mitogen-activated protein kinase (MAPK, also known as extracellular signal regulated kinase (ERK)) and phosphoinositide 3-kinase (PI3K). MAPK functions as the major effector of the RAS oncoprotein, which is a member of the large family of GTPases. The RAS oncogenes were initially discovered in murine sarcoma viruses^{159,160} and three human RAS genes have been described to date (H-RAS, N-RAS and K-RAS). They are anchored to the plasma membrane and function as molecular switches. Adaptors, such as the growth factor receptor bound protein 2 (Grb2) and the nucleotide exchange factor son of sevenless (SOS) enable the conversion of guanosine diphosphate (GDP) bound to RAS to guanosine triphosphate (GTP) in the cytoplasm¹⁶¹. This exchange allows the subsequent activation of RAF, MEK and MAPK. In its inactive form, MAPK presents a catalytic site that is blocked by a segment of amino acids, termed the lip. The binding of MEK to MAPK destabilizes the lip structure, leading to the exposure of a tyrosine and threonine that are subsequently phosphorylated by MEK. This results in conformational changes and enables the binding of ATP in the catalytic site of MAPK and its dimerization. Exclusively in this dimerized form, MAPK can be translocated to the nucleus where it regulates the activation of numerous transcription factors, such as MYC.

PI3K phosphorylates inositol membrane lipids to generate phosphatidylinositol 3,4,5-trisphosphate (PIP3) which in turn modulates the activity of intracellular protein effectors. PIP3 recruits pleckstrin homology (PH) domain containing proteins to the membrane, such as AKT1 and PDK1, which subsequently activate signaling cascades involved in cell growth and proliferation. Both pathways actively cross-talk. They can negatively regulate each other's activity, a phenomenon that is frequently observed when one of the two pathways is actively inhibited with a drug. For instance, MEK inhibition leads to the EGF-mediated hyperactivation of AKT¹⁶². Both pathways can also cross activate each other by regulating common downstream nodes, such as the TSC1/2 complex (Figure 12).

The MAPK and PI3K pathways are frequently deregulated in cancer because they play central roles in the control of cell proliferation, apoptotic pathways, oncogenic kinase signaling, and transcriptional regulation. Recent large-scale genomic sequencing initiatives of thousands of tumors through the TCGA consortium made the significance of both pathways very apparent on a global scale¹. This uncovered BRAF, KRAS, PIK3CA and PTEN as the most frequently mutated genes in cancer. In endometrial carcinoma, for example, PIK3CA showed somatic mutations in 53% of 240 tumors. Similarly, KRAS is the most frequently altered signaling node of the MAPK pathway, being mutated in about 40% of colorectal carcinomas¹⁶³ and BRAF mutations were found in 60% of thyroid carcinomas¹⁶⁴. These key regulators show the most significant 'mutation hot spots' in cancer¹⁶⁵, defined as the enriched occurrence of specific point mutations within the gene and characteristic for its cancer driving function. The most frequent hot spots are V600E in BRAF, and E545K in the helical and H1047R in the kinase domains of PIK3CA.

2.4.3. Development of kinase inhibitors

Protein kinases have become a major class of drug targets and today about 37 kinase inhibitors are FDA approved with an additional 250 in clinical trials¹⁶⁶. Among those, small molecules targeting the key nodes of the MAPK and PI3K pathways have been most intensely pursued for cancer treatment. While some, such as RAS cannot be directly targeted yet (#YetToBeDrugged)¹⁶⁷, others have advanced into the preclinical stage as validated targets. Specific MAPK and PI3K inhibitors include cobimetinib (GDC-0973), pictilisib (GDC-0941), and taselisib (GDC-0032) which are used in articles 3, 4 and 5. The oral, potent, and selective MEK inhibitor cobimetinib has been approved for the treatment of melanoma in combination with BRAF inhibitors^{168,169}. Pictilisib, is an oral, highly specific, ATP-competitive small-molecule class I pan-PI3K inhibitor¹⁷⁰. It has demonstrated significant antitumor activity in a wide array of cancer models in preclinical studies¹⁷¹⁻¹⁷³. Similarly, taselisib is an oral, potent and selective inhibitor of mutant PIK3CA. It is currently in clinical phases

I, II and III for the treatment of advanced solid tumors, lung and breast cancer, respectively^{174–176}.

Besides kinase inhibitors, other anticancer strategies use targeted protein degradation mechanisms, such as the proteolysis targeting chimeras (PROTAC) technology, where small molecules simultaneously bind a target protein and a ubiquitin ligase, enabling its ubiquitination and degradation in a generic way^{177,178}.

While targeted therapies that block signaling through the MAPK and PI3K pathways have shown clinical efficacy in several tumor types, the underlying signaling phosphorylation cascades often remain poorly understood. The core signaling axes of the two pathways span intricate networks, in which only a few substrates and connections are well characterized so far.

2.4.4. MS-based phosphoproteomics of the MAPK and PI3K pathways

Studying the dynamics of signal transduction networks in response to EGF stimulation revealed that different phosphorylation sites of the same protein often show distinct kinetics⁸. Hence, accurate modelling of a signaling pathway requires treating it not only as a network of proteins but also as a network of interconnected phosphorylation sites. Modifying specific nodes in this network by altering the activities of associated protein kinases or phosphatases, may significantly influence the entire signaling web and lead to unexpected effects. Only a few studies have attempted to explore the global spectrum of phosphorylation signaling downstream of the pivotal regulators of the MAPK or PI3K pathways^{179–182}. These efforts demonstrated the ability of MS-based proteomics to identify and quantify thousands of phosphorylation sites and explore perturbed signaling webs. While these studies primarily focused on the signal transduction through a specific protein kinase under fixed conditions, in articles 4 and 5 we compared the regulation of phosphoproteomes by mutating or inhibiting multiple gatekeepers of the most

important cancer pathways, and examined the dependence on inhibitor class, dosage, cell type, and type of kinase modulation (activation versus inhibition). In addition, we studied the impact of dual inhibition, since combinatorial therapies are thought to be superior to single agent treatment. This revealed interesting and previously unknown effects on feedback mechanisms. In the clinic, the efficacy of combined therapies, remains to be determined¹⁸³. MS-based analyses of convergence, crosstalk, and feedback associated phosphorylation patterns are poised to provide important scientific insights into the optimal uses of combination strategies to effectively block oncogenic pathways and overcome feedback mechanisms.

2.5. Tissue level: 'from bench to bedside'

2.5.1. Current cancer diagnosis in clinical practice

In a clinical context, blood sampling and tissue biopsies are the two most common means of biological sample collection. Today, the analysis of blood samples is dominated by protein-based clinical laboratory tests using single-protein immunoassays¹⁸⁴. About 42% of analyses requested by clinicians are based on proteins, compared to 35% for small molecules and 17% for cells. It is apparent that already today, proteins are the most frequently assayed analytes in clinical routine. Although the cancer field is dominated by genetic analyses, there are a number of established biomarkers for several common cancers. These were discovered and are routinely measured by non-proteomic technologies such as ELISAs. The only possible exception was the FDA approved blood test 'OVA1', which is approved in a narrow indication in the context of ovarian cancer (OvCa). OVA1 relies on a multiplexed immunoassay to measure the concentration of five individually non-specific proteins, including apolipoprotein 1 (APOA1), beta-2 macroglobulin (B2M), serum transferrin (TF), pre-albumin (ALB), and the known OvCa marker CA-125 (MUC16)¹⁸⁵⁻¹⁸⁸. These proteins, however, are in some cases not detected specifically enough using immunoassays and it is debated whether they are sufficiently specific to OvCa^{186,189,190}. Of note, using high pH reversed-phase fractionation and state of the art MS we can easily identify all OVA1 protein members with high accuracy in only 1 μ l of serum from an OvCa patient (Figure 13).

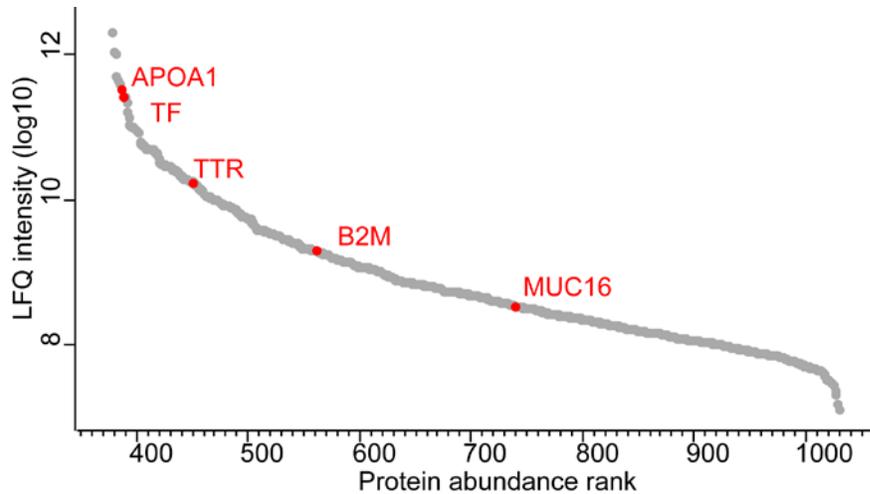


Figure 13 Protein abundance rank in OvCa serum. The five proteins used in the OVA1 test are highlighted in red (Data generated by the author).

For the detection and monitoring of the progression of prostate cancer, prostate-specific antigen (PSA) levels - a protein produced exclusively by normal and malignant prostate gland cells - are measured in the blood. PSA levels above 4 ng/ml are usually followed by a prostate biopsy and the higher the PSA levels, the higher the risk of diagnosing prostate cancer in the biopsy. Higher PSA levels, however, also correlate with benign enlargement of the prostate or infections, meaning that these test are not very specific (high false positive rates).

Tissue samples are generally obtained by needle biopsy or surgical excision and are subsequently analyzed by diagnostic pathology using hematoxylin and eosin (H&E) staining (Figure 14). Hematoxylin colors nuclei of cells in dark purple, whereas eosin mainly stains eosinophilic structures, including the cytoplasm, intra- and extracellular proteins in pink. Blood cells are stained intensively red. H&E staining enables the detection of irregular cell proliferation, stromal infiltration, and overall abnormal cellular morphology. In most cases, H&E alone enables the diagnosis and grading of cancer. Grading systems based on H&E staining, such as the Gleason score for prostate cancer exist since the 1960s and are still routinely used as an indicator of how quickly the tumor is likely to grow and spread.

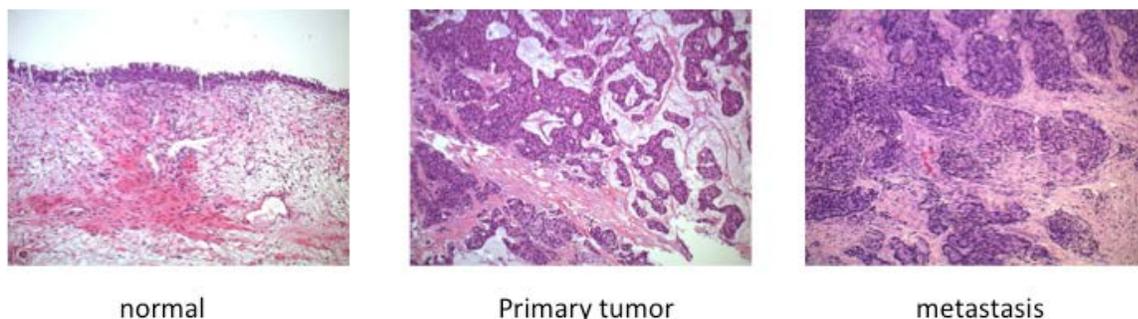


Figure 14 H&E stainings of healthy, primary tumor and metastatic tissues. Proliferating cells are colored in dark purple, such as the healthy epithelium (left panel) of cancer cells (middle and right panels). The prominent pink staining in the middle and right panels is representative of prominent stromal infiltration.

In some cases, further tissue stainings are performed to diagnose a particular cancer type/subtype or measure the likelihood of a patient to respond to a particular therapeutic treatment. To this end, immunohistochemistry (IHC), which utilizes specific antibodies for the detection of proteins in tissue sections is applied. For instance, current tissue-based FDA approved protein biomarkers include EGFR for the therapy selection of colon cancer or HER2 for breast cancer diagnosis and therapy selection. While these blood- and tissue-based tests are routinely used in the clinic, they may suffer from lack of specificity, are based on outdated technologies in some case, and generally do not enable the early detection of cancer development.

2.5.2. Cancer FFPE tissues proteomics

The most frequent method for human tissue preservation is formalin fixation and paraffin-embedding (FFPE). It is routinely used in tissue banks due to its long-term preservation capabilities, and amenability to downstream IHC. FFPE is an economical choice since samples can be stored at room temperature and at great density. It is estimated that about half a billion archived FFPE cancer tissue samples exist to date¹⁹¹. These immense archives of material in principle present an

invaluable resource for studying the underlying molecular mechanisms of cancer, testing known biomarkers and uncovering new ones. The main use of FFPE tissues in the clinic today is H&E and IHC stainings for the detection of known cancer aberrations. It is semiquantitative at best and allows the evaluation of only a few targets at a time. FFPE cohorts have been challenging to use in gene expression studies due to the difficulty in isolating nucleic acids, often resulting in samples with poor RNA quality that are not usable for next generation sequencing¹⁹². This was also observed in article 7, where RNA sequencing could not be performed because most of the RNA was degraded after extraction from FFPE and even from fresh frozen cancer tissue. Proteins are more stable than RNA or DNA, therefore protein profiling in high-throughput platforms, in principle holds great promise for uncovering new biomarkers and improving prognostic and predictive power for clinicians.

Taking advantage of the stability and ease-of-handling of proteins, protein extraction from FFPE material is possible in a robust manner for MS-based analysis. Proteomic analysis of almost 30-year old FFPE tissues have been carried out successfully¹⁹³ and the comparison of FFPE to fresh tissues did not reveal major quantitative or qualitative differences at the protein or PTM level¹⁹⁴. We have developed techniques to reverse the cross-links induced upon fixation to ensure deep and quantitative proteome profiling to a depth of 10,000 proteins in colorectal cancer tissues^{195–197}. Our group has also shown that a combination of tissue proteomics and machine-learning classified patients with diffuse large B-cell lymphoma depending on the cell of origin¹⁹⁸. To enable a more streamlined analysis of FFPE samples, we recently demonstrated that in StageTip sample preparation is possible in a rapid manner and holds great promise for future applications in the clinic (Article 7). Regarding the sample collection and processing, MS-based tissue analysis can be performed on whole, macro- or micro-dissected tissues. Currently, the majority of large-scale tissue studies are based on whole-tissue specimens, and thus contain a mix of heterogeneous tumor cells, their respective tissue microenvironment (stroma), and most likely some non-cancer cells. The stroma, composed mainly of fibroblasts, immune and endothelial cells has been shown to play driving roles during all phases of tumorigenesis and can influence therapy and

patient outcome^{199–201}. Laser-capture microdissection (LCM) is particularly powerful to isolate homogeneous cell populations for subsequent investigation of proteomic differences between tumoral and stromal cells. Procedures in our laboratory can now easily quantify several thousands of proteins starting with as little as 10,000 micro-dissected cells in single MS runs using state of the art MS acquisition methods.

2.5.3. In the quest of more reliable and early cancer biomarkers using MS-based proteomics

A biomarker is a biological molecule that can be used to define a normal or abnormal condition or disease. It may also be used to monitor therapeutic treatment. Biomarkers can thus be grouped into prognostic, predictive, and pharmacodynamics types²⁰². Prognostic biomarkers provide information about the patient's overall cancer outcome, regardless of therapy. A classical predictive biomarker, which estimates the effect of a therapeutic intervention, is the HER2 protein amplification in breast cancer, which indicates the clinical efficacy of anti-HER2 antibodies such as Trastuzumab (Herceptin). HER2 is also a prognostic biomarker, as HER2+ breast cancers are associated with worse outcome. Similarly, EGFR mutated lung tumors have been shown higher sensitivity to erlotinib or gefitinib than wild type EGFR lung tumors²⁰³.

Biomarkers can be found at the DNA, RNA, or protein level and several analytical platforms have been developed over the years for biomarker discovery. As protein reflect the 'real time' status of a human body, are actively involved in disease onset and are the main targets of currently available cancer therapies, they represent an ideal target for biomarker discovery, in addition to the tumor driving DNA mutations. Biomarkers for early cancer detection represent one of the most promising approaches to fight cancer and improve clinical outcomes for cancer patients²⁰⁴. Protein biomarker discovery, however, remains challenging compared to DNA and RNA mainly due to the immaturity of the technology and the large dynamic range of

proteins found in body fluids and tissues. Recently, our group has developed an automated, robust, and highly reproducible workflow to quantify hundreds of plasma proteins. Among those there were more than 40 FDA-approved biomarkers⁶². In a similar manner, we are now developing a very rapid and sensitive single run mass spectrometric workflow for LCM and macrodissected cancer tissues. The latest technological developments now make DIA tissue measurements increasing competitive with DDA measurements, with the further attraction of improved measurement reproducibility.

2.5.4. Clinical cancer proteomics consortia

Similar to the TCGA project, which sequenced, characterized and catalogued cancer-specific alterations for thousands of tumors, the National Cancer Institute's (NCI) clinical proteomic tumor analysis consortium (CPTAC), launched in 2011, aims to systematically identify and characterize cancer-relevant proteins and their underlying biological pathways. CPTAC also integrates both proteomic and genomic data (termed proteogenomics), which are starting to attract interest in the cancer community. For example, the proteogenomic characterization of colon and colorectal cancer suggested novel proteomic tumor subtypes associated with clinical outcome²⁰⁵, similar to previous studies focusing on ovarian²⁰⁶, and breast⁹⁵ cancers. Of note, protein levels could not be predicted from genomic or transcriptomic data, emphasizing the importance of studying the actual molecular actors within a cellular system. Another goal of CPTAC is to enhance precision oncology and share data collected with scientists and physicians. Other cancer proteomic efforts have been based on DIA SWATH technologies. For instance, the MS company SCIEX as well as the Human Proteome Organization (HUPO) contribute to the large-scale cancer tissue analyses of multiple tumors with the goal of reproducibly identifying and quantifying at least a few thousand proteins. Such studies may generate new insights into oncogenesis beyond the genomic and transcriptional level.

II) Aims of the thesis

The aim of my thesis was to investigate system-wide alterations in cancer at the protein level. Starting from cell line models, I moved to translational tissue cancer proteomics and showed that MS-based proteomics can be used for therapeutic target identification and characterization (Figure 15).

Similar to the layout above, my journey started in the nucleus of cells, where I investigated the epigenetic changes in lymphoma at the biotechnology company Genentech Inc. We found that EZH2, which is the most significantly over-expressed epigenetic regulator in cancer, is co-regulated with parts of the network driving the cell cycle. We also uncovered that IDH1 mutation leads to perturbations of the histone code, altering histone tail acetylation and methylation in glioblastoma cell line models.

Zooming out from the nuclear to the cellular level, I analyzed phosphorylation-signaling changes in primary and secondary glioblastoma cell line models at the University of California, San Francisco (UCSF). Here, I focused on the interplay between the MAPK and PI3K signaling cascades and showed that dual inhibition is superior to single MEK inhibition and that it reverses phosphorylation-signaling patterns driven by oncogenic RAS overexpression. We also published an update of where the MS-based PTM analysis field stands, with a focus on current enrichment strategies and technological advances.

At the Max Planck Institute (MPI) of Biochemistry, I moved from cellular to translational proteomics, investigating human tissues. A major challenge was to overcome the high dynamic range of protein concentration in particular tissues, which masks the identification of low-abundant proteins. In this context, I gained much experience in the course of my exploration of the human heart proteome (Article 6). The heart is basically a sophisticated muscle, which is a particularly difficult tissue due to the overwhelming contribution of the proteins of the contractile apparatus. We combined strategies and technologies such as peptide library

matching, the peptide fractionation with the 'loss-less nano fractionator', and efficient use of the precursor ions with 'BoxCar' runs. This combined strategy yielded a dramatic increase in the total number of quantified proteins, reaching a depth of more than 10,000 proteins in the human heart.

Based on these technological advances, the final aim of the thesis was to demonstrate that MS-based proteomics can be applied to large tissue cohorts and that it is possible to gain biologically and medically relevant information. I optimized a MS-based proteomic workflow for the rapid screening of clinical tissue samples. By applying this workflow to a chemorefractory cancer patient, I uncovered a potential therapeutic target at the proteome level. Altogether, my results provide unique resources of the identity and quantity of proteins and their (dis)regulation, phosphosites and histone PTMs, revealing new insights into oncogenesis in different types of cancer.

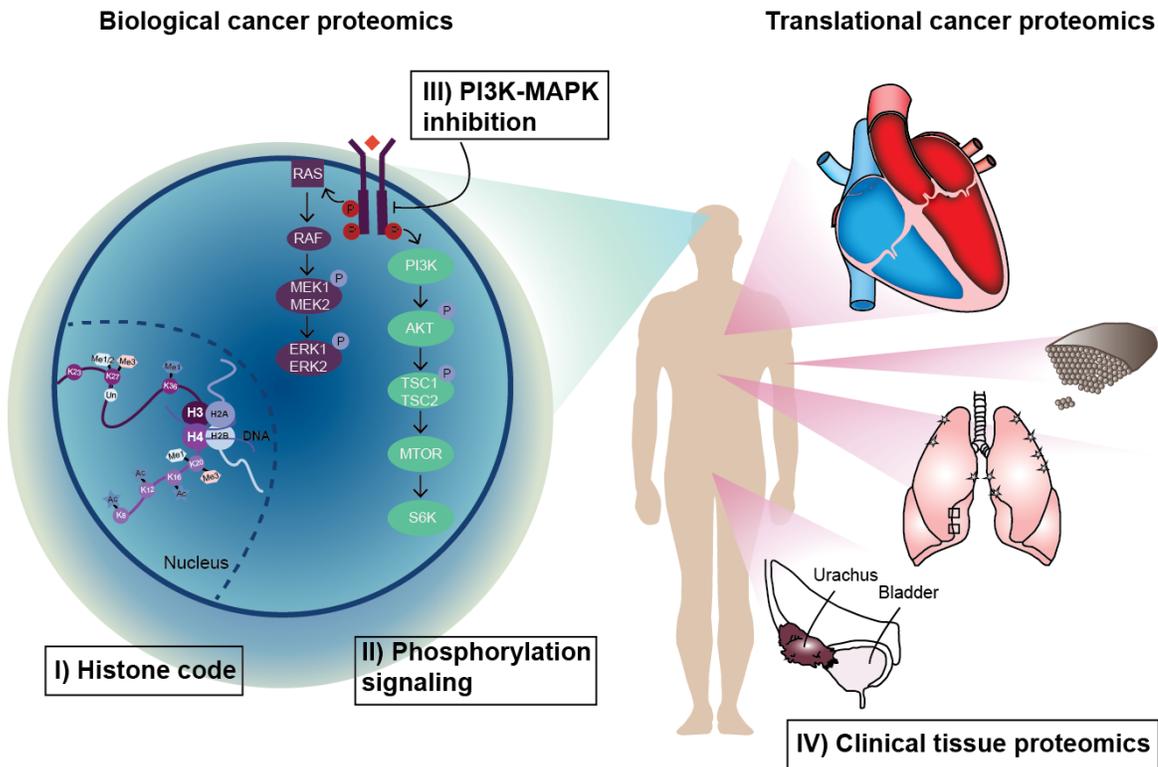


Figure 15 Overview of my PhD Thesis - A journey through the proteomic landscape of cancer. Moving from the nuclear (Genentech Inc.) and cellular (UCSF) levels to clinical tissue (MPI) cancer proteomics.

III) Publications

Article 1: Mass spectrometry-based detection and assignments of protein posttranslational modifications

Article 2: Bioinformatics analysis of thousands of TCGA tumors to determine the involvement of epigenetic regulators in human cancer

Article 3: Quantitative proteomics reveals fundamental regulatory differences in oncogenic HRAS and Isocitrate Dehydrogenase (IDH1) driven astrocytoma

Article 4: Phosphoproteome analysis of the MAPK pathway reveals previously undetected feedback mechanisms

Article 5: Quantitative phosphoproteomic analysis of the PI3K-regulated signaling network

Article 6: Region and cell-type resolved quantitative proteomic map of the human heart

Article 7: Rapid proteomic workflow for solid tumors reveals LSD1 as a drug target in an end stage cancer patient (in preparation)

Article 1: Mass spectrometry-based detection and assignment of protein posttranslational modifications²⁰⁷

ACS Chemical Biology

Sophia Doll^{1,2} and Alma L. Burlingame¹

¹Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158-2517

²Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, 82152, Germany

Great advances in MS-based proteomics now allow the identification and quantitation of thousands of posttranslational modification (PTM) sites in a single experiment. Recent developments in chromatography, PTM enrichment strategies, and mass spectrometry have vastly increased the known inventory of many protein modifications. The classes most actively investigated and discussed in this review include phosphorylation, ubiquitination, O-GlcNAcylation, methylation, and acetylation. More recently, succinylation, SUMOylation, and citrullination have been investigated globally. Characterization of key regulatory roles of PTMs in multiple cellular activities, including cancer development, have made PTMs a very attractive field of study over the last decade.

In this review, we provide an update of where the MS-based PTM analysis field stands, with a focus on current enrichment strategies and technological advances. We discuss affinity purification enrichment strategies based on charge properties, and antibody recognition. Furthermore, we review different fragmentation methods for high confidence sequence identification and site localization of different PTMs. We also highlighted a few examples of the discovery of previously unknown biological roles of PTMs. Finally, we addressed the challenge of defining site-specific functions.

My review has already been cited approximately 50 times in the last two years.



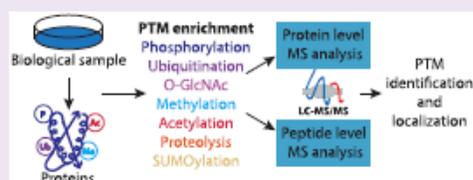
Mass Spectrometry-Based Detection and Assignment of Protein Posttranslational Modifications

Sophia Doll^{†,‡} and Alma L. Burlingame^{*†}

[†]Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158-2517, United States

[‡]Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, 82152, Germany

ABSTRACT: Recent advances in mass spectrometry (MS)-based proteomics allow the identification and quantitation of thousands of posttranslational modification (PTM) sites in a single experiment. This follows from the development of more effective class enrichment strategies, new high performance instrumentation and bioinformatic algorithms with rigorous scoring strategies. More widespread use of these combined capabilities have led to a vast expansion in our knowledge of the complexity of biological processes mediated by PTMs. The classes most actively pursued include phosphorylation, ubiquitination, O-GlcNAcylation, methylation, and acetylation. Very recently succinylation, SUMOylation, and citrullination have emerged. Among the some 260 000 PTM sites that have been identified in the human proteome thus far, only a few have been assigned to key regulatory and/or other biological roles. Here, we provide an update of MS-based PTM analyses, with a focus on current enrichment strategies coupled with revolutionary advances in high performance MS. Furthermore, we discuss examples of the discovery of recently described biological roles of PTMs and address the challenges of defining site-specific functions.



The human genome project revealed only approximately 20 000 protein-coding genes.¹ The proteome, however, is far more complex and diverse because of post-translational modifications (PTMs) and to some extent isoform variations.² While RNA sequencing detects the expression and sequence variations of the entire transcriptome,³ mass spectrometry (MS)-based proteomics has the advantage of being able to detect and structurally define any covalent changes in a protein after translation. A daunting number of such changes confer altered physiological activity, and many are reversible. There is a growing need to carry out accurate measurements of site-specific dynamics due to the lack of immunoaffinity reagents for the large numbers of newly identified proteins and their PTM analogs in rewired signaling networks, for example. Thus, the field is seeing an increase in use and further optimization of multiplexed targeted, selected-component quantitation by spectral acquisition in millisecond time frames. In fact, studies of large scale PTM dynamics will be driven by mass spectral-based quantitation—the methodology of choice. PTMs increase the functional diversity of proteins by adding covalent modifications such as phosphorylation, ubiquitination, glycosylation, methylation, and acetylation. Beside single PTMs, proteins are often modified through a combination of post-translational hydrolytic cleavages and the addition of functional groups through a stepwise processes leading to protein maturation or activation. Protein modifications influence and many times even define a large variety of normal and pathogenic cell biology functions. Therefore, identifying and understanding PTMs is critical for gaining a comprehensive understanding of cell biology, the detection and delineation of molecular defects underlying human and other diseases, drug

target discovery and validation, and the eventual treatment and prevention of diseases.

A comprehensive treatment of our earlier level of knowledge of over 300 types of PTMs, which are known to occur physiologically, can be found in the Walsh monograph.⁴ Since then, revolutionary advances in enrichment strategies and improved performances of capillary liquid chromatography (LC) and new MS instrumentation have driven our growing knowledge of many PTMs. In fact, the delineation of the actual complexity of many PTMs has emerged mostly through the past decade. Thus, by significant enrichment of classes of modified peptides before MS-analysis, thousands of precise sites can now be identified with high confidence.^{5–13}

In high-resolution tandem MS, two stages of mass analysis are used in a single experiment. The MS1 scan refers to the m/z of the precursor ion (peptide or protein), whereas the MS2 scans refer to the m/z values recorded for their fragmented ionic products. Major developments have led to new instrumentation that provides both high resolution MS and high mass measurement accuracies for both MS1 and MS2 levels simultaneously. The selection of appropriate energy deposition methods, however, is necessary to ensure generation of sequence ion series required for unambiguous site assignments. Having high mass spectral measurement quality has increased the reliability and efficiency of PTM identification at the peptide level and, in addition, has permitted the precise

Special Issue: Post-Translational Modifications

Received: November 5, 2014

Accepted: December 26, 2014

Published: December 26, 2014

localization of modified sites for some intact protein sequences.¹⁴ In particular, MS fragmentation strategies that generate sufficient peptide fragmentation information are essential for precise PTM identification and localization, by definition. Among the different fragmentation strategies commonly employed, electron capture and transfer-based fragmentation techniques have proven to be essential for the localization of labile modifications as well as in dramatically expanding our experimental capability to carry out sequence analyses on large peptides or even medium size intact proteins. Labile structures include many serine or threonine phosphorylation sites as well as very labile modifications such as *O*-linked *N*-acetylglucosamine (*O*-GlcNAc), γ -carboxyglutamic acid, and others.^{15–18}

In this review, we first describe how enrichment strategies and the revolutionary advances in mass spectrometry have contributed to high confidence PTM identification as well as site-specific localization. We then emphasize the discovery of previously unknown biological roles of PTMs including the signaling pathway activation role of arginine methylation and multiple cross-talk forms between PTMs. Finally, we address the biological challenges of defining individual site-specific functions. The complex fields of extracellular glycosylation and protein lipidation are not covered in this review, but the interested reader is referred to recently published reviews on these topics.^{19,20}

PTM Enrichment Strategies. Since regulatory PTMs feature substoichiometric site occupancies in many cases, specific enrichment techniques are essential to achieve detection and characterization of low relative abundance components in digests of cell lysates (Table 1). Common

Table 1. PTM Enrichment Strategies

enrichment strategy	PTM
antibody-based	tyrosine phosphorylation ^{22,33}
	arginine/lysine methylation ³⁶
	lysine acetylation ³⁰
	ubiquitin-like ^{17,35}
ionic interaction-based IMAC, TiO ₂ , SIMAC	serine/threonine/arginine phosphorylation ^{24,25}
metabolic tagging	lectin
	<i>N</i> -glycosylation/ <i>O</i> -GlcNAc ¹¹
	His10 SUMO2
	SUMO ¹³
iodoTMT	S-nitrosylation ⁴³
	citrullination ⁴¹
glyoxal derivat	
enzymatic-based	
	subtiligase
	proteolysis ³⁸
PNGase	
<i>N</i> -glycosylation ¹¹	
phospholipase	
GPI-anchor ⁴²	

PTM enrichment strategies use affinity purification based on charge properties or antibody recognition. They are applied at the peptide level usually ensuring higher accessibility of PTMs and allowing quite specific binding. Prior to enrichment, proteins are digested into peptides. The most commonly used enzyme for peptide level PTM analysis is trypsin, due to its high cleavage specificity after lysine and arginine.²¹ In some cases, however, the resulting tryptic peptides might be too long or too short for effective MS analysis. Instead of using trypsin alone, other proteases such as chymotrypsin, AspN, Lys-N, or

endoproteinase Glu-C may be employed to increase the PTM coverage.²²

Ionic Interaction-Based PTM Enrichment Strategies. Among all PTMs phosphorylation is the most extensively studied and approximately 197 000 human phosphorylation sites have been reported.²³ Global analysis of serine- and threonine-phosphorylation is commonly achieved by metal ion-based enrichment methods such as immobilized metal affinity chromatography (IMAC) and titanium dioxide (TiO₂).^{24,25} Both approaches use metal cations to bind the negatively charged phosphopeptides, and protocols designed to achieve higher enrichment efficiencies have been constantly improved over the past decade. Space does not permit discussion of the extensive literature here, but the interested reader is referred to the recent Engholm-Keller and Larsen review.²⁷

Multiply phosphorylated peptides, however, remain challenging to detect and assign unambiguously. The combination of both IMAC and TiO₂ enrichment methods, which is termed sequential elution from IMAC (SIMAC) enables efficient separations of monophosphorylated from multiply phosphorylated peptides and thus higher numbers of monophosphorylation site identifications.²⁸ The use of graphite powder and titanium(IV) has been reported to increase the identification of multiply phosphorylated peptides.^{29,30} An alternative approach that holds promise to achieve higher phosphopeptide enrichment levels involve gallium complexes that stabilize the weak interaction between phosphoryl moiety and serine or threonine residues.³¹

Antibody-Based PTM Enrichment Methods. Global phosphorylation analysis revealed that 86%, 12%, and 2% of phosphorylation events occur on serine, threonine, and tyrosine, respectively.³² Although phosphotyrosine residues represent only a small percentage of all phosphosites the signal-initiating role of receptor tyrosine kinase (RTK) phosphorylation initiates key signaling cascades and plays a driving role in multiple diseases including cancer. To achieve higher identification coverage of phosphotyrosine-containing peptides, antibody-based enrichments in combination with LC-MS/MS analyses have been applied yielding quantitative profiling of hundreds of phosphorylated tyrosine residues.^{33,34}

Antibody-based enrichment applications have been extended to quantitative profiling of ubiquitination, methylation, and acetylation. In particular, the development of antiglycine-K antibodies led to the detection of more than 19 000 ubiquitination sites in a single proteomic workflow.^{12,35} Furthermore, antibodies targeting mono-, di-, and trimethylated lysine moieties and mono- and dimethyl arginine side chains have been applied to map the human methylome in depth.³⁶

Enzymatic-Based PTM Enrichment. In addition to ionic interaction- or antibody-based enrichment strategies, enzymatic-based enrichments such as that, based upon application of the protein ligase, subtiligase, have been developed to study cellular substrates formed during intracellular proteolysis. During proteolytic cleavage, new free N-termini are generated, which are specifically biotinylated by the subtiligase, while native acetylated N-termini (present on almost 90% of human proteins) are not recognized. Subtiligase does not recognize ϵ -amino functions on lysine side chains. Upon labeling, avidin beads capture the biotinylated N-terminal amino functions of newly cleaved proteins. These labeled truncated proteins are digested with trypsin. The N-terminal peptide of the substrate is subsequently released from the avidin beads at the built in cleavage site by Tobacco etch virus (TEV) and analyzed by LC-

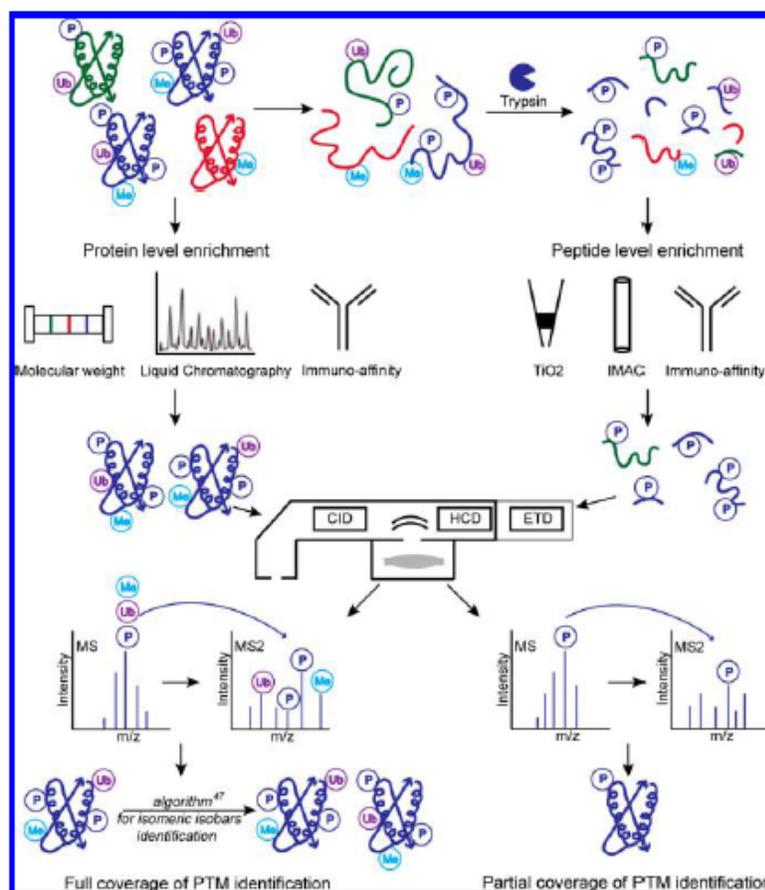


Figure 1. Overview of PTM analysis at the protein and peptide level. The principal steps for protein level PTM enrichment and MS analysis are represented on the left. Peptide level phosphorylation enrichment strategies are illustrated on the right. Isomeric isobars are proteins or peptides that may reveal the same amino acid sequence with equal numbers of PTMs but with different PTM configurations. The development of algorithms that search for non-redundant ions representing all possible PTM configurations enable the identification, quantification, and localization of all PTMs.

MS/MS.³⁷ This enrichment method allowed the identification of over 8000 proteolytic sites including more than 1700 caspase cleavages sites in human cells.³⁸

Very recently, the wild type α -lytic protease was shown to cleave small ubiquitin-like modifier (SUMO) modified proteins to form a GG-K remnant, thus providing a new strategy for mapping of SUMO-modified proteins without the need for introduction of mutations.³⁹ In addition, it has been reported that over 4000 SUMO sites have been uncovered using the decahistidine (His₁₀)-tagged SUMO-2 strategy prior to MS measurement.¹³

Sequential Enrichment. The development of sequential enrichment strategies, which consist of using the flow-through of a first PTM enrichment step for a second or third enrichment step has become particularly desirable to reveal potential regulatory relationships from cross-talk among multiple PTMs from the same biological sample.^{8,40}

In summary, enrichment strategies have been successfully established for proteome-wide identification of phosphorylation, ubiquitination, acetylation, methylation, proteolytic cleavages, SUMOylation, as well as lectin-based enrichment for O- and N-linked glycopeptides¹¹ and iodoTMT-based S-nitrosylation enrichment.⁴³ It is noted that neither analytical

nor enrichment strategies exist yet for the vast majority of the other known PTMs.⁴

PTM Enrichment at the Protein Level. In contrast to peptide level approaches, direct PTM analysis of intact proteins preserves the intact structure of the protein that is mostly destroyed in peptide level approaches. Prior to MS analysis, intact proteins are separated from complex protein mixtures using a variety of enrichment techniques.^{44,45} These include gel-eluted liquid fraction entrapment electrophoresis, which separates proteins on the basis of their molecular weights, and LC techniques such as affinity, ion-exchange chromatography, size-exclusion chromatography, reverse phase chromatography, and online reversed phase LC tandem MS approach.^{1,4,46,48} Immunoaffinity methods can also be used as effective and specific protein purification protocols.⁴⁹ Intact MS is a powerful technique to reveal global purity and relative stoichiometries and localize PTMs in highly modified but small proteins (10 to 50 kDa) without prior knowledge of targeted PTMs (Figure 1).^{44–46}

MS Fragmentation Strategies. In MS-based PTM analysis, it is essential to generate enough peptide fragmentation information (MS2 scans) for high confidence sequence identification and site localization of PTMs. Several fragmenta-

tion strategies including collision induced dissociation (CID), higher energy collisional dissociation (HCD), electron capture dissociation (ECD), and electron transfer dissociation (ETD) have been employed over the past decade for this purpose.

Collision Induced Dissociation. CID is the most common and widely applied unimolecular dissociation technique for MS-based proteome identification and quantification analysis. Under CID, the peptide or protein precursor ions are positively charged due to protonation of basic amino-acid residues and undergo collisions by interactions with neutral gas molecules (vibronic activation). In accordance with the mobile proton model,⁵⁰ peptide bonds dissociate generating primary N-terminal b- and C-terminal y-type ions.⁵¹ The CID process is generally more effective for small and low-charge state peptides but is strongly influenced by the amino acid sequence and the distribution of the positive charges across the peptide backbone.

Higher Energy Collisional Dissociation. A similar strategy of fragmentation is the beam-type CID, also called HCD. HCD fragmentation is characterized by higher activation energy compared to CID. The higher fragmentation efficiency for HCD produces predominantly y-type fragment ions compared to b-type ions.^{52,53} In the HCD mode, Fourier transform detection in the Orbitrap analyzer results in better quality of MS2 mass spectra, but spectral acquisition times are longer compared to the greater acquisition speed recording CID spectra (MS/MS data collected in the ion trap).^{54,55} Both collisionally activated dissociation (CAD) methods, CID and HCD, are almost universal for analyses of stable PTMs and provide a high probability to generate and detect the modification-specific peptide sequence ion series.

Electron Capture and Electron Transfer Dissociations. In certain cases, however, CAD methods do not provide site-specific modification ions for long amino acid sequences. Moreover, the localization of the phosphorylation event within identified peptides that harbor more than one serine, threonine, or tyrosine is sometimes ambiguous. In addition, it has proven to be very challenging to decipher the complex histone patterns of modifications or assign PTMs that are labile in the gas phase using CAD energy deposition methods. For these cases, electron-based fragmentation methods such as ECD and ETD can be applied, achieving fragmentation through neutralization of backbone protonation sites with thermal electrons (ECD) or radical anions (ETD).^{14,17,56} The resulting nonergodic cleavages of N- α backbone bonds generate c- and z- type fragment ions without losing the PTM localization information.^{14,57} While ECD can only be implemented on Fourier transform ion cyclotron resonance (FTICR) MS instruments, ETD can be implemented on high resolution tandem MS instruments and is able to achieve higher detection sensitivity of labile PTM sites as well as complex PTM occupancies than with ECD-based approaches.^{17,58–60} ECD and ETD are complementary to CAD because they perform better with highly charged state analytes, whereas CAD is more efficient with low-charge state peptides.^{57,61} ECD and ETD, however, have major advantages over CAD for detecting unstable PTMs because peptide backbone fragmentation is virtually independent of the amino acid sequence, neutral losses such as phosphate groups are reduced, and O-GlcNAc elimination does not occur.^{8,17,61,62}

HCF-1 is a transcriptional coregulator of cell proliferation and has been previously described as one of the most highly O-GlcNAcylated proteins. Nearly 30 HCF-1 O-GlcNAc sites have

previously been reported.¹⁷ Using ETD and HCD, the site-specific localization of nearly 20 additional O-GlcNAc on HCF-1 could be identified with high confidence, allowing better understanding of the transcriptional regulating role of O-GlcNAc modification of HCF-1.¹⁷ In contrast, α B-crystalline has been found to be O-GlcNAcylated at only one serine residue.⁶³ O-GlcNAc transferase (OGT) was found to be phosphorylated by Glycogen Synthase Kinase 3 β (GSK3 β), and the phosphorylation site has been localized to serine 3 or 4.⁶⁴

Intact Protein Isoforms Profiling. Protein isoforms share a high percentage of amino acid sequence homology but often dramatically differ in their cellular concentration and biological roles.⁶⁵ While MS on a digest mixture may not detect the peptide carrying a given particular isoform modification, intact protein level analysis has the advantage of mapping the complete amino acid sequence. In intact MS/MS analysis, precursor ions are conventionally selected by quadrupole or ion trap device before fragmentation. These methods, however, suffer from low mass selectivity. The stored waveform inverse Fourier transform (SWIFT) method enables high mass selectivity and achieves better isoform detection but can only be implemented on an FTICR instrument.⁶⁶ Although FTICR mass spectrometers are expensive and less easily accessible than other mass spectrometers, traditionally, intact MS analysis has been most successful with FTICR-MS, since it provides the highest possible resolution for intact protein sample analysis.^{67,68} Early work combining ECD and the SWIFT technology on an FTICR-MS instrument has revealed the PTM occupancies of intact histone variants H2B.1 and H2B.2 from tetrahymena as well as novel H3 protein isoforms in rat.^{69,70}

Dynamics-Quantitation. New MS designs have proven to provide powerful tools to quantify selected components of protein and PTM networks that provide new insight into cellular dynamics. The dynamics of histone modifications have been recently quantitatively established. Metabolic labeling of human cells using ¹³C glucose has been shown to enable monitoring of the dynamics of ¹³C-labeled acetyl groups' incorporation on specific histone lysine. In this work, the turnover of acetylation was determined to be generally faster than methylation but slower than phosphorylation. Moreover, the modification rate varied depending on the histone type, targeted residues, as well as neighboring modifications.⁷¹

Caspase cleavage dynamics have also been quantitatively assessed by MS. Applying the N-termini subtiligase enrichment strategy, MS experiments across three human cell lines have revealed that the cellular cleavage kinetics of over 500 caspase substrates vary strongly between cell types and cytotoxic drug treatments. Furthermore, common caspase cleavage substrates that can be used to monitor the pro-apoptotic effects of cancer drug treatment have been identified.⁷²

Uncovering PTM-Mediated Biological Processes.
Arginine Methylation Initiates Smad Signaling. Phosphorylation is arguably the most common modification and a central mechanism for cell transduction. Indeed many transmembrane receptors are kinases or act through cytoplasmic kinase domains. In some cases, however, the low kinetics of substrate phosphorylation suggests a preceding step in the signal initiation. Recently, arginine methylation has been shown to activate the bone morphogenetic proteins (BMP) induced Smad-signaling pathway. Upon BMP binding inhibitory Smad6 is methylated by PRMT1. Smad6 in its unmethylated form

dissociates from the BMP kinase receptor complex allowing activation of regulatory Smads through phosphorylation (Figure 2A).⁷³ This observation raises the possibility that ligand-induced methylation may play a role in the activation of other signaling pathways.

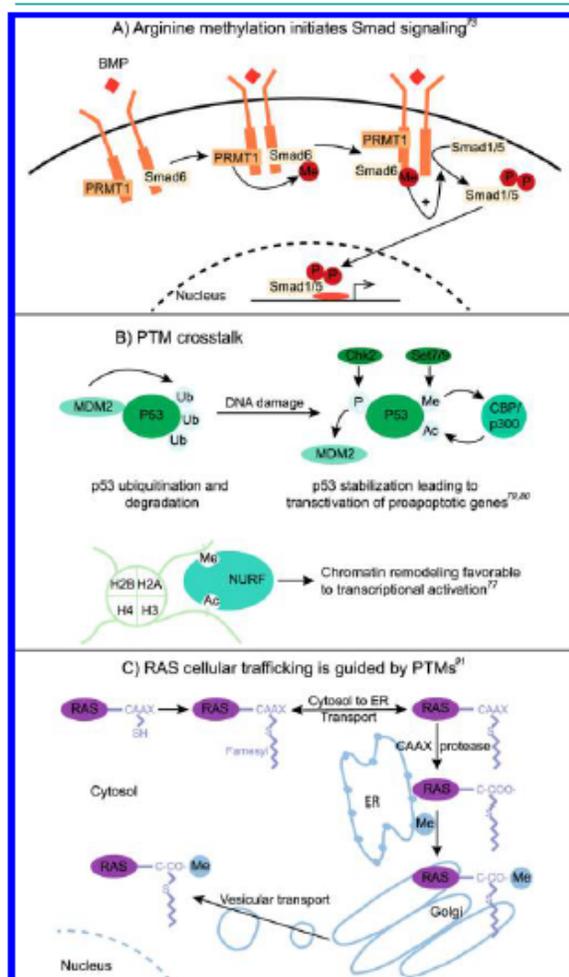


Figure 2. Examples of PTM-mediated biological processes. (A) Arginine methylation-mediated Smad signaling pathway. (B) p53 interaction with the CBP/p300 complex through methylation and acetylation exchange after DNA damage. p53 is subsequently stabilized and transactivates pro-apoptotic genes. The lower portion of panel B represents a protein-mediated histone crosstalk between methylation and acetylation, leading to chromatin remodeling favorable to transcriptional activation. (C) PTM-mediated cellular trafficking of RAS.

PTM Crosstalk. Since recent studies have provided a global view of the widespread occurrence of many PTMs, the next challenge is to understand the interplay among different PTM classes and specific sites of regulation. PTM-mediated crosstalk has been classified as positive or negative.⁷⁴ In the case of positive crosstalk, one PTM serves as an active signal for the addition or removal of another PTM (e.g., phosphorylation dependent ubiquitination⁷⁵ and SUMOylation⁷⁶) or as a trigger for binding proteins that carry out a second modification (e.g.,

histone mark binding proteins⁷⁷). In contrast, negative crosstalk can include direct competition of different modifications targeting the same amino acid or one modification can mask the recognition site for another PTM (e.g., the acetylation-mediated inhibition of kinase phosphorylation⁷⁸).

Crosstalk between phosphorylation and ubiquitination is one of the most studied relationships. For example, phosphorylation can either promote or inhibit ubiquitination by regulating the activity of the E3 ubiquitin ligase controlling proteasomal degradation. For example, following DNA damage, p53 becomes phosphorylated, thus decreasing the interaction affinity with E3 ligase MDM2. p53 in its deubiquitinated form further interacts with the CBP/p300 complex through methylation and acetylation exchange. p53 is consequently stabilized and transactivates pro-apoptotic genes (Figure 2B).^{79,80} Crosstalk between phosphorylation and ubiquitination is also essential for the cancer-implicated epidermal growth factor (EGF)-mediated extracellular-signal-regulated kinases signaling pathway.⁷⁵ Here, E3 ligase Cbl binds phosphotyrosine residues of the activated EGF receptor (EGFR). Subsequent proteasome-independent ubiquitination of EGFR and endocytic adaptor proteins mediate the EGFR internalization. Interestingly, deubiquitination directs EGFR through the recycling pathway back to the cell membrane.⁸¹

The histone code presents one of the most important examples for extensive interplay between different PTMs. ETD-based MS revealed several H3 lysine residues including lysine 9 (H3K9) that can be exclusively methylated or acetylated with different biological outcomes. While H3K9 methylation correlates with transcriptional repression, H3K9 acetylation induces transcriptional activation. H3K9 can be mono-, di-, or trimethylated with varying biological output depending on the degree of methylation.⁸² Other histone 3 crosstalk forms include the recently described acetylation-dependent SUMOylation.¹³ This new type of crosstalk is of particular interest, since histone SUMOylation has been previously associated with transcriptional repression, whereas histone acetylation is linked to transcriptional activation. Another form of crosstalk describes proteins that bind to different types of PTMs. For example, the nucleosome remodeling factor (NURF) contains two domains that bind the H3K4 di- or trimethylation and H4K16 acetylation sites, leading to the transcriptional activation of homeotic genes (Figure 2B).⁷⁷ An additional form of histone crosstalk involves the phosphorylation of H3 (H3S10), which leads to the acetylation of H4. H3S10 creates a binding site for 14-3-3, a phosphoserine binding protein. 14-3-3 recruits a histone acetyltransferase MOF, which subsequently acetylates H4 on lysine 16. Acetylated H4K16 in turn forms a binding site for a kinase that phosphorylates RNA Polymerase II to facilitate transcriptional elongation of FOSL1, a gene activated in response to serum.⁸³ Histone modifications can also prevent the recruitment of binding proteins. For example, the heterochromatin protein 1 (HP1) is not able to bind H3K9 when the adjacent serine 10 is phosphorylated during mitosis or gene activation.⁸⁴ In general, the interplay among PTMs on histones appears to be context and time specific, increasing the challenge of understanding the regulated changes.

A few examples have shown that O-GlcNAcylation interacts with phosphorylation, acetylation, methylation, and ubiquitination.^{64,85–88} While the crosstalk between O-GlcNAcylation and phosphorylation has been hypothesized to be particularly important for multiple cellular processes, including the

regulation of enzyme activity, cell division, and cytoskeletal functions,^{17,85,89} large-scale experiments revealed that only 7% of all O-GlcNAcylation sites were found to be phosphorylated also and the frequency of negative crosstalk between these two PTMs is virtually equal to the frequency expected by chance alone.⁸ Over the past decade, however, many particular examples of crosstalk have been described, showing that PTMs can work in concert to determine the final biological read-out.^{64,74,90}

PTMs Regulate the Cellular Ras Trafficking. While the activation status of Ras proteins is regulated by the exchange of GTP, for GDP their cellular localization is mediated by PTMs. Ras traffics between various subcellular compartments guided by modifications such as farnesylation, proteolysis, methylation, palmitoylation, and phosphorylation. While farnesylation of Ras proteins increases its affinity toward the endoplasmic reticulum (ER), proteolysis and subsequent methylation and palmitoylation trap the two Ras isoforms HRas and NRas in the Golgi apparatus. Ras proteins are subsequently transported to the plasma membrane via vesicles (Figure 2C).⁹¹

These PTMs present potential therapeutic targets for the development of small molecule Ras inhibitors in cancer.

Biological Challenge of Defining Site-Specific Functions. The revolutionary development of PTM enrichment methods and new MS strategies has enabled the identification and quantification of more than 260 000 PTMs.^{2,3} Only a small proportion of PTM sites, however, have been associated with a particular biological function. One major challenge for the discovery of site-specific functions is to select a small number of sites from a large-scale data set for follow-up experiments. In general, any given PTMs are selected depending on the identification accuracy, reasonable stoichiometry, and their potential regulatory role in the process of interest. Evolutionary conservation usually points to important functionality and can therefore be used as a guide in selection criteria.⁹² Quantitative large scale MS is an additional powerful tool that can be employed to initially screen a subset of modified proteins that are regulated by the pathway of interest. For example, large-scale MS revealed that less than 15% of phosphorylated sites are modulated by EGF treatment.³²

Functional follow-up experiments often include *in vitro* enzyme assays to determine specific enzyme substrates or lysine point mutations to investigate the functional role of a particular PTM. The core components of the Clustered Regulatory Interspaced Palindromic Repeats (CRISPR) system includes the Cas9 nuclease, which is able to create double-strand breaks in DNA and guide RNA (gRNA), which directs the CRISPR complex to a target sequence complementary to gRNA.⁹³ Using CRISPR, point mutations have been generated in the genome of mice, which led to single amino acid substitutions in proteins of interest to probe site-specific PTM functions under *in vivo* conditions.⁹⁴ Furthermore, the combination of CRISPR and Chromatin Affinity Purification (ChAP)-MS provides a new tool to study epigenetic regulation. Applying CRISPR-ChAP-MS, a specific section of chromatin can be purified for subsequent identification of associated histone PTMs and proteins by high resolution MS. Unlike Chromatin Immunoprecipitation (ChIP), this proteomic approach does not depend on prior knowledge of the targeted protein or PTM.⁹⁵ Other useful MS follow up experiments rely on short hairpin RNA (shRNA) knockdown of specific regulators to investigate protein network interactions.⁹⁶

MS-based proteomic measurements with additional conditions, such as targeted kinase or receptor inhibition, perturbing the PTMs of interest can also be used as functional follow-up experiments. For instance, large-scale proteomics experiments have elucidated the mTOR-regulated phosphoproteome⁹⁷ and cellular outputs of the fibroblast growth factor receptor.⁹⁸ In addition, MS analyses revealed new SUMOylated protein sites uncovering the regulatory role of SUMOylation in all nuclear processes.¹³ Despite these technological advances, follow up experiments are often limited to selected PTMs and/or sites on particular proteins and therefore would not reveal potential functionality for the majority of PTMs that can be characterized. Hence, it is reasonable to call into question the biological relevance of at least some fraction of the 260 000 reported human PTM sites.⁹⁹

In summary, enrichment strategies and revolutionary advances in mass spectrometry have enabled rigorous identification and quantification of large numbers of PTMs, which certainly constitute the most complex regulatory networks in eukaryote cells. Future challenges for PTM proteomics include the optimization of PTM enrichment strategies and the development of PTM screens that are faster, more sensitive, and reproducible. Finally, the emerging importance of PTM multisite occupancies and potential modes of PTM crosstalk in proteins require the development of novel methods for full-spectrum PTM identification not only at the peptide level but also at the intact protein level.

AUTHOR INFORMATION

Corresponding Author

*E-mail: alb@cgl.ucsf.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support was provided by National Institutes of Health/National Institute of General Medical Sciences (NIH/NIGMS) grant 8P41GM103481 and the Adelson Medical Research Foundation. We thank R. J. Chalkley and F. Gnäd for helpful input.

REFERENCES

- (1) Consortium, I. H. G. S. (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.
- (2) Jensen, O. N. (2004) Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* 8, 33–41.
- (3) Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63.
- (4) Walsh, C. T. (2006) *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, 1st ed., pp 1–490, Roberts and Company, Publishers, Greenwood Village, CO.
- (5) Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphal, M. S., and Coon, J. J. (2014) The one hour yeast proteome. *Mol. Cell. Proteomics* 13, 339–347.
- (6) Imamura, H., Sugiyama, N., Wakabayashi, M., and Ishihama, Y. (2014) Large-scale identification of phosphorylation sites for profiling protein kinase selectivity. *J. Proteome Res.* 13, 3410–3419.
- (7) Masuda, T., Sugiyama, N., Tomita, M., and Ishihama, Y. (2011) Microscale phosphoproteome analysis of 10 000 cells from human cancer cell lines. *Anal. Chem.* 83, 7698–7703.
- (8) Trinidad, J. C., Barkan, D. T., Gullledge, B. F., Thalhammer, A., Sali, A., Schoepfer, R., and Burlingame, A. L. (2012) Global identification and characterization of both O-GlcNAcylation and

- phosphorylation at the murine synapse. *Mol. Cell. Proteomics* 11, 215–229.
- (9) Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S., and Mann, M. (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* 3, ra3 DOI: 10.1126/scisignal.2000475.
- (10) Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834–840.
- (11) Zielinska, D. F., Gnad, F., Schropp, K., Wisniewski, J. R., and Mann, M. (2012) Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol. Cell* 46, 542–548.
- (12) Kim, W., Bennett, E. J., Huttlin, E. L., Guo, A., Li, J., Possemato, A., Sowa, M. E., Rad, R., Rush, J., Comb, M. J., Harper, J. W., and Gygi, S. P. (2011) Systematic and quantitative assessment of the ubiquitin modified proteome. *Mol. Cell* 44, 325–340.
- (13) Hendriks, I. A., D'Souza, R. C. J., Yang, B., Verlaan-de Vries, M., Mann, M., and Vertegaal, A. C. O. (2014) Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat. Struct. Mol. Biol.* 21, 927–936.
- (14) Elnik, S. M., Maltby, D., Panning, B., and Burlingame, A. L. (2010) High resolution electron transfer dissociation studies of unfractionated intact histones from murine embryonic stem cells using on-line capillary LC separation. *Mol. Cell. Proteomics* 9, 824–837.
- (15) Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9528–9533.
- (16) Jünger, M. A., and Aebersold, R. (2014) Mass spectrometry-driven phosphoproteomics: Patterning the systems biology mosaic. *Wiley Interdiscip. Rev. Dev. Biol.* 3, 83–112.
- (17) Myers, S. A., Dao, S., Affar, E. B., and Burlingame, A. (2013) Electron transfer dissociation (ETD): The mass spectrometric breakthrough essential for O-GlcNAc protein site assignments—a study of the O-GlcNAcylated protein host cell factor C1. *Proteomics* 13, 982–991.
- (18) Ramström, M., and Sandberg, H. (2011) Characterization of γ -carboxylated tryptic peptides by collision-induced dissociation and electron transfer dissociation mass spectrometry. *Eur. J. Mass Spectrom.* 17, 497–506.
- (19) Moremen, K. W., Tiemeyer, M., and Naim, A. V. (2012) Vertebrate protein glycosylation: Diversity, synthesis, and function. *Nat. Rev. Mol. Cell Biol.* 13, 448–462.
- (20) Han, X., Yang, K., and Gross, R. W. (2012) Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. *Mass Spectrom. Rev.* 31, 134–178.
- (21) Olsen, J. V., Ong, S.-E., and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* 3, 608–614.
- (22) Gilmore, J. M., Kettenbach, A. N., and Gerber, S. A. (2012) Increasing phosphoproteomic coverage through sequential digestion by complementary proteases. *Anal. Bioanal. Chem.* 402, 711–720.
- (23) www.phosphosite.org.
- (24) Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faergeman, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* 4, 310–327.
- (25) Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. R. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* 76, 3935–3943.
- (26) Schmidt, A., Trentini, D. B., Spiess, S., Fuhrmann, J., Ammerer, G., Mechtler, K., and Clausen, T. (2014) Quantitative phosphoproteomics reveals the role of protein arginine phosphorylation in the bacterial stress response. *Mol. Cell. Proteomics* 13, 537–550.
- (27) Engholm-Keller, K., and Larsen, M. R. (2013) Technologies and challenges in large-scale phosphoproteomics. *Proteomics* 13, 910–931.
- (28) Thingholm, T. E., Jensen, O. N., Robinson, P. J., and Larsen, M. R. (2008) SIMAC (Sequential Elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol. Cell. Proteomics* 7, 661–671.
- (29) Larsen, M. R., Graham, M. E., Robinson, P. J., and Roepstorff, P. (2004) Improved detection of hydrophilic phosphopeptides using graphite powder microcolumns and mass spectrometry: Evidence for *in vivo* doubly phosphorylated dynamin I and dynamin III. *Mol. Cell. Proteomics* 3, 456–465.
- (30) Zhou, H., Low, T. Y., Hennrich, M. L., van der Toorn, H., Schwend, T., Zou, H., Mohammed, S., and Heck, A. J. R. (2011) Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV)-based IMAC enrichment. *Mol. Cell. Proteomics* 10, M110–006452.
- (31) Svane, S., Kryuchkov, F., Lennartson, A., McKenzie, C. J., and Kjeldsen, F. (2012) Overcoming the instability of gaseous peptide phosphate ester groups by dimetal protection. *Angew. Chem. Int. Ed. Engl.* 51, 3216–3219.
- (32) Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635–648.
- (33) Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X.-M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.* 23, 94–101.
- (34) Pighi, C., Gu, T.-L., Dalai, I., Barbi, S., Parolini, C., Bertolaso, A., Pedron, S., Parisi, A., Ren, J., Cecconi, D., Chilosi, M., Menestrina, F., and Zamo, A. (2011) Phospho-proteomic analysis of mantle cell lymphoma cells suggests a pro-survival role of B-cell receptor signaling. *Cell. Oncol. (Dordr.)* 34, 141–153.
- (35) Udesi, N. D., Svinkina, T., Mertins, P., Kuhn, E., Mani, D. R., Qiao, J. W., and Carr, S. A. (2013) Refined preparation and use of anti-glycine remnant (K-E-GG) antibody enables routine quantification of 10 000s of ubiquitination sites in single proteomics experiments. *Mol. Cell. Proteomics* 12, 825–831.
- (36) Carlson, S. M., and Gozani, O. (2014) Emerging technologies to map the protein methylome. *J. Mol. Biol.* 426, 3350–3362.
- (37) Yoshihara, H. A. I., Mahrus, S., and Wells, J. A. (2008) Tags for labeling protein N-termini with subtiligase for proteomics. *Bioorg. Med. Chem. Lett.* 18, 6000–6003.
- (38) Wüta, A. P., Seaman, J. E., and Wells, J. A. (2014) Global analysis of cellular proteolysis by selective enzymatic labeling of protein N-termini. *Methods Enzymol.* 544, 327–358.
- (39) Meyer, J. G., Yang, B., Bennett, E., and Komives, E. A. (2014) A novel comprehensive discovery approach for SUMO modified proteins. *Mol. Cell. Proteomics* 13, S41–S43.
- (40) Mertins, P., Qiao, J. W., Patel, J., Udesi, N. D., Clauser, K. R., Mani, D. R., Burgess, M. W., Gillette, M. A., Jaffe, J. D., and Carr, S. A. (2013) Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat. Methods* 10, 634–637.
- (41) Christophorou, M. A., Castelo-Branco, G., Halley-Stott, R. P., Oliveria, C. S., Loos, R., Radzsheuskaya, A., Mowen, B. A., Bertone, P., Silva, J. C. R., Zernicka-Goetz, M., Nielsen, M. L., Gurdon, J. B., and Kouzarides, T. (2014) Citullination regulates pluripotency and histone H1 binding to chromatin. *Nature* 507, 104–108.
- (42) Deeg, M. A., and Davitz, M. A. (1995) Glycosylphosphatidylinositol-phospholipase D: A tool for glycosylphosphatidylinositol structural analysis. *Methods Enzymol.* 250, 630–640.
- (43) Qu, Z., Meng, F., Bomgardner, R. D., Viner, R. L., Li, J., Rogers, J. C., Cheng, J., Greenleaf, C. M., Cui, J., Lubahn, D. B., Sun, G. Y., and Gu, Z. (2014) Proteomic quantification and site-mapping of S-nitrosylated proteins using isobaric iodoTMT reagents. *J. Proteome Res.* 13, 3200–3211.
- (44) Ge, Y., Rybakova, I. N., Xu, Q., and Moss, R. L. (2009) Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12658–12663.

- (45) Boyne, M. T., Pesavento, J. J., Mizzen, C. A., and Kelleher, N. L. (2006) Precise characterization of human histones in the H2A gene family by top down mass spectrometry. *J. Proteome Res.* 5, 248–253.
- (46) Zhang, H., and Ge, Y. (2011) Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circ. Cardiovasc. Genet.* 4, 711–711.
- (47) Guan, S., and Burlingame, A. L. (2010) Data processing algorithms for analysis of high resolution MS/MS spectra of peptides with complex patterns of posttranslational modifications. *Mol. Cell. Proteomics* 9, 804–810.
- (48) Fisk, J. C., Li, J., Wang, H., Aletta, J. M., Qu, J., and Read, L. K. (2013) Proteomic analysis reveals diverse classes of arginine methylproteins in mitochondria of trypanosomes. *Mol. Cell. Proteomics* 12, 302–311.
- (49) Zhang, J., Dong, X., Hacker, T. A., and Ge, Y. (2010) Deciphering modifications in swine cardiac troponin I by top-down high-resolution tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* 21, 940–948.
- (50) Paizs, B., and Suhai, S. (2005) Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* 24, 508–548.
- (51) Lau, K. W., Hart, S. R., Lynch, J. A., Wong, S. C. C., Hubbard, S. J., and Gaskell, S. J. (2009) Observations on the detection of b- and y-type ions in the collisionally activated decomposition spectra of protonated peptides. *Rapid Commun. Mass Spectrom.* 23, 1508–1514.
- (52) Frese, C. K., Altelea, A. F. M., Henrich, M. L., Nolting, D., Zeller, M., Griep-Raming, J., Heck, A. J. R., and Mohammed, S. (2011) Improved peptide identification by targeted fragmentation using CID, HCD, and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* 10, 2377–2388.
- (53) Cui, W., Thompson, M. S., and Reilly, J. P. (2005) Pathways of peptide ion fragmentation induced by vacuum ultraviolet light. *J. Am. Soc. Mass Spectrom.* 16, 1384–1398.
- (54) Nagaraj, N., D'Souza, R. C. J., Cox, J., Olsen, J. V., and Mann, M. (2010) Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J. Proteome Res.* 9, 6786–6794.
- (55) Jedrychowski, M. P., Huttlin, E. L., Haas, W., Sowa, M. E., Rad, R., and Gygi, S. P. (2011) Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol. Cell. Proteomics* 10 (12), M1111–009910.
- (56) Young, N. L., DiMaggio, P. A., Plazas-Mayorca, M. D., Baliban, R. C., Floudas, C. A., and Garcia, B. A. (2009) High throughput characterization of combinatorial histone codes. *Mol. Cell. Proteomics* 8, 2266–2284.
- (57) Huang, T.-Y., and McLuckey, S. A. (2010) Gas-phase chemistry of multiply charged bioions in analytical mass spectrometry. *Annu. Rev. Anal. Chem.* 3, 365–385.
- (58) McAlister, G. C., Berggren, W. T., Griep-Raming, J., Hoving, S., Makarov, A., Phanstiel, D., Stafford, G., Swaney, D. L., Syka, J. E. P., Zabrouskov, V., and Coon, J. J. (2008) A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J. Proteome Res.* 7, 3127–3136.
- (59) Sobott, F., Watt, S. J., Smith, J., Edlmann, M. J., Kramer, H. B., and Kessler, B. M. (2009) Comparison of CID versus ETD-based MS/MS fragmentation for the analysis of protein ubiquitination. *J. Am. Soc. Mass Spectrom.* 20, 1652–1659.
- (60) Scott, N. E., Parker, B. L., Connolly, A. M., Paulech, J., Edwards, A. V. G., Crossett, B., Falconer, L., Kolarich, D., Djordjevic, S. P., Hojrup, P., Packer, N. H., Larsen, M. R., and Cordwell, S. J. (2011) Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Mol. Cell. Proteomics* 10 (2), M000031–MCP201.
- (61) Zubarev, R. A., Horn, D. M., Fridriksson, E. K., Kelleher, N. L., Kruger, N. A., Lewis, M. A., Carpenter, B. K., and McLafferty, F. W. (2000) Electron capture dissociation for structural characterization of multiply charged protein cations. *Annu. Rev. Mass Spectrom.* 72, 563–573.
- (62) Witze, E. S., Old, W. M., Resing, K. A., and Ahn, N. G. (2007) Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* 4, 798–806.
- (63) Chalkley, R. J., and Burlingame, A. L. (2001) Identification of GlcNAcylation sites of peptides and α -crystallin using Q-TOF mass spectrometry. *J. Am. Soc. Mass Spectrom.* 12, 1106–1113.
- (64) Kaasik, K., Kivimäe, S., Allen, J. J., Chalkley, R. J., Huang, Y., Baer, K., Kissel, H., Burlingame, A. L., Shokat, K. M., Ptáček, L. J., and Fu, Y.-H. (2013) Glucose sensor O-GlcNAcylation coordinates with phosphorylation to regulate circadian clock. *Cell. Metab.* 17, 291–302.
- (65) Stastna, M., and Van Eyk, J. E. (2012) Analysis of protein isoforms: Can we do it better? *Proteomics* 12, 2937–2948.
- (66) Guan, S., and Burlingame, A. L. (2010) High mass selectivity for top-down proteomics by application of SWIFT technology. *J. Am. Soc. Mass Spectrom.* 21, 455–459.
- (67) Marshall, A. G., Hendrickson, C. L., and Jackson, G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrom. Rev.* 17, 1–35.
- (68) Tipton, J. D., Tran, J. C., Catherman, A. D., Ahlf, D. R., Durbin, K. R., and Kelleher, N. L. (2011) Analysis of intact protein isoforms by mass spectrometry. *J. Biol. Chem.* 286, 25451–25458.
- (69) Medzihradsky, K. F., Zhang, X., Chalkley, R. J., Guan, S., McFarland, M. A., Chalmers, M. J., Marshall, A. G., Diaz, R. L., Allis, C. D., and Burlingame, A. L. (2004) Characterization of tetrahymena histone H2B variants and posttranslational populations by electron capture dissociation (ECD) Fourier transform ion cyclotron mass spectrometry (FT-ICR MS). *Mol. Cell. Proteomics* 3, 872–886.
- (70) Garcia, B. A., Thomas, C. E., Kelleher, N. L., and Mizzen, C. A. (2008) Tissue-specific expression and post-translational modification of histone H3 variants. *J. Proteome Res.* 7, 4225–4236.
- (71) Everetts, A. G., Zee, B. M., DiMaggio, P. A., Gonzales-Cope, M., Collier, H. A., and Garcia, B. A. (2013) Quantitative dynamics of the link between cellular metabolism and histone acetylation. *J. Biol. Chem.* 288, 12142–12151.
- (72) Shimbo, K., Hsu, G. W., Nguyen, H., Mahrus, S., Trinidad, J. C., Burlingame, A. L., and Wells, J. A. (2012) Quantitative profiling of caspase-cleaved substrates reveals different drug-induced and cell-type patterns in apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12432–12437.
- (73) Xu, J., Wang, A. H., Osés-Prieto, J., Makhijani, K., Katsuno, Y., Pei, M., Yan, L., Zheng, Y. G., Burlingame, A., Brückner, K., and Derynck, R. (2013) Arginine methylation initiates BMP-induced Smad signaling. *Mol. Cell* 51, 5–19.
- (74) Hunter, T. (2007) The age of crosstalk: Phosphorylation, ubiquitination, and beyond. *Mol. Cell* 28, 730–738.
- (75) Nguyen, L. K., Kolch, W., and Kholodenko, B. N. (2013) When ubiquitination meets phosphorylation: A systems biology perspective of EGFR/MAPK signalling. *Cell Commun. Signal.* 11, 52.
- (76) Sun, Y., Kaneko, S., Li, X. K., and Li, X. (2014) The PI3K/Akt signal hyperactivates Eya1 via the SUMOylation pathway. *Oncogene* 33, 1038–1044.
- (77) Ruthenburg, A. J., Li, H., Patel, D. J., and David Allis, C. (2007) Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* 8, 983–994.
- (78) Mukherjee, S., Keitany, G., Li, Y., Wang, Y., Ball, H. L., Goldsmith, E. J., and Orth, K. (2006) Yersinia YopJ acetylates and inhibits kinase activation by blocking phosphorylation. *Science* 312, 1211–1214.
- (79) Moll, U. M., and Petrenko, O. (2003) The MDM2-p53 interaction. *Mol. Cancer Res.* 1, 1001–1008.
- (80) Coutts, A. S., Adams, C. J., and La Thangue, N. B. (2009) p53 ubiquitination by Mdm2: A never ending tail? *DNA Repair* 8, 483–490.
- (81) Meijer, I. M. J., and van Leeuwen, J. E. M. (2011) ERBB2 is a target for USP8-mediated deubiquitination. *Cell. Signal.* 23, 458–467.
- (82) Latham, J. A., and Dent, S. Y. R. (2007) Cross-regulation of histone modifications. *Nat. Struct. Mol. Biol.* 14, 1017–1024.
- (83) Lee, J.-S., Smith, E., and Shilatifard, A. (2010) The language of histone crosstalk. *Cell* 142, 682–685.

- (84) Fischle, W., Tseng, B. S., Domann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., Hunt, D. F., Funabiki, H., and Allis, C. D. (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* 438, 1116–1122.
- (85) Hart, G. W., Slawson, C., Ramirez-Correa, G., and Lagerlof, O. (2011) Cross talk between O-GlcNAcylation and phosphorylation: Roles in signaling, transcription, and chronic disease. *Annu. Rev. Biochem.* 80, 825–858.
- (86) Allison, D. F., Wamsley, J. J., Kumar, M., Li, D., Gray, L. G., Hart, G. W., Jones, D. R., and Mayo, M. W. (2012) Modification of RelA by O-linked N-acetylglucosamine links glucose metabolism to NF- κ B acetylation and transcription. *Proc. Natl. Acad. Sci. USA* 109, 16888–16893.
- (87) Sakabe, K., and Hart, G. W. (2010) O-GlcNAc transferase regulates mitotic chromatin dynamics. *J. Biol. Chem.* 285, 34460–34468.
- (88) Ruan, H.-B., Nie, Y., and Yang, X. (2013) Regulation of protein degradation by O-GlcNAcylation: Crosstalk with ubiquitination. *Mol. Cell. Proteomics* 12, 3489–3497.
- (89) Ma, J., and Hart, G. W. (2014) O-GlcNAc profiling: From proteins to proteomes. *Clin. Proteomics* 11, 8. DOI: 10.1186/1559-0275-11-8.
- (90) Venne, A. S., Kollipara, L., and Zahedi, R. P. (2014) The next level of complexity: Crosstalk of posttranslational modifications. *Proteomics* 14, 513–524.
- (91) Ahearn, I. M., Haigis, K., Bar-Sagi, D., and Philips, M. R. (2011) Regulating the regulator: Post-translational modification of Ras. *Nat. Rev. Mol. Cell Biol.* 13, 39–51.
- (92) Beltrao, P., Bork, P., Krogan, N. J., and van Noort, V. (2013) Evolution and functional cross-talk of protein post-translational modifications. *Mol. Syst. Biol.* 9, 714.
- (93) Mali, P., Esvelt, K. M., and Church, G. M. (2013) Cas9 as a versatile tool for engineering biology. *Nat. Methods* 10, 957–963.
- (94) Inui, M., Miyado, M., Igarashi, M., Tamano, M., Kubo, A., Yamashita, S., Asahara, H., Fukami, M., and Takada, S. (2014) Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Sci. Rep.* 4, 5396.
- (95) Waldrip, Z. J., Byrum, S. D., Storey, A. J., Gao, J., Byrd, A. K., Mackintosh, S. G., Wahls, W. P., Taverna, S. D., Raney, K. D., and Tackett, A. J. (2014) A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics* 9, 1207–1211.
- (96) Zheng, Y., Zhang, C., Croucher, D. R., Soliman, M. A., St-Denis, N., Pasculescu, A., Taylor, L., Tate, S. A., Hardy, W. R., Colwill, K., Dai, A. Y., Bagshaw, R., Dennis, J. W., Gingras, A.-C., Daly, R. J., and Pawson, T. (2013) Temporal regulation of EGF signalling networks by the scaffold protein Shc1. *Nature* 499, 166–171.
- (97) Hsu, P. P., Kang, S. A., Rameseder, J., Zhang, Y., Ottina, K. A., Lim, D., Peterson, T. R., Choi, Y., Gray, N. S., Yaffe, M. B., Marto, J. A., and Sabatini, D. M. (2011) The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* 332, 1317–1322.
- (98) Francavilla, C., Rigbolt, K. T. G., Emdal, K. B., Carraro, G., Vernet, E., Bekker-Jensen, D. B., Streicher, W., Wikström, M., Sundström, M., Bellusci, S., Cavallaro, U., Blagoev, B., and Olsen, J. V. (2013) Functional proteomics defines the molecular switch underlying FGF receptor trafficking and cellular outputs. *Mol. Cell* 51, 707–722.
- (99) Lienhard, G. E. (2008) Non-functional phosphorylations? *Trends Biochem. Sci.* 33, 351–352.

Article 2: Bioinformatics analysis of thousands of TCGA tumors to determine the involvement of epigenetic regulators in human cancer²⁰⁸

BMC Genomics

Florian Gnad^{1*}, Sophia Doll^{2,3}, Gerard Manning¹, David Arnott², Zemin Zhang^{1,4}

¹Departments of Bioinformatics and Computational Biology, Genentech, USA.

²Protein Chemistry, Genentech Inc., South San Francisco, CA 94080, USA.

³Proteomics and Signal Transduction, Max-Planck-Institute for Biochemistry, Martinsried, Germany.

⁴Biodynamic Optical Imaging Center, College of Life Sciences, Peking University, Beijing 100871, China.

In this manuscript, we investigated the epigenetic landscape in cancer at a large scale. We first aimed to uncover somatic alterations in the epigenetic machinery in thousands of tumors from The Cancer Genome Atlas (TCGA) at the genomic level. To this end, a classification model that predicts the likelihood of epigenetic regulator genes to be an oncogene, tumor suppressor, or neutral gene was trained using 3,356 tumors from seven cancer types. We found several tumor suppressor genes among epigenetic regulators and gene expression and correlation network analysis showed that EZH2 was the most significantly over-expressed epigenetic regulator in cancer.

We followed up on this finding at the proteomic level, where we quantified global proteomic changes by mass spectrometry after EZH2 inhibition. We treated lymphoma cells with the EZH2 inhibitor EPZ-6438 (Tazemetostat) and performed quantitative MS-based proteomics on these samples. We found that the repressive epigenetic mark H3K27me3 was significantly downregulated and that EZH2 is co-regulated with parts of the cell cycle network. These results suggest that EZH2 drives the development of cancer via deregulated cell cycle regulation.

RESEARCH

Open Access

Bioinformatics analysis of thousands of TCGA tumors to determine the involvement of epigenetic regulators in human cancer

Florian Gnad^{1*}, Sophia Doll^{2,3}, Gerard Manning¹, David Amott², Zemin Zhang^{1,4*}

From VarI-SIG 2014: Identification and annotation of genetic variants in the context of structure, function and disease

Boston, MA, USA. 12 July 2014

Abstract

Background: Many cancer cells show distorted epigenetic landscapes. The Cancer Genome Atlas (TCGA) project profiles thousands of tumors, allowing the discovery of somatic alterations in the epigenetic machinery and the identification of potential cancer drivers among members of epigenetic protein families.

Methods: We integrated mutation, expression, and copy number data from 5943 tumors from 13 cancer types to train a classification model that predicts the likelihood of being an oncogene (OG), tumor suppressor (TSG) or neutral gene (NG). We applied this predictor to epigenetic regulator genes (ERGs), and used differential expression and correlation network analysis to identify dysregulated ERGs along with co-expressed cancer genes. Furthermore, we quantified global proteomic changes by mass spectrometry after EZH2 inhibition.

Results: Mutation-based classifiers uncovered the OG-like profile of *DNMT3A* and TSG-like profiles for several ERGs. Differential gene expression and correlation network analyses revealed that *EZH2* is the most significantly over-expressed ERG in cancer and is co-regulated with a cell cycle network. Proteomic analysis showed that EZH2 inhibition induced down-regulation of cell cycle regulators in lymphoma cells.

Conclusions: Using classical driver genes to train an OG/TSG predictor, we determined the most predictive features at the gene level. Our predictor uncovered one OG and several TSGs among ERGs. Expression analyses elucidated multiple dysregulated ERGs including *EZH2* as member of a co-expressed cell cycle network.

Background

The epigenetic landscape has become an important research topic within oncology. Epigenetic regulatory mechanisms include DNA methylation, covalent histone modification, and chromatin remodeling mediated by the SWI/SNF complex. DNA methylation typically reduces gene expression and is catalyzed by three major DNA methyltransferases (DNMTs) [1]. In comparison, a larger and more diverse panel of proteins regulates gene expression as writers, readers or erasers of posttranslational histone modifications [2,3] (Additional file 1). Acetyl marks

are written by histone acetyltransferases (HATs), read by bromodomain containing proteins, and erased by histone deacetylases (HDACs). Analogously, histone methyl marks are written by methyltransferases (HMTs) and erased by demethylases (HDMTs). The multi-subunit SWI/SNF chromatin-remodeling complex modulates gene expression via nucleosome repositioning [4].

Perturbing the epigenetic machinery can lead to uncontrolled cellular proliferation and altered apoptosis [5,6]. Consequently, alterations of epigenetic regulators and histone marks are frequently observed in cancer and numerous compounds have been reported to be effective against cancer cells by inhibiting epigenetic proteins and reversing the effect of epigenetic modifications [7,8]. Clinically approved epigenetic drugs include the DNMT

* Correspondence: gnad@gene.com; zemin@pkuedu.cn

¹Departments of Bioinformatics and Computational Biology, Genentech, USA
Full list of author information is available at the end of the article

inhibitors vidaza and decitabine [9,10], and HDAC inhibitors vorinostat and romidepsin [11,12] for treatment of myelodysplastic syndrome and cutaneous T cell lymphoma, respectively. Multiple pharmaceutical companies are targeting the histone methyltransferase EZH2 for cancer treatment. EZH2 inhibitors EPZ-5687 [13] from Epizyme[®] and GSK-2816126 [14] from GlaxoSmithKline[®] for the treatment of non-Hodgkin's lymphoma are currently in clinical phases I/II and I, respectively.

The increasing interest in the role of epigenetic mechanisms in cancer has been accompanied by technological breakthroughs and large-scale initiatives to profile large numbers of human tumors. The TCGA network has produced genome and transcriptome sequencing data for thousands of tumors, allowing systematic analysis of molecular defects in cancer [15-17]. Integrative analyses as seen in the TCGA Pan-Cancer project [18] can uncover OGs and TSGs, identify novel biomarkers, and classify molecular subtypes. Most of the current driver identification approaches aim to uncover somatic alterations, point mutations in particular, that occur at a statistically significant rate in cancer. Alternatively, using genomic profiles of known OGs and TSGs as a reference, machine learning based predictors can be trained to identify cancer genes [19].

Both methodologies are founded on features of classical drivers, which are mainly characterized by significant mutation or copy number patterns. Many other genes show consistent deregulated expression in cancer, but are not classified as drivers, because their impact on the development of cancer is not clear. Mutation based analyses might therefore underestimate the roles of genes that drive cancer via increased or decreased expression. Here we analyze the genomic landscapes of thousands of tumors to pinpoint molecular aberrations within ERG families. We used mutation, expression, and copy number alterations as features to predict OGs and TSGs among 187 epigenetic regulators based on both published and reprocessed TCGA data. Differential gene expression analysis revealed ERGs with frequent distorted expression in cancer. We further aimed to identify genes that are co-expressed with ERGs.

Materials and methods

Definition of ERGs and application of predictors

We classified ERG families by the presence of domains associated with writing, reading and erasing epigenetic marks, and defined the relationships between their members by sequence similarity. For deacetylases, methyltransferases, demethylases, and bromodomain-containing proteins, the amino acid sequences of the corresponding domains were used to determine conservation by multiple sequence alignment. Domain annotations were retrieved from the UniProt database ([\[www.uniprot.org\]\(http://www.uniprot.org\)\) \[20\]. Sequences of 'SET', 'SAM-dependent MTase PRMT-type' and 'DOT1' domains were derived for methyltransferases. 'JmjC' and 'SWIRM' domains were characteristic for demethylases. Deacetylases contained 'Histone deacetylase' or 'Deacetylase sirtuin-type' domains, while each bromodomain containing protein contained at least one domain described as 'Bromo' or 'Bromo 1' in UniProt. When proteins had multiple copies of a domain, the N-terminal domain was used. Full-length sequences were used for acetyltransferases and members of the SWI/SNF complex, since their catalytic domains are not clearly defined. We created multiple sequence alignments for each family with ClustalW2 \(<http://www.ebi.ac.uk/Tools/msa/clustalw2/>\) \[21\] using default parameters. Phylogenetic trees were calculated with Jalview 2.8 \[22\] based on average distance minimization and visualized in iTOL 2.1 \(<http://itol.embl.de>\) \[23,24\].](http://</p></div><div data-bbox=)

Mutation and copy number data

To create gene-alteration profiles for all human genes, mutation and copy number data from tumors across the following published TCGA cancer types were retrieved using cBioPortal (<http://cbioportal.org>) [25,26]: urothelial bladder carcinoma (BLCA) [27], breast carcinoma (BRCA) [28], colon and rectal carcinoma (COAD, READ) [29], glioblastoma (GBM) [30], chromophobe renal cell carcinoma [31] (KICH), clear cell renal carcinoma (KIRC) [32], acute myeloid leukemia (LAML) [33], lung adenocarcinoma (LUAD) [34], lung squamous cell carcinoma (LUSC) [35], ovarian carcinoma (OV) [36], gastric adenocarcinoma (STAD) [37], papillary thyroid carcinoma (THCA) [38], and endometrial carcinoma (UCEC) [39]. The CGDSR R package functions *getMutationData* and *getProfileData* were recursively applied for all RefSeq genes. We distinguished between missense mutations with high (HiFI) or low (LoFI) functional impact based on MutationAssessor [40]. Mutations with predicted "medium" or "high" functional impacts were defined as HiFI mutations, while mutations with predicted "neutral" or "low" functional impacts were defined as LoFI mutations. Loss of function (LOF) mutations were determined as the sum of nonsense and frameshift mutations. In addition to non-synonymous mutations from cBioPortal, we retrieved silent mutations directly from the TCGA Data Portal (<https://tcga-data.nci.nih.gov/tcga/>). Benign mutations were defined as the combination of silent and LoFI mutations. Copy number levels from cBioPortal were classified as 'deep loss', 'single-copy loss', 'diploid', 'low-level gain' or 'high-level gain' by GISTIC [41]. The extents of copy number deletions and amplifications for each gene in each cancer study were determined as the proportions of tumors with 'deep loss' and 'high-level gain' changes, respectively. R [42] was used to format

mutation and copy number data for annotation of trees in iTOL.

Expression data and differential gene expression analysis

To identify differential gene expression between tumors and healthy tissues, TCGA RNAseq raw reads were downloaded for available tumor types (BLCA, BRCA, COAD, KICH, KIRC, LUAD, LUSC, STAD, TCHA, UCEC) and processed by our GSNAP [43] based transcriptome analysis pipeline [44]. RNAseq data for both tumors and healthy tissues were not available for GBM, LAML and OV. RNAseq reads were first aligned to ribosomal RNA sequences to remove ribosomal reads. Remaining reads were aligned to the human reference genome (NCBI Build 37) using GSNAP version '2012-01-11', allowing maximum of 2 mismatches per 75 base sequence (parameters: "-M 2 -n 10 -B 2 -i 1 -N 1 -w 200000 -E 1 -pairmax-rna = 200000"). Gene expression was quantified with RPKM values (reads mapping to a gene per kilobase of transcript per million reads sequenced) and variance stabilized counts derived from the number of reads mapped to each RefSeq gene. The DESeq R package [45] was applied to estimate size factors, obtain dispersion estimates, and measure differential gene expression between tumors and healthy tissues using default parameters. Results were reported as fold changes and associated adjusted p-values. In addition to DESeq based negative binomial generalized linear models for differential expression significance, we defined genes with tumor exclusive expression (genes that are expressed in tumors but not in healthy tissues), if their 90% quantile expression levels in all healthy tissues were equal to the expression levels of pseudo counts, but minimum 1 RPKM in the tumors of at least one cancer type.

Prediction of OGs and TSGs

Following the methodology for parameter tuning as described in the TUSON explorer [19], we applied the Lasso approach [46] to identify the most reliable features for predicting OGs and TSGs. Lasso minimizes the residual sum of squares (RSS) with a constraint ("L1 penalty") on the sum of the absolute values of the coefficients β_j for all predictors p :

$$RSS + \lambda \sum_{j=1}^p |\beta_j|$$

The L1 penalty has the effect of shrinking some of the coefficients to zero when the tuning parameter λ is sufficiently large. As a result, lasso models select the most predictive subsets of features at specified λ values.

For both feature selection and training, we used 49 OGs and 49 TSGs from the Cancer Gene Census (CGC) [47] with experimentally validated involvement in tumorigenesis as provided by TUSON. Genes that have

not been associated with cancer development according to CGC or the Entrez gene database formed a set of 10,900 NGs. Using TCGA data we employed 48 features associated with mutation, expression, or copy number alterations for each human gene (Additional file 2). To prevent imbalanced classifications, we created 1000 random NG sets of size 150 each. Feature selections and predictions were conducted for OGs and TSGs separately.

Using the 'cv.glmnet' function from the R package glmnet [46], we trained lasso based binomial classification models for each random NG set against all OGs or TSGs. We used 20-fold cross validations to determine tuning parameter λ yielding minimum cross-validated errors. Features were defined as reliable for OG or TSG prediction, respectively, if the associated β coefficients were not zero in at least 90% of the 1000 resulting classifiers. While TUSON applied the lasso approach for feature selection only, we also used the resulting fitted logistic regression models for prediction. We applied glmnet's 'predict' function to each of the 1000 fitted models based on optimal λ values and the respective optimal feature subsets. This resulted into 1000 sets of predicted OGs and TSGs. Using a bagging based ensemble classification approach, we applied binary classifications of all human genes based on a 90% majority vote.

Notably, we used all 49 OGs and 49 TSGs as positive sets for training. In the absence of a separate test set, prediction accuracies were therefore measured as average 20-fold cross validation based areas under the curve (AUC) across the 1000 classifiers from the training step.

Co-expression analysis

To estimate the strength of the pairwise linear relationship between the expression levels all human genes in healthy tissues, Pearson's correlation coefficients were calculated based on WGCNA, an R package for weighted correlation network analysis [48]. Using DESeq [45], variance stabilized RNAseq count data were used as a measure of gene expression. Expression data of all non-tumor samples were merged and analyzed in a combined approach. We applied hard thresholding ($R > 0.85$) to convert the resulting $19,115 \times 19,115$ similarity matrix into an adjacency matrix. Using R we turned the adjacency matrix into a network file that can be imported in Cytoscape [49]. Known cancer genes were defined by the Cancer Gene Census (CGC) [47]. In total 25 out of 501 CGC genes were ERGs.

Sample preparation and mass spectrometry analysis

To analyze the effect of EZH2 inhibition on the proteome, we applied quantitative mass spectrometry based proteomics to a non-Hodgkin's lymphoma B cell line, WSU-DLCL2. Cells were cultivated in SILAC RMP1

1640 medium containing $^{13}\text{C}6^{15}\text{N}2$ -lysine (Lys8) and $^{13}\text{C}6^{15}\text{N}4$ -arginine (Arg10), as described [50]. After fully labeling, as assessed by quantitative mass spectrometry, cells were treated with the EZH2 inhibitor EPZ-6438 (Epizyme[®], Cambridge, MA) (provided by LT Pharma-Tech Inc[®]) (250 nM) for 2, 4, 6 or 8 days.

Cell pellets were lysed in 8 M Urea, 20 mM HEPES buffer by sonication and clarified by centrifugation at $16,000 \times g$ for 10 min. Protein content was measured using the Pierce BCA protein assay (Thermo Scientific) by fluorescence spectrometry. SILAC-labeled proteins were combined with an equal amount of unlabeled proteins. Proteins were reduced with dithiothreitol and alkylated with iodoacetamide prior to tryptic in-gel digestion. 100 μg of the heavy/light protein mix was loaded and separated by SDS-PAGE on a 4-12% NuPAGE Bis-Tris gel (Invitrogen) and stained with SimplyBlue Coomassie (Invitrogen). Gel bands were excised, separated into 16 fractions, and destained followed by overnight trypsin digestion at 37°C in 50 mM ammonium bicarbonate.

Nanoflow LC-MS/MS analysis of tryptic peptides was conducted on an LTQ-Orbitrap XL (ThermoFisher) in combination with a Waters nanoAcquity UPLC system, as described [50]. The mass spectrometer was operated in data-dependent mode and tandem mass spectra were searched against the UniProt human database using Mascot and a maximum false positive rate of 2% for proteins.

Histone purification and H3K27me3 quantification

H2A, H2B, H3, and H4 histones were purified with a commercially available histone purification kit (Active Motif) accordingly to the manufacturer's instruction. Histone concentrations were measured using the Direct Detect[®] Spectrometer (EMD Millipore). Heavy and light amino acid-labeled histones were mixed in a 1:1 ratio. Histones were propionylated, quenched by hydroxylamine followed by tryptic digestion overnight and phenyl isocyanate labeling. Histone peptides were then analyzed by capillary reverse phase ultra high-pressure liquid chromatography-electrospray ionization tandem mass spectrometry on an Orbitrap mass spectrometer. Briefly, 1 μg of desalted histone peptides were injected on 1.7 μm BEH-C18 column (Waters) and eluted over the course of 90 minutes with an acetonitrile gradient. Spectra were acquired in a "top-15" data-dependent experiment. Data were further processed with Fishtones (<http://research-pub.gene.com/fishtones-js/howto/>).

Clustering of time courses

Using the R package Mfuzz [51], log₂ ratios of protein intensity time profiles were clustered based on the fuzzy c-means (FCM) soft partitioning clustering algorithm.

We used $c = 3$ and $m = 1.7$ as parameters, where c is the number of clusters and m is the fuzzification parameter. Membership values ranging from 0 to 1 reflect the similarities of each time profile to its associated cluster.

Gene ontology analysis

We used Cytoscape [49] and BinGO [52] to derive biological functions that were significantly overrepresented in co-expressed gene networks or proteins with intensity changes after EZH2 inhibition. The significance of overrepresented gene ontology annotations in these sets compared to entire human proteome was calculated on the basis of hypergeometric models and Benjamini Hochberg false discovery rate correction.

Results

Definition of ERG families

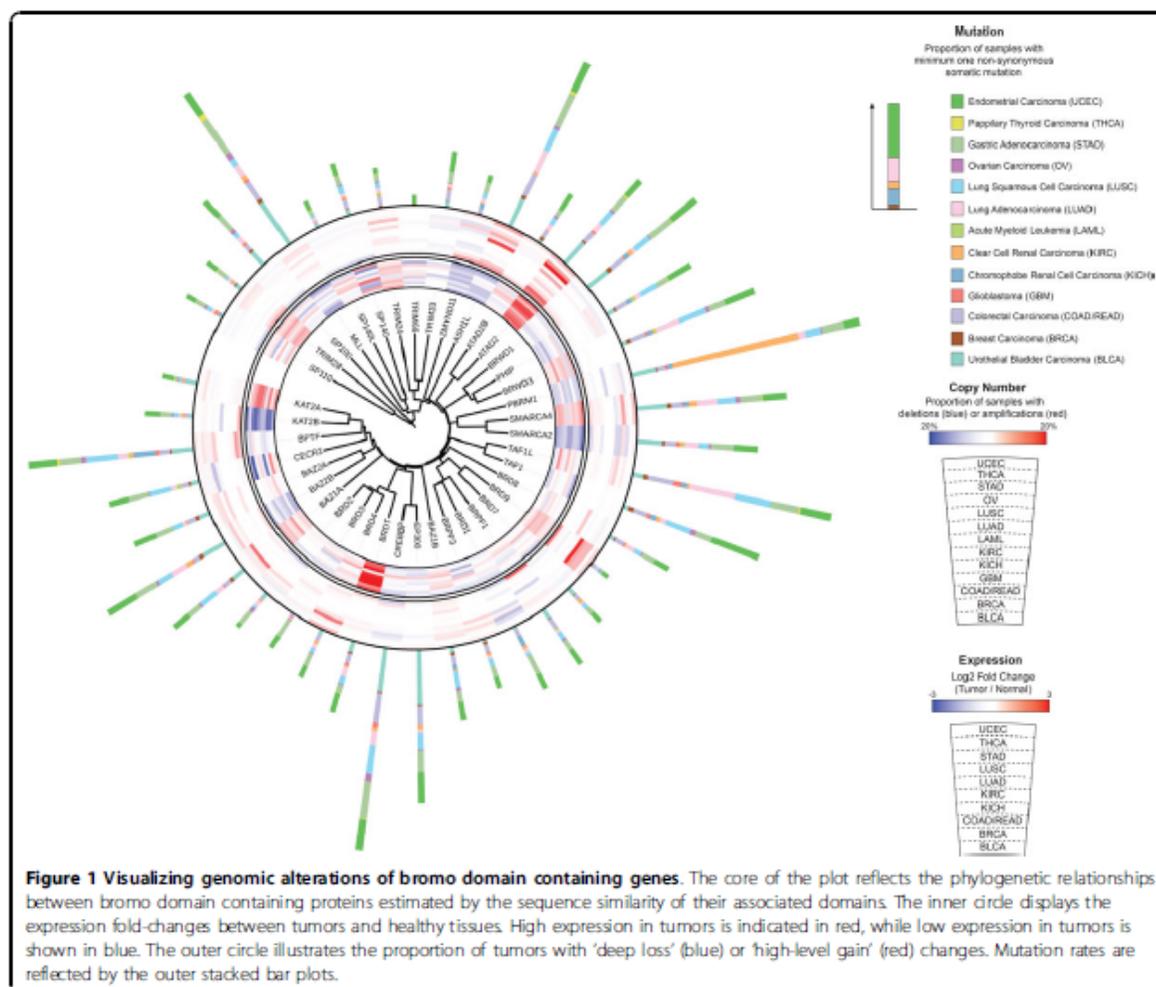
We defined ERG families and their members by the presence of domains associated with writing, reading and erasing epigenetic marks as described [53]. The resulting panel of 187 epigenetic regulators comprised 3 DNMTs, 58 HMTs, 32 HDMTs, 18 HATs, 18 HDACs, 41 bromodomain proteins, and 20 members of the SWI/SNF complex (Additional files 1, 3 and 4). Their phylogenetic relationships were estimated by the sequence similarity of associated domains. Using mutation, copy number and expression data from 5943 tumors across 13 TCGA cancer types, we set out to determine the involvement of the defined ERGs in human cancer by OG/TSG prediction, differential expression and correlation network analysis.

Prediction of cancer driver genes

The most common approach to uncover cancer driver genes is to identify somatic alterations that occur at a statistically significant rate. As an alternative approach, machine learning based classifications use characteristics of known OGs and TSGs as a reference (training set) to predict cancer genes [19]. We implemented a similar approach to test its applicability in our tumor panel and to identify cancer drivers within ERG families.

Construction of gene-alteration profiles

To characterize known cancer drivers and to identify genes with similar features, we constructed 'alteration profiles' for all human genes in each individual cancer type as well as in the combined set of tumors (Additional file 5, Materials and Methods). Each gene profile contained 48 features measuring various types of alterations in cancer (Additional file 2). For members of the defined ERG gene families, we applied iTOL to visualize four of the 48 established features (Figure 1 and Additional file 6). These include the proportions of tumors with significant copy number alterations, non-synonymous



sequence mutations within the gene coding region, and the degree of differential expression between tumors and adjacent normal tissues. The frequencies of copy number deletions or amplifications for each gene were determined as the proportions of tumors with 'deep loss' or 'high-level gain' changes based on GISTIC calculations [41], respectively. To measure the degree of dysregulated expression for each gene, we developed fold-change and p-value based scores reflecting the significance of differential gene expression based on negative binomial generalized linear models (Materials and Methods). The majority (44 of 48) of the integrated features, however, describe the frequencies of various sequence mutation classes. To exclude the effect of protein size [54], we normalized mutation frequencies by the background mutation rate or coding sequencing length. To distinguish between missense mutations with high (HiFI) or low (LoFI) functional impact, we used MutationAssessor [40], which is known to have high

accuracy [55]. Loss of function (LOF) mutations were defined as the combination of nonsense and frameshift mutations. Benign mutations (as the combination of silent and LoFI) mutations reflect the background mutation rate of each gene. As a measure of the preferred occurrence of specific point mutations within a gene, termed 'mutation hot spots', we calculated entropy based 'mutation selection scores' as described [19].

Selection of features reliable for OG and TSG prediction

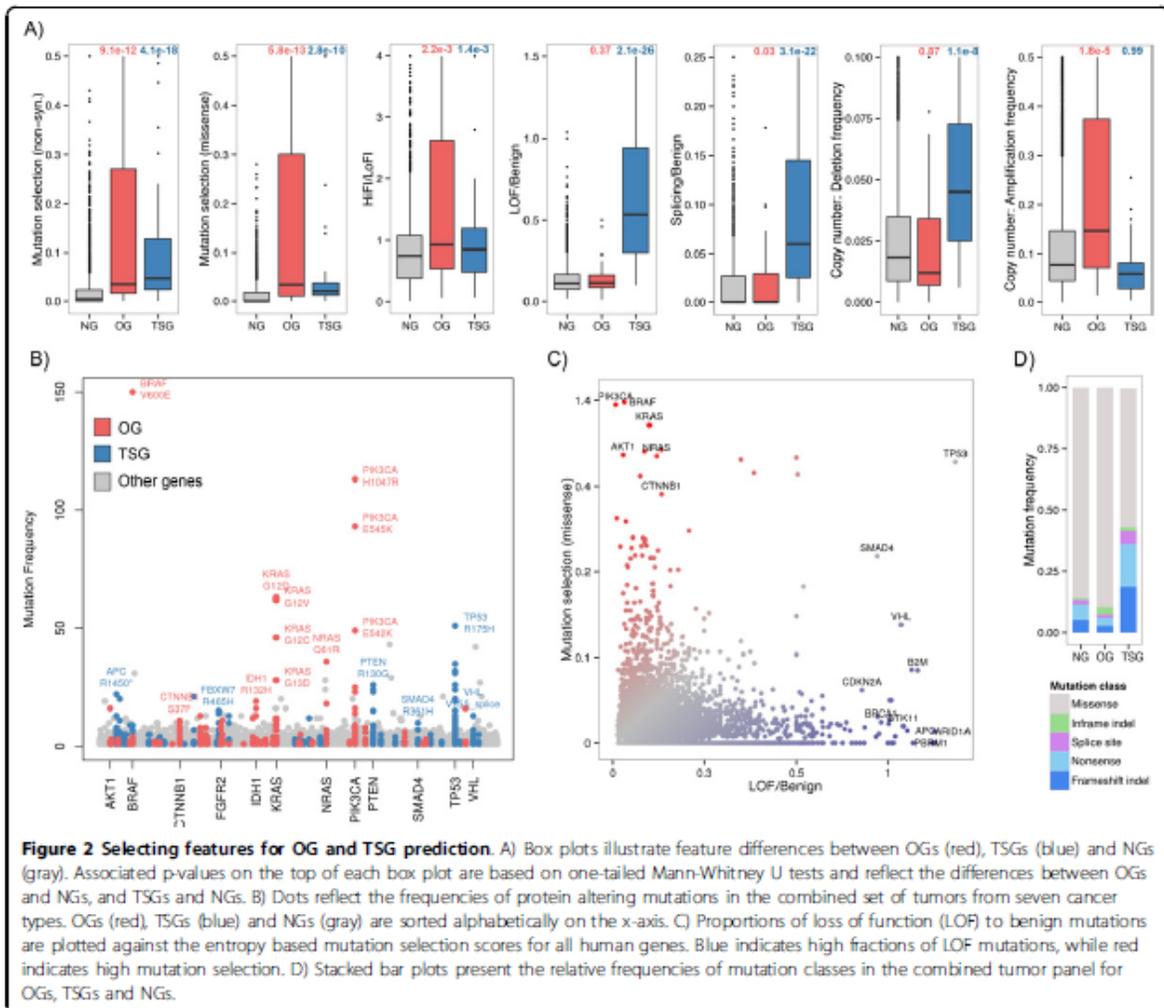
For the training of binary classifiers and for the selection of predictive features, we obtained OGs and TSGs from the Cancer Gene Census (CGC) [47] as well as NGs, as described [19]. To select features from the generated gene-alteration profiles that distinguish cancer drivers from NGs, we followed the methodology for parameter tuning from the TUSON (TUMor Suppressor and ONcogene) explorer [19] (Materials and Methods). We used the least absolute shrinkage and selection operator

(Lasso) method [46] to identify the most reliable out of 48 parameters for predicting cancer genes.

For OG prediction, the most reliable parameters were the occurrence of mutation hot spots, represented by the mutation selection score ($p = 5.8 \times 10^{-13}$, one-tailed Mann-Whitney U test, β coefficient = 5.3), the ratio of HiFI to LoFI missense mutations ($p = 2.2 \times 10^{-3}$, $\beta = 0.06$), and the amplification frequency ($p = 1.8 \times 10^{-5}$, $\beta = 2.48$) (Figure 2A). These features indicate that canonical OGs are characterized by copy number amplifications or recurrent missense mutations with high impact on protein function. Examples for such somatic mutation hot spots include V600E in BRAF (265 tumors), H1047R in PIK3CA (113 tumors), or G12D in KRAS (63 tumors) (Figure 2B, Additional file 7). Overall, BRAF ($S_m = 2.71$), PIK3CA ($S_m = 1.25$), KRAS ($S_m = 1.11$), and IDH1 ($S_m = 0.82$) showed the highest selection scores

for missense mutations (S_m) among OGs. Interestingly, known copy number driven OGs including MYC ($S_m = 0$), ALK ($S_m = 0$), and SOX2 ($S_m = 0$) showed significantly low preference for mutation hot spots ($p = 2.7 \times 10^{-18}$, one-tailed Mann-Whitney U test comparing amplified versus non-amplified OGs based on CGC annotation). Notably, none of the expression parameters was selected as predictive feature ($p > 0.4$) implying that the OGs from the training set are not consistently over-expressed in cancer.

The most reliable feature set for TSG prediction included the ratio of LOF to benign mutations ($p = 2.1 \times 10^{-26}$, $\beta = 2.13$), splicing to benign mutations ($p = 3.1 \times 10^{-22}$, $\beta = 1.85$) and the frequency of homozygous copy number losses ($p = 1.1 \times 10^{-8}$, $\beta = 1.08$) (Figures 2A and 2D). In addition, given the significant underrepresentation of TSGs in amplicons, the Lasso approach also



selected the amplification frequency as predictive ($p = 2.8 \times 10^{-3}$, $\beta = -1.06$). This indicates that canonical TSGs are characterized by copy number loss or mutations that have deleterious effects on protein function. Interestingly, multiple TSGs showed significantly recurrent LOF or splice site mutations including APC (R1450*; 22 tumors) and VHL (V155splice; 13 tumors) as well as missense mutations including TP53 (R175H; 51 tumors) and PTEN (R130G; 26 tumors) (Figure 2B, Additional file 7). Consequently, the selection scores for non-synonymous mutations ($p = 4.1 \times 10^{-18}$, $\beta = 1.89$) were high in the training set (Figures 2A and 2C) and thus selected for TSG prediction by Lasso. Expression parameters were not selected as reliable features ($p > 0.4$) for TSG prediction.

Applying cancer gene classifiers to ERGs reveals more TSGs than OGs

To uncover cancer drivers among ERGs, we applied the trained classification models that were used for feature selection. We used all 49 known OGs and TSGs for the training and feature selection step, because the size of the positive set was relatively small for machine learning. Therefore, the assessment of our predictors relied on 20-fold cross-validations instead of an independent test set. The average areas under the curve (AUC) as measure of prediction accuracy for OG and TSG classifications were 84.21% and 92.17%, respectively (Materials and Methods).

To identify cancer genes that are driven by mutation, we applied the predictors to the defined ERGs using the most predictive mutation parameters only. Overall five ERGs including the SWI/SNF complex subunits PBRM1, ARID1A, and SMARCD1 were classified as TSGs (Figure 3A). Exclusion of copy number data from the feature set yielded the same set. *ARID1A* had the highest ratio of LOF to benign mutations among ERGs, and was mutated in 25.4% of urothelial bladder tumors, 31.1% of gastric tumors, and 33.5% of endometrial tumors. Overall 72.2% of all non-synonymous mutations in *ARID1A* were LOF. *PBRM1* was mutated in 36.5% of clear cell renal carcinomas, of which 75.0% were LOF. *SETD2* was also classified as TSG with 39.9% of all non-synonymous mutations classified as LOF. Consistent with the mutation profiles of TSGs in the training set, multiple LOF mutations had a non-random distribution within predicted TSGs in more than one tumor (Figures 3B and 3C). *ARID1A*, for example, showed a frameshift mutation at position 1848 in 20 tumors (Figure 3C). Overall, with the exception of alterations in *DNMT3A*, the most recurrent mutations within ERG families were associated with loss of function (Figure 3B).

Using the missense mutation selection score and the ratio of HiFi to LoFi mutations, only *DNMT3A* was predicted as OG. This result reflects the lack of recurrent

and potentially activating hotspot missense mutations within ERG families in our tumor panel. With the exception of *DNMT3A*, we detected nonsense mutations and indels, but no missense point mutations among ERGs that occurred in more than four tumors. The driver classification of *DNMT3A* by our OG predictor can be attributed to the occurrence of a mutation hot spot in acute myeloid leukemia. In total 28 (14.4%) of the 195 tumors showed a missense mutation on position 882 resulting in an overall mutation selection score of 0.55.

We expected *EZH2* to be classified also as an OG, since it is a validated target pursued by multiple pharmaceutical companies. Activating mutations within the SET domain of *EZH2* are frequent in non-Hodgkin's lymphoma [56], but were not found as recurrent in the analyzed cancer types.

With copy number data as an additional feature, completing the set of selected predictive parameters, *ACTL6A* and *ATAD2* were the only predicted OG among ERGs. However, since the amplified genomic regions harboring these genes were typically very large, with an average length exceeding 50 Mb, it is equally likely that both genes are only amplified as a passenger genes.

Detection of ERGs with consistent over- or under-expression in cancer

The Lasso-based feature selection for OG/TSG prediction showed that canonical cancer drivers are usually characterized by significant mutation patterns or copy number alterations (Figure 2A), but not by consistent gene expression patterns. Consequently, while our machine learning approach enabled us to uncover cancer driver-like mutation and copy number alterations among ERGs, significant gene expression patterns could not be detected by prediction.

To pinpoint ERGs with consistently higher or lower expression in cancer, we determined the differential expression significance across the ten cancer types with available RNAseq data using negative binomial generalized linear models (GLM) [45] (Materials and Methods). To assess the overall significance of differential expression in cancer for each gene, we combined the p-values resulting from the cancer type specific analyses using Fisher's probability test. Overall 11 ERGs showed consistent up-regulation in all cancer types with combined p-values (p_F) lower than 0.001 (Figure 4A, Additional file 8). *EZH2* showed the most significant over-expression in tumors ($p_F = 3.18 \times 10^{-112}$) (Figure 5A) not correlating with copy number alterations ($p = 0.87$ based on linear regression between RPKM expression levels and total copy number) (Materials and Methods). The *MYC* cofactor

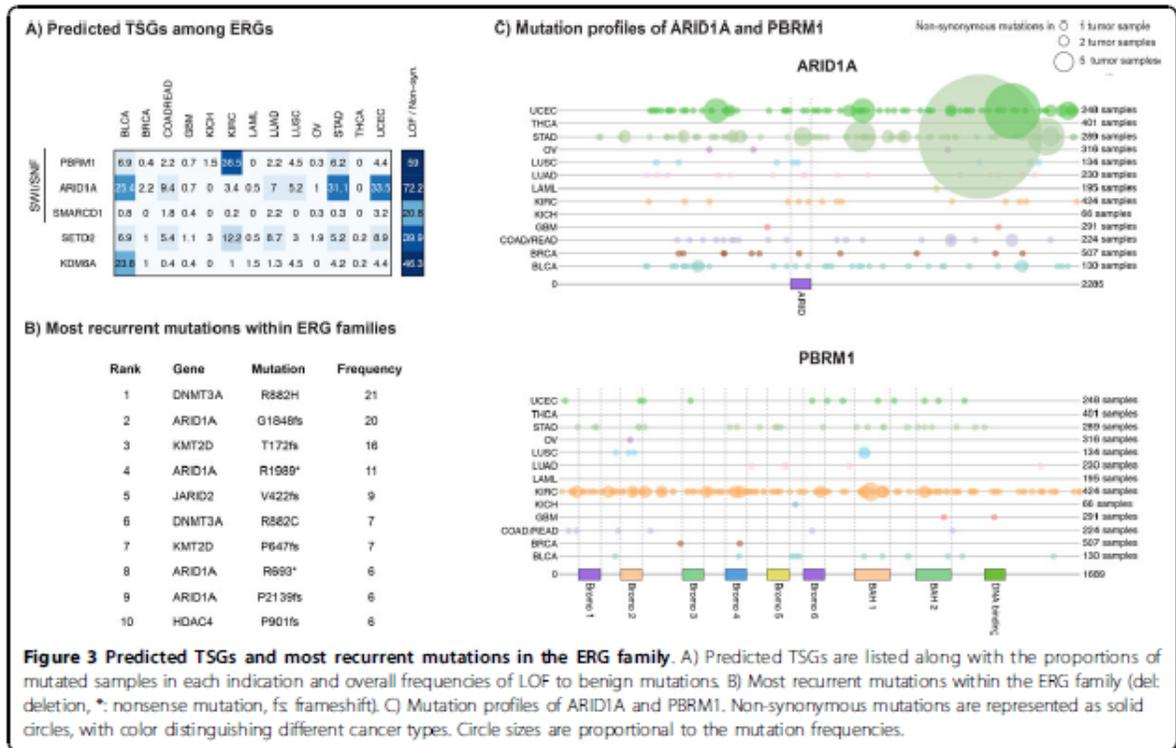
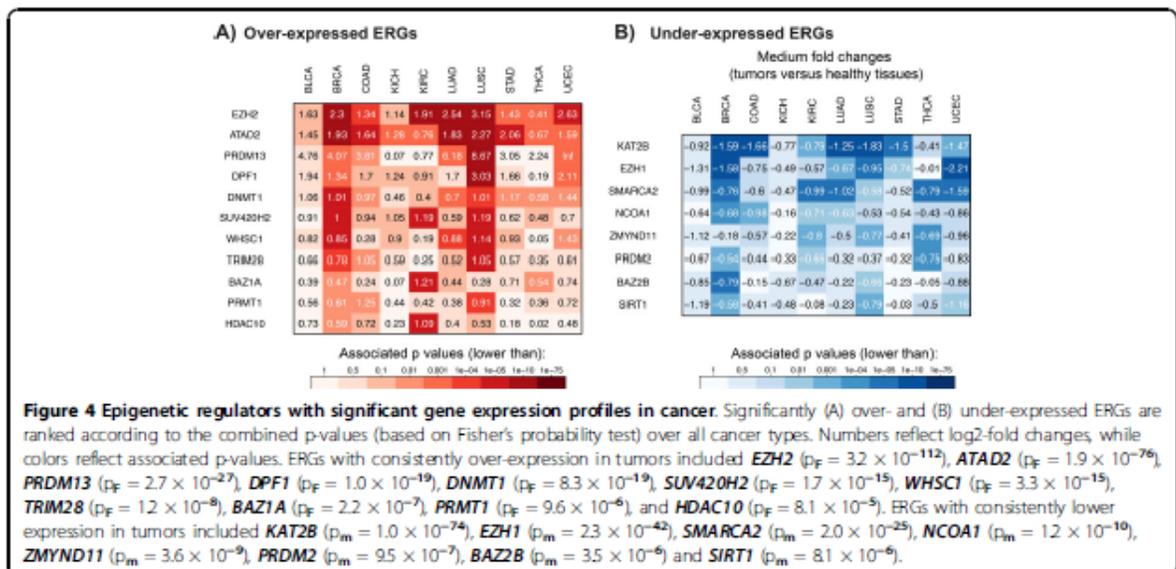
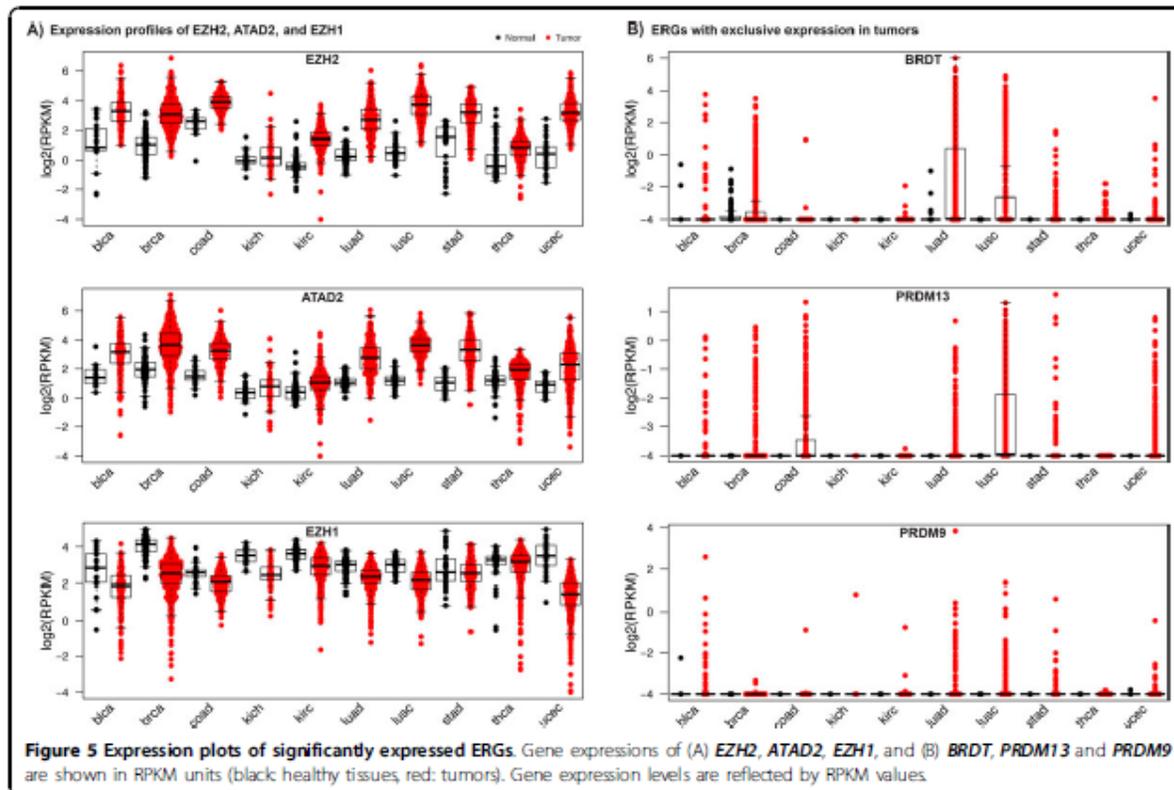


Figure 3 Predicted TSGs and most recurrent mutations in the ERG family. A) Predicted TSGs are listed along with the proportions of mutated samples in each indication and overall frequencies of LOF to benign mutations. B) Most recurrent mutations within the ERG family (del: deletion, *: nonsense mutation, fs: frameshift). C) Mutation profiles of ARID1A and PBRM1. Non-synonymous mutations are represented as solid circles, with color distinguishing different cancer types. Circle sizes are proportional to the mutation frequencies.

[57,58] bromodomain reader *ATAD2* (ATPase family, AAA domain containing 2) was the second most significantly over-expressed ERG ($p_F = 2.6 \times 10^{-25}$). Expression levels of *ATAD2* correlated significantly with copy number changes ($p = 3.1 \times 10^{-12}$) consistent

with the length of the *MYC* amplicon that spans the genomic region of *ATAD2* in 98% of the cases. Other ERGs with homogeneous over-expression in cancer included *PRDM13*, *DPF1*, *DNMT1*, *SUV420H2*, *WHSC1*, *TRIM28*, *BAZ1A*, *PRMT1*, and *HDAC10*.





Quantile-based differential expression analysis (Materials and Methods) revealed that *BRDT*, *PRDM9*, and *PRDM13* were exclusively expressed in tumors but not in paired healthy tissues (Figure 5B). With high expression levels across several cancer types, the testis-specific *BRDT* gene [59,60] displayed characteristics of a cancer-testis (CT) gene. CT genes are genes with normal expression restricted to adult testicular germ cells, and yet are aberrantly activated and expressed in various cancer types [61]. As such CT genes are interesting targets in cancer therapy. As previously observed in non-small cell lung cancers [62], differential expression analysis between *BRDT* expressing and non-expressing LUSC tumors revealed co-expression with canonical CT genes such as *MAGE-A11*, *GAGE4*, *GAGE5*, *GAGE6* and *GAGE12I* (Additional file 9, Materials and Methods). Almost all co-expressed genes also showed tumor-specific expression in LUSC. *PRDM9* and *PRDM13* also showed exclusive expression in cancer, but their biological roles in these tumors are unclear.

Consistently down-regulated ERGs included *KAT2B*, *EZH1*, *SMARCA2*, *NCOA1*, *ZMYND11*, *PRDM2*, *BAZ2B* and *SIRT1*, which showed significantly lower expression ($p_F < 0.001$) in tumors compared to healthy tissues (Figure 4B). Comparing the sets of over- or under-expressed ERGs

showed that closely related genes such as *KAT2A* and *KAT2B* exhibited different expression profiles. *EZH2* and *EZH1*, for example, form PRC2 (Polycomb repressive complex 2) complexes with similar functionalities [63], but opposite expression profiles. As another example, while *PRDM9* and *PRDM13* were exclusively expressed in tumors, *PRDM2* was consistently down-regulated in tumors.

Taken together the resulting panel of significantly over- or under-expressed ERGs form an interesting candidate set of genes that potentially drive the development of cancer via dysregulated expression. This model is generally not applicable to classical OGs and TSGs, but might hold true for ERGs.

Co-expression network analysis

In addition to the identification of significant expression patterns in tumors, we used the expression levels in healthy tissues to detect co-expressed genes under normal conditions. The main objective of this analysis was to uncover the involvement of ERGs in co-expression networks, which frequently form jointly regulated functional modules [64]. Co-expressed genes can have similar biological activities and even physically interact. In some cases co-expression may reflect that one gene encoding protein regulates the activity of the other gene.

Using variance stabilized TCGA RNAseq count data as a measure of gene expression, we analyzed co-expression networks by estimating pairwise linear relationships between all protein coding human genes (Materials and Methods).

Based on hard thresholding ($r > 0.85$) we converted the resulting $19,115 \times 19,115$ similarity matrix into an adjacency matrix, which contains binary information (0: no co-expression; 1: co-expression) about pairwise co-expression. Transforming the adjacency matrix into nodes (genes) and edges (co-expression) resulted in one major network with 2465 genes including 37 ERGs, and 11 separate networks with 8 to 112 genes (Figure 6A).

EZH2 is a member of a cell cycle network

Using gene ontology enrichment (Materials and Methods) we found that the discrete co-expression networks, which were not connected to any node of the main network, were associated with specific biological functions such as muscle contraction ($p_{GO} = 8.9 \times 10^{-11}$), collagen fibril organization ($p_{GO} = 4.4 \times 10^{-10}$), tissue development ($p_{GO} = 1.4 \times 10^{-5}$), oxidative phosphorylation ($p_{GO} = 3.1 \times 10^{-24}$), or regulation of secretion ($p_{GO} = 2.3 \times 10^{-3}$) (Additional file 10). Two co-expression modules contained 37 members of the protocadherin family representing tightly linked gene clusters α and γ , consequently associated with cell-cell adhesion

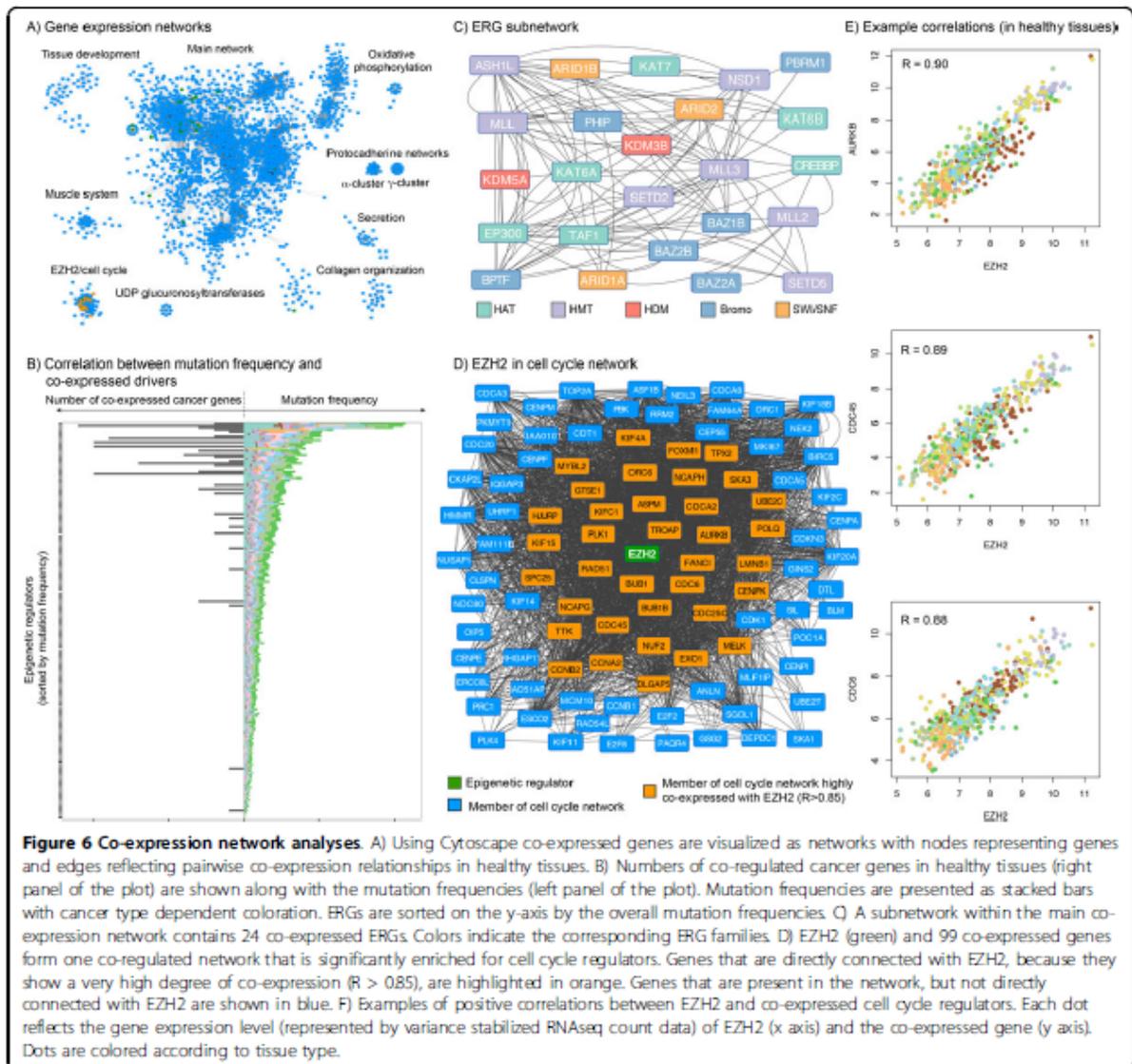


Figure 6 Co-expression network analyses. A) Using Cytoscape co-expressed genes are visualized as networks with nodes representing genes and edges reflecting pairwise co-expression relationships in healthy tissues. B) Numbers of co-regulated cancer genes in healthy tissues (right panel of the plot) are shown along with the mutation frequencies (left panel of the plot). Mutation frequencies are presented as stacked bars with cancer type dependent coloration. ERGs are sorted on the y-axis by the overall mutation frequencies. C) A subnetwork within the main co-expression network contains 24 co-expressed ERGs. Colors indicate the corresponding ERG families. D) EZH2 (green) and 99 co-expressed genes form one co-regulated network that is significantly enriched for cell cycle regulators. Genes that are directly connected with EZH2, because they show a very high degree of co-expression ($R > 0.85$), are highlighted in orange. Genes that are present in the network, but not directly connected with EZH2 are shown in blue. F) Examples of positive correlations between EZH2 and co-expressed cell cycle regulators. Each dot reflects the gene expression level (represented by variance stabilized RNAseq count data) of EZH2 (x axis) and the co-expressed gene (y axis). Dots are colored according to tissue type.

($p_{GO} = 2.7 \times 10^{-34}$, 1.8×10^{-25}). Another network included nine members of the UDP glucuronosyltransferase 1 family, significantly associated with various metabolic processes.

Strikingly, we identified a distinct network with 100 genes, almost exclusively associated with the regulation of the cell cycle ($p_{GO} = 4.4 \times 10^{-57}$) (Figure 6D). *EZH2* is the only epigenetic regulator in this module. In total 62 genes in the network were annotated cell cycle regulators with consistent up-regulation in tumors, including cell division cycle genes *CDC6*, *CDC45*, and *CDC25C*, cyclins *CCNA2*, *CCNB1* and *CCNB2*, genes encoding for aurora kinase B (*AURKB*) and its interaction partner *NUF2*, mitotic checkpoint protein kinase *TTK*, *RAD51*, checkpoint activator *FANCI*, *DLGAP5*, polo-like kinases (*PLK*) 1 and 4 along with interacting cyclin regulator *FOXM1*. Other essential cell cycle genes included *BUB1*, *BUB1B*, *CHEK2*, *CDK1*, and several members of the kinesin family.

Multiple members of the cell cycle network are known to regulate or physically interact with each other. For example, the expression of *EZH2* is known to be regulated by the co-expressed transcription factor *E2F2* [65]. The exact role of *EZH2* as the only ERG in the cell cycle network, however, is not clear.

Co-expression patterns in the main network

Analyzing the main network revealed multiple pairwise co-expressions between ERGs and cancer genes. While we took only a subset of genes of the CGC to train our predictors, we defined all genes from the CGC as cancer drivers in the co-expression analysis. Some examples for positive correlations between ERGs and drivers are illustrated in Figure 6E and Additional file 11. Interestingly, genes encoding longer proteins showed more co-expressed genes, presumably because they provide increased surfaces for interaction. Consequently, without normalizing for coding sequence length, frequently mutated ERGs correlated with more cancer genes than rarely mutated ERGs ($p = 2.3 \times 10^{-5}$ using permutation test) (Figure 6B).

Overall, we found seven cases, where the expression of one ERG was negatively correlated with the expression of another gene ($R < -0.85$) (Additional file 12). The transcription factor *BUD31* was involved in three of the seven instances including *ASH1L*, *KAT6A*, and *KDM3B*. Without known functional causalities, however, it is difficult to interpret these negative correlations.

Identifying sub-networks by investigating highly co-expressed gene pairs (directly linked nodes) within the major network revealed 24 inter-connected co-expressed ERGs (Figure 6C). This sub-network was composed of members from different ERG families. Similar to all observed co-expression patterns, this finding may not only imply common functionality, but also reflect that

the epigenetic machinery is partially controlling itself or is commonly controlled by another regulatory mechanism.

Proteomic analysis of the antiproliferative effect of EZH2 inhibition in mutant lymphoma cells

While the exact role of *EZH2* in the identified cell cycle network is not clear, *EZH2* is known as direct transcription repressor or activator of several cell cycle regulators (Additional file 13). As member of the Polycomb-group family, *EZH2* acts as transcription repressor of several cell cycle-related tumor suppressor genes such as *CDKN1C* through methylation of histone H3 on lysine 27 (H3K27) [66,67]. In an alternative model for *EZH2* mediated regulation, *EZH2* promotes tumorigenicity by direct activation of OGs such as *STAT3* [68].

Inhibition of *EZH2* has been suggested to induce cell cycle arrest in G1 phase and antiproliferative response in the mutant-bearing lymphoma cell line WSU-DLCL2 (*EZH2*^{Y641F}) [13]. The associated study further showed that proliferation of *EZH2* wildtype cells was not affected by the same treatment. After only 2 days of compound treatment cell cycle genes were found significantly down-regulated in the mutant cell line based on microarray experiments. Overall, we identified 11 out of the 30 most down-regulated cell cycle genes from this study in our co-expressed network (*CDC6*, *BUB1*, *CDC25C*, *BUB1B*, *TTK*, *CCNB1*, *CCNA2*, *PKMYT1*, *E2F2*, *CDC20*, *PLK1*).

To analyze the effect of *EZH2* inhibition on the proteome, we treated WSU-DLCL2 cells with the selective *EZH2* small molecule inhibitor EPZ-6438 (Epizyme[®], Cambridge, MA) [69,70] and measured global proteomic changes after 2, 4, 6 and 8 days using SILAC (stable isotope labeling by amino acids in cell culture) based mass spectrometry (Materials and methods).

Consistent with previous findings [13], we observed decreased viability of WSU-DLCL2 cells after *EZH2* inhibition. Concordant with *EZH2* as member of the PRC2 complex, which trimethylates histone 3 on lysine 27, the level of the H3K27me3 histone mark decreased by a factor of 2 and 3.3 after 2 and 5 days respectively (Additional file 14).

We identified 2530 proteins on average and quantified their intensity changes between EPZ-6438 treated cells (heavy labeled) and their respective non-treated control cells (light labeled) (Additional file 15). The combined proteome profiles over all time points comprised 3066 proteins. Overall 1852 proteins were commonly identified in all time point experiments. Clustering the associated time courses revealed three different profiles representing up-, down-, and non-regulated proteins (Additional file 16).

In total 267 and 202 proteins showed minimum 2-fold increase or decrease in expression levels respectively. Based on gene ontology (GO) analysis, the set of down-regulated proteins was significantly enriched for genes associated with cell cycle ($p = 2.25 \times 10^{-9}$) and DNA replication ($p = 6.97 \times 10^{-17}$) (Additional file 17). Among the 59 down-regulated cell cycle proteins were CDK2, CND1, MCM7, RFC2 and several regulators that were co-expressed in the EZH2 cell cycle network including CDK1, CND3, FANCI, BUB1, KIF11, TOP2A, TOPK, and UHRF1 (Figures 7A,B). Overall, 24 cell cycle associated proteins were up-regulated after EZH2 inhibition including tumor suppressors ATM, BRCA2 and cell cycle inhibitor CDN2C.

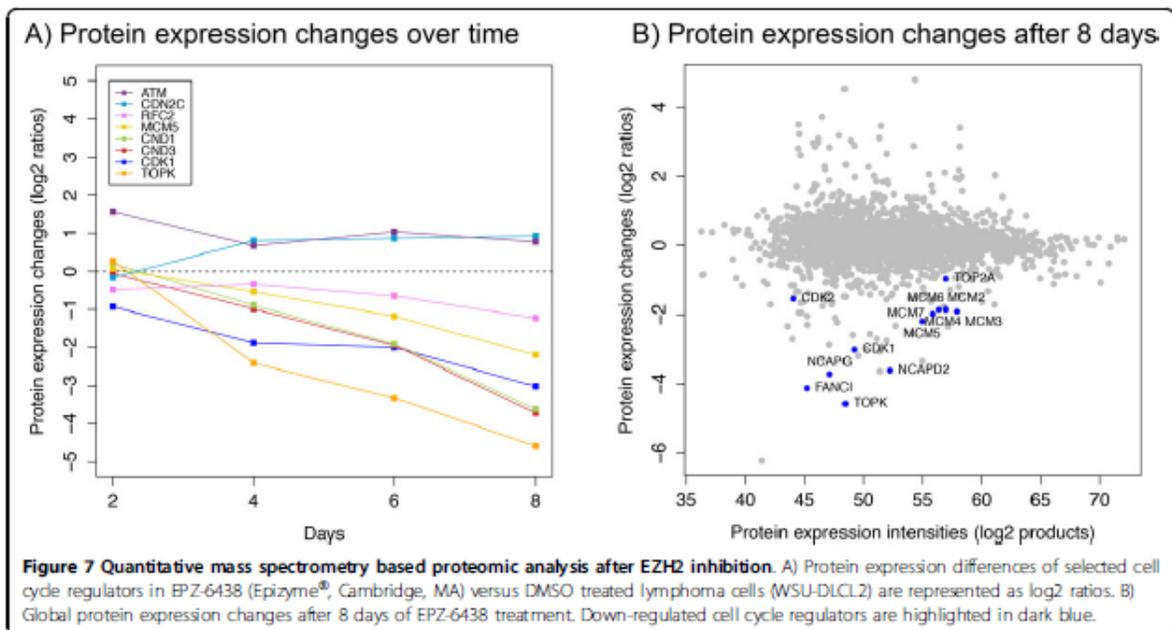
Our proteomics results do not distinguish whether EZH2 inhibition acts directly on the expression of cell cycle genes or more generally induces cell cycle arrest. However, coupled with the known regulatory roles of EZH2 as a member of the PRC2 complex and results from previous studies [13], these data suggest a regulatory function of EZH2 in controlling it's co-expressed cell cycle network.

Discussion

Alterations in the epigenetic machinery that lead to uncontrolled cellular proliferation have become an important research topic in the field of oncology. By training cancer gene predictors based on TCGA data, we found that classical cancer drivers are characterized by significant mutation or copy number patterns, but

not by altered expression. Among ERGs we identified multiple TSGs with significant proportions of loss of function mutations. Given the lack of recurrent mutation hot spots within the ERG families in the tumor panel, DNMT3A was the only ERG that showed an OG-like alteration profile. However, whether DNMT3A acts as OG or TSG has been debated, and additional studies are required to understand the exact role of DNMT3A in cancer. The classification of *DNMT3A* as OG driver can be attributed to the identification of a mutation hot spot on position 882 in acute myeloid leukemia. This shows that our predictor, which was trained on the combined set of all tumors, was capable to detect significant alterations within a single cancer type. It also makes clear, however, that the approach is biased towards included cancer types. Activating mutations within the catalytic SET domain of EZH2, for example, are known in non-Hodgkin's lymphoma [71], but were absent in our tumor cohort.

Many ERGs, which were not predicted as drivers, had dysregulated expression in cancer. The role of dysregulated genes in cancer is generally difficult to determine [72], but the discovered expression profiles among ERGs were remarkable. *EZH2* was the most significantly up-regulated gene. Strikingly, co-expression network analysis uncovered *EZH2* as the only ERG in a co-expressed cell cycle module. Selective inhibition of *EZH2* has been shown to decrease expression of multiple cell cycle regulators [13], many of which are in our co-expressed network. Despite the limitations of mass spectrometry to



identify a subset of the whole human proteome only, we confirmed the down-regulation of cell cycle proteins and showed a decrease of the PRC2-associated H3K27 methylation mark after EZH2 inhibition. We cannot determine, however, whether EZH2 inhibition acts directly on the expression of cell cycle genes or more generally induces cell cycle arrest. Interestingly, the EZH2 homolog EZH1 has been proposed to form PRC2 complexes with similar functions [63], EZH1 is commonly down-regulated in cancer, which contradicts its involvement as a cell cycle promoting PRC2 subunit.

Several other ERGs were significantly over-expressed in tumors. *BRDT*, *PRDM9* and *PRDM13*, for example, were exclusively expressed in tumors. *BRDT* showed characteristics of a CT gene, and was co-expressed with other known CT genes. The underlying mechanisms that induce the co-expression of these genes or the effect on the cancer cell are not known, but *BRDT* may present a potential candidate for initializing their expression as an epigenetic regulator.

While overexpression of *ATAD2* and *ACTL6A* are related to their genomic locations on large frequently amplified chromosome regions, *DNMT3B*, *KAT2A*, *SUV420H2* and several other ERGs showed ubiquitous significant up-regulation in cancer, therefore presenting an interesting candidate set for potential therapeutic targets.

Taken together, our prediction method identified several ERGs with mutation alteration profiles characteristic of classical TSGs. *DNMT3A* was the only predicted OG-like ERG with mutation hot spots in acute myeloid leukemia. Expression analysis further supports the role of EZH2 as an OG. Our study provides the first systematic analysis of the epigenetic regulators, thus providing basis for further prioritization of such players as candidates for therapeutic target discovery.

Additional material

Additional file 1: Epigenetic regulators of gene expression as writers, erasers and readers of covalent DNA and histone modifications. The upper panel provides an overview of writers (DNMTs, HATs, and HMTs), erasers (DNDMTs, HDACs, and HDMTs), and readers (bromo domain containing and methyl binding proteins) of epigenetic marks. Epigenetic regulators can be identified by the presence of specific associated domains, which are listed on the right of the lower panel. The sequence similarities between contained domains or total protein sequences formed the phylogenetic trees for each epigenetic gene family as shown on left.

Additional file 2: Description of genomic features.

Additional file 3: List of members of ERG families.

Additional file 4: Illustration of the SWI/SNF complex.

Additional file 5: Overview: Cancer gene prediction applied to ERGs.

Additional file 6: Genomic alterations of HATs, HDACs, HMTs, HDMs and members of the SWI/SNF complex. The compositions of the plots are explained in Figure 1.

Additional file 7: List of frequent mutations in TCGA.

Additional file 8: Differential gene expression analysis results.

Additional file 9: Co-expression of cancer testis genes. A) Volcano plot resulting from the differential expression analysis between *BRDT* expressing and *BRDT* non-expressing LUSC tumors. B) Gene expression levels of co-expressed cancer testis in LUSC (black: healthy tissues, red: tumors).

Additional file 10: Gene ontology enrichment analysis of identified co-expression networks. For some of the identified networks, gene ontology enrichment analyses were performed. "N" is the total number of annotated genes in the given network, while "x" is the number of annotated genes in the network that are associated with the given gene ontology accession (GO-ID). "n" is the number of annotated genes in the background set, while "n'" is the number of genes from the background set that are associated with the given gene ontology accession (GO-ID).

Additional file 11: Examples of co-expression between ERGs and other genes in healthy tissues. Each dot reflects the gene expression levels (represented by variance stabilized RNAseq count data) of the ERG (x axis) and the co-expressed gene (y axis). Dots are colored according to the associated tissue indication.

Additional file 12: Negative correlations between expression levels of ERGs and other genes. Analogously gene expression (variance stabilized RNAseq count data) of the epigenetic regulator (x axis) and the co-expressed gene (y axis). Colors indicate the associated tissue indication.

Additional file 13: Known models for EZH2 as cell cycle regulator. Two established models describe a cell cycle regulating role of EZH2: With its transcription repressing role as member of PRC2 complex (left panel), EZH2 enhances the expression of cell cycle regulators indirectly by repressing associated tumor suppressors such as CDKN1C. In an alternative model, EZH2 acts as a direct activator (right panel). Phosphorylated EZH2 activates STAT3 via methylation, which in turn activates the cyclin D1/CDK2 complex. Interestingly CDK1 and CDK2 have been shown to phosphorylate EZH2. In addition EZH2 has been shown to inhibit BRCA1 phosphorylation presumably via interaction with Akt-1 resulting into increase of cell cycle promoting CDC25C.

Additional file 14: Mass spectrometry based quantitation of H3K27me following EZH2 inhibition.

Additional file 15: Mass spectrometry results. List of identified proteins and corresponding quantitative results.

Additional file 16: Clustered time series. Using fuzzy c-means clustering, time course profiles formed three clusters of down-, up-, and non-regulated proteins. Colors reflect the similarities between specific time series and the associated cluster.

Additional file 17: Gene ontology enrichment analysis of regulated proteins.

Competing interests

All authors were employed by Genentech, Inc. during the time the study was done.

Authors' contributions

FG and ZZ were responsible for the study design. FG performed data analysis. SD and DA performed proteomic analysis. FG, GM and ZZ wrote the manuscript.

Acknowledgements

We thank Jens Reeder and Gregoire Pau for development of the transcriptome sequencing analysis pipeline. Special thanks go to Anneleen Daemen and Sasha Sharma for processing RNAseq data from TCGA. We thank Richard Bourgon, David Stokoe and Jeff Settleman for helpful discussions. Publication funds for this article came from Genentech. This article has been published as part of *BMC Genomics* Volume 16 Supplement 8, 2015: Vari-SIG 2014: Identification and annotation of genetic variants in the context of structure, function and disease. The full contents

of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/16/S8>.

Authors' details

¹Departments of Bioinformatics and Computational Biology, Genentech, USA.
²Protein Chemistry, Genentech Inc., South San Francisco, CA 94080, USA.
³Proteomics and Signal Transduction, Max-Planck-Institute for Biochemistry, D-82152 Martinsried, Germany. ⁴Biodynamic Optical Imaging Center, College of Life Sciences, Peking University, Beijing 100871, China.

Published: 18 June 2015

References

1. Bird A: DNA methylation patterns and epigenetic memory. *Genes & Development* 2002, **16**(1):6-21.
2. Kouzarides T: Chromatin modifications and their function. *Cell* 2007, **128**(4):693-705.
3. Strahl BD, Allis CD: The language of covalent histone modifications. *Nature* 2000, **403**(6765):41-45.
4. Wilson BG, Roberts CW: SWI/SNF nucleosome remodellers and cancer. *Nature Reviews Cancer* 2011, **11**(7):481-492.
5. Jones PA, Baylin SB: The epigenomics of cancer. *Cell* 2007, **128**(4):683-692.
6. Portela A, Esteller M: Epigenetic modifications and human disease. *Nature Biotechnology* 2010, **28**(10):1057-1068.
7. Rodriguez-Paredes M, Esteller M: Cancer epigenetics reaches mainstream oncology. *Nature Medicine* 2011, **17**(3):330-339.
8. Simo-Riuadibas L, Esteller M: Cancer genomics identifies disrupted epigenetic genes. *Human Genetics* 2013, **133**(6):713-725.
9. Kaminskas E, Farrell AT, Wang YC, Sridhara R, Pazdur R: FDA drug approval summary: azacitidine (5-azacytidine, Vidaza) for injectable suspension. *Oncologist* 2005, **10**(3):176-182.
10. Garcia JS, Jain N, Godley LA: An update on the safety and efficacy of decitabine in the treatment of myelodysplastic syndromes. *Oncotargets and Therapy* 2010, **3**:1-13.
11. O'Connor OA, Heaney ML, Schwartz L, Richardson S, Willim R, MacGregor-Cortelli B, et al: Clinical experience with intravenous and oral formulations of the novel histone deacetylase inhibitor suberoylanilide hydroxamic acid in patients with advanced hematologic malignancies. *J Clin Oncol* 2006, **24**(1):166-173.
12. Plekarcz RL, Frye R, Tumer M, Wright JJ, Allen SL, Kirschbaum MH, et al: Phase II multi-institutional trial of the histone deacetylase inhibitor romidepsin as monotherapy for patients with cutaneous T-cell lymphoma. *J Clin Oncol* 2009, **27**(32):5410-5417.
13. Krutson SK, Wagle TJ, Warholic NM, Sneeringer CJ, Allain CJ, Klaus CR, et al: A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nature Chemical Biology* 2012, **8**(11):890-896.
14. McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, et al: EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* 2012, **492**(7427):108-112.
15. Omberg L, Elliott K, Yuan Y, Kandath C, Wong C, Kellen MR, et al: Enabling transparent and collaborative computational analysis of 12 tumor types within The Cancer Genome Atlas. *Nature Genetics* 2013, **45**(10):1121-1126.
16. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garaway LA, Golub TR, et al: Discovery and saturation analysis of cancer genes across 21 tumor types. *Nature* 2014, **505**(7484):495-501.
17. Criollo G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C: Emerging landscape of oncogenic signatures across human cancers. *Nature Genetics* 2013, **45**(10):1127-1133.
18. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al: The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics* 2013, **45**(10):1113-1120.
19. Davoli T, Xu AW, Mengwasser KE, Sack LM, Yoon JC, Park PJ, Elledge SJ: Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* 2013, **155**(4):948-962.
20. UniProt Consortium: Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Research* 2013, **41**(Database issue):D43-D47.
21. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al: Clustal W and Clustal x version 2.0. *Bioinformatics* 2007, **23**(21):2947-2948.
22. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ: Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009, **25**(9):1189-1191.
23. Letunic I, Bork P: Interactive Tree Of Life (ITOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 2007, **23**(1):127-128.
24. Letunic I, Bork P: Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* 2011, **39**(Web Server issue):W475-W478.
25. Gao J, Aksoy BA, Dogrusoz U, Drieschner G, Gross B, Sumer SO, et al: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling* 2013, **6**(269):pl1.
26. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* 2012, **2**(5):401-404.
27. Cancer Genome Atlas Research Network: Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 2014, **507**(7492):315-322.
28. Cancer Genome Atlas Research Network: Comprehensive molecular portraits of human breast tumours. *Nature* 2012, **490**(7418):61-70.
29. Cancer Genome Atlas Research Network: Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012, **487**(7407):330-337.
30. Brennan CW, Verhaak RG, McKenna A, Campos B, Noursheh H, Salama SR, et al: The somatic genomic landscape of glioblastoma. *Cell* 2013, **155**(2):462-477.
31. Davis CF, Ricketts CJ, Wang M, Yang L, Cherniack AD, Shen H, et al: The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* 2014, **26**(3):319-330.
32. Cancer Genome Atlas Research Network: Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 2013, **499**(7456):43-49.
33. Cancer Genome Atlas Research Network: Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England Journal of Medicine* 2013, **368**(22):2059-2074.
34. Cancer Genome Atlas Research Network: Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014, **511**(7511):543-550.
35. Cancer Genome Atlas Research Network: Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012, **489**(7417):519-525.
36. Eberl HC, Mann M, Vermeulen M: Quantitative proteomics for epigenetics. *ChemBiochem* 2011, **12**(2):224-234.
37. Cancer Genome Atlas Research Network: Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 2014, **513**(7517):202-209.
38. Cancer Genome Atlas Research Network: Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 2014, **159**(3):676-690.
39. Kandath C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, et al: Integrated genomic characterization of endometrial carcinoma. *Nature* 2013, **497**(7447):67-73.
40. Riva B, Antipin Y, Sander C: Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research* 2011, **39**(17):e118.
41. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, et al: Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**(50):20007-20012.
42. RCT: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2014 (<http://www.R-project.org>).
43. Wu TD, Nacu S: Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 2010, **26**(7):873-881.
44. Liu J, Lee W, Jiang Z, Chen Z, JhunJhunwala S, Haverty PM, et al: Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. *Genome Research* 2012, **22**(12):2315-2327.
45. Anders S, Huber W: Differential expression analysis for sequence count data. *Genome Biology* 2010, **11**(10):R106.
46. Friedman J, Hastie T, Tibshirani R: Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 2010, **33**(1):1-22.

47. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: A census of human cancer genes. *Nature Reviews Cancer* 2004, **4**(3):177-183.
48. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008, **9**:559.
49. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003, **13**(11):2498-2504.
50. Anania VG, Bustos DJ, Lill JR, Kirkpatrick DS, Coscoy L: A Novel Peptide-Based SILAC Method to Identify the Posttranslational Modifications Provides Evidence for Unconventional Ubiquitination in the ER-Associated Degradation Pathway. *International Journal of Proteomics* 2013, **2013**:857918.
51. Kumar L, M EF: Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2007, **2**(1):5-7.
52. Maere S, Heymans K, Kuiper M: BINGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 2005, **21**(18):3448-3449.
53. Arrowsmith CH, Bounta C, Fish PV, Lee K, Schapira M: Epigenetic protein families: a new frontier for drug discovery. *Nature Reviews Drug Discovery* 2012, **11**(5):384-400.
54. Nehrt NL, Peterson TA, Paik D, Kann MG: Domain landscapes of somatic mutations in cancer. *BMC Genomics* 2012, **13**(Suppl 4):S9.
55. Gnad F, Baucom A, Mukhyala K, Manning G, Zhang Z: Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics* 2013, **14**(Suppl 3):S7.
56. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al: Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 2011, **476**(7360):298-303.
57. Fourret R, Laffaire J, Hofman P, Beau-Faller M, Mazieres J, Validire P, et al: A comparative and integrative approach identifies ATPase family, AAA domain containing 2 as a likely driver of cell proliferation in lung adenocarcinoma. *Clin Cancer Res* 2012, **18**(20):5606-5616.
58. Ciro M, Prosperini E, Quarto M, Grazini U, Walfordsson J, McBlane F: ATAD2 is a novel cofactor for MYC, overexpressed and amplified in aggressive tumors. *Cancer Research* 2009, **69**(21):8491-8498.
59. Shang E, Nickerson HD, Wen D, Wang X, Wolgemuth DJ: The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation. *Development* 2007, **134**(19):3507-3515.
60. Matzuk MM, McKeown MR, Filippakopoulos P, Li Q, Ma L, Agno JE, et al: Small-molecule inhibition of BRDT for male contraception. *Cell* 2012, **150**(4):673-684.
61. Caballero OL, Chen YT: Cancer/testis (CT) antigens: potential targets for immunotherapy. *Cancer Science* 2009, **100**(11):2014-2021.
62. Scanlan MJ, Altorki NK, Gure AO, Williamson B, Jungbluth A, Chen YT, Old LJ: Expression of cancer-testis antigens in lung cancer: definition of bromodomain testis-specific gene (BRDT) as a new CT gene, CT9. *Cancer Letters* 2000, **150**(2):155-164.
63. Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD, Reinberg D: Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Molecular Cell* 2008, **32**(4):503-518.
64. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N: Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genetics* 2003, **34**(2):166-176.
65. Bracken AP, Pasini D, Capra M, Prosperini E, Cilli E, Helin K: EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* 2003, **22**(20):5323-5335.
66. Yoo KH, Hennighausen L: EZH2 methyltransferase and H3K27 methylation in breast cancer. *International Journal of Biological Sciences* 2012, **8**(1):59-65.
67. Velichutina I, Shakhovich R, Geng H, Johnson NA, Gascoyne RD, Melnick AM, Elemento O: EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. *Blood* 2010, **116**(24):5247-5255.
68. Kim E, Kim M, Woo DH, Shin Y, Shin J, Chang N, et al: Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. *Cancer Cell* 2013, **23**(6):839-852.
69. Knutson SK, Warholic NM, Wigle TJ, Klaus CR, Allain CJ, Raimondi A, et al: Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**(19):7922-7927.
70. Knutson SK, Kawano S, Minoshima Y, Warholic NM, Huang KC, Xiao Y, et al: Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. *Molecular Cancer Therapeutics* 2014, **13**(4):842-854.
71. McCabe MT, Graves AP, Ganji G, Diaz E, Halsey WS, Jiang Y, et al: Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypermethylation of histone H3 on lysine 27 (H3K27). *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(8):2989-2994.
72. Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS: A census of amplified and overexpressed human cancer genes. *Nature Reviews Cancer* 2010, **10**(1):59-64.

doi:10.1186/1471-2164-16-S8-S5

Cite this article as: Gnad et al.: Bioinformatics analysis of thousands of TCGA tumors to determine the involvement of epigenetic regulators in human cancer. *BMC Genomics* 2015 **16**(Suppl 8):S5.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Article 3: Quantitative proteomics reveals fundamental regulatory differences in oncogenic HRAS and Isocitrate Dehydrogenase (IDH1) driven astrocytoma²⁰⁹

Molecular and Cellular Proteomics

Sophia Doll[‡], Anatoly Urisman[‡], Juan A. Oses-Prieto[‡], David Arnott[§], and Alma L. Burlingame[‡]

[‡]Department of Pharmaceutical Chemistry, University of California, San Francisco, 94158–2517 California

[§]Department of Protein Chemistry, Genentech Inc, South San Francisco, 94158 –2517 California

Glioblastoma are the most frequent brain tumors and can be classified into primary and secondary glioblastomas. Current therapies combine surgery, radiotherapy, chemotherapy, and temozolomide treatment, but less than 5% of the patients survive longer than 5 years after diagnosis. Genomic analyses showed that more than 70% of secondary glioblastomas harbor IDH1 mutation. The downstream effects of IDH1 mutation on the proteome, phosphoproteome, and epigenome in glioma, however, remain poorly understood. In this publication, we applied a SILAC labeling methodology and high-resolution mass spectrometry to identify the main differences between oncogenic HRAS driven cells and mutant IDH1-driven glioma cells, mimicking primary and secondary glioblastoma. We found that primary glioblastoma cell lines are mainly characterized by the overexpression of the MEK and PI3K signaling pathways. Dual inhibition of MEK and PI3K was clearly superior to single MEK inhibition in the cell line model; however, clinical efficacy remains to be determined. In contrast, secondary glioblastoma cell lines showed epigenetic reprogramming at the histone code level. Applying a histone hybrid chemical labeling method and high-resolution MS, we identified significant histone methylation, acetylation, and butyrylation changes. Our results suggest a global transcriptional repressive state, consistent with the down-regulation of the proteome,

transcriptome, whereas DNA is hyper-methylated and there is an increase in histone tri-methylation marks. This work provides a unique resource of the identity of altered proteins, phosphosites, and histone PTMs in RAS and IDH1 mutant astrocytoma cell lines, revealing new insight into oncogenesis in glioma. I also illustrated the January 2017 cover of Molecular and Cellular Proteomics (MCP).

Quantitative Proteomics Reveals Fundamental Regulatory Differences in Oncogenic HRAS and Isocitrate Dehydrogenase (IDH1) Driven Astrocytoma^{*§}

Sophia Doll[‡], Anatoly Urisman[‡], Juan A. Oses-Prieto[‡], David Arnott[§], and Alma L. Burlingame^{‡¶}

Glioblastoma multiformes (GBMs) are high-grade astrocytomas and the most common brain malignancies. Primary GBMs are often associated with disturbed RAS signaling, and expression of oncogenic HRAS results in a malignant phenotype in glioma cell lines. Secondary GBMs arise from lower-grade astrocytomas, have slower progression than primary tumors, and contain IDH1 mutations in over 70% of cases. Despite significant amount of accumulating genomic and transcriptomic data, the fundamental mechanistic differences of gliomagenesis in these two types of high-grade astrocytoma remain poorly understood. Only a few studies have attempted to investigate the proteome, phosphorylation signaling, and epigenetic regulation in astrocytoma. In the present study, we applied quantitative phosphoproteomics to identify the main signaling differences between oncogenic HRAS and mutant IDH1-driven glioma cells as models of primary and secondary GBM, respectively. Our analysis confirms the driving roles of the MAPK and PI3K/mTOR signaling pathways in HRAS driven cells and additionally uncovers dysregulation of other signaling pathways. Although a subset of the signaling changes mediated by HRAS could be reversed by a MEK inhibitor, dual inhibition of MEK and PI3K resulted in more complete reversal of the phosphorylation patterns produced by HRAS expression. In contrast, cells expressing mutant IDH1 did not show significant activation of MAPK or PI3K/mTOR pathways. Instead, global downregulation of protein expression was observed. Targeted proteomic analysis of histone modifications identified significant histone methylation, acetylation, and butyrylation changes in the mutant IDH1 ex-

pressing cells, consistent with a global transcriptional repressive state. Our findings offer novel mechanistic insight linking mutant IDH1 associated inhibition of histone demethylases with specific histone modification changes to produce global transcriptional repression in secondary glioblastoma. Our proteomic datasets are available for download and provide a comprehensive catalogue of alterations in protein abundance, phosphorylation, and histone modifications in oncogenic HRAS and IDH1 driven astrocytoma cells beyond the transcriptomic level. *Molecular & Cellular Proteomics* 16: 10.1074/mcp.M116.063883, 39–56, 2017.

Gliomas are the most common human brain tumors with ~250,000 cases per year worldwide (1). Gliomas arise from glial cells, which are non-neuronal cells that provide support and protection for neurons. Astrocytomas are the most common form of glioma (2) and are histologically categorized into four grades (I–IV), of which grade III (anaplastic astrocytoma), and grade IV (glioblastoma multiforme (GBM)¹), are malignant (3). GBMs are among the deadliest human cancers, and despite the use of aggressive multimodality therapy combining surgery, radiotherapy, and chemotherapy, less than 5% of patients survive longer than 5 years after diagnosis (4). Although most GBMs (90%) develop *de novo* (primary glioblastoma) and typically have rapid progression, some (10%) progress more slowly after initially presenting as low-grade gliomas (secondary glioblastoma) (supplemental Fig. S1) (5).

Most common alterations in primary glioblastomas include telomerase reactivation, p53 and pRB pathway deactivation, PTEN loss, and EGFR amplification leading to RAS signaling activation (6). RAS network alterations (other than by mutation, which is rare in gliomas) are commonly observed in malignant astrocytomas (7–10). Importantly, overexpression of HRAS in normal human astrocytes (NHAs) results in the formation of intracranial tumors strongly resembling human

From the [‡]Department of Pharmaceutical Chemistry, University of California, San Francisco, 94158–2517 California; [§]Department of Protein Chemistry, Genentech Inc, South San Francisco, 94158–2517 California

Received September 12, 2016, and in revised form, November 4, 2016

Published, MCP Papers in Press, November 10, 2016, DOI 10.1074/mcp.M116.063883

Author contributions: S.D. designed research; S.D. performed research; S.D. analyzed data; S.D. wrote the paper; A.U. helped with the biological interpretation; J.A.O. mass spectrometry support; D.A. supervised the histone ptm analysis; A.L.B. supervised the project.

¹ The abbreviations used are: GBM, glioblastoma multiforme; NHA, normal human astrocytes; PTM, post-translational modification; hTERT, human telomerase reverse transcriptase.

Signaling and Epigenetics in Astrocytoma

grade III anaplastic astrocytoma after injection into mice. In contrast, activation of AKT or EGFR does not trigger this transformation (7). RAS remains difficult to target and several small molecule inhibitors targeting downstream nodes of the RAS regulated core axis have been developed for potential cancer treatment. For example, cobimetinib (Genentech Inc., South San Francisco, CA) and pictilisib (Genentech Inc.) are both highly selective and potent inhibitors of MEK1/2 and PI3K, respectively (11, 12). MEK or PI3K inhibitors as single agents in GBM therapy, however, lack of efficacy and require the development of more effective therapies (13, 14).

Over 70% of secondary glioblastomas harbor isocitrate dehydrogenase 1 (IDH1) mutations, involving Arg132 in nearly all cases, whereas primary glioblastoma rarely show IDH mutations (15–17). It has been shown that IDH1 Arg132 mutation suppresses the biochemical ability of IDH1 to convert isocitrate into α -Ketoglutarate (α -KG) by further converting α -KG into 2-hydroxyglutarate (2-HG) (18). As a result, the oncometabolite 2-HG accumulates at high levels in IDH1 mutant tumors (up to 100-fold of normal) and inhibits α -KG-dependent histone and DNA demethylases, affecting epigenetic regulation and associated gene expression (supplemental Fig. S2) (19, 20). Although 60 other human α -KG-dependent dioxygenases exist (21), α -KG-dependent histone demethylases are the most sensitive to 2-HG-mediated inhibition (20). Other epigenetic alterations induced by mutant IDH1 include the CpG island methylator phenotype (CIMP), characterized by increased global DNA methylation (22). However, more detailed mechanisms of tumorigenesis produced by accumulation of 2-HG and CIMP in IDH1-mutant gliomas, including specific alterations in signaling pathways and key epigenetic factors responsible for transformation and progression, remain largely unknown.

To uncover possible mechanisms driving the cellular and molecular transformation of primary and secondary glioblastoma, we investigated changes of protein abundance, phosphorylation, and histone post-translational modifications (PTMs) in an *in vitro* model. Direct analysis of proteins and PTMs provides an important advantage over genomic and transcriptomic approaches, because mRNA expression often does not correlate with protein expression and cannot predict functional state of proteins determined by PTMs (23, 24). Only a few studies have attempted to investigate the alterations in phosphorylation within signaling networks in human gliomas (25–27), but none have addressed the potential involvement of an altered histone code by targeted quantitative mass spectrometry (MS). MS-based proteomics provides a platform for in-depth identification and quantification of thousands of proteins and their PTMs, including histone PTMs, and allows the quantitation of perturbed signaling networks (28).

To study these mechanisms *in vitro*, we chose a previously developed cellular model (7, 29). This system consists of NHAs, which are immortalized by the reactivation of telomerase activity via the expression of human telomerase reverse

transcriptase (hTERT) and the introduction of virally-encoded E6 and E7 to inhibit the transcription of p53 and pRb, respectively. NHA E6/E7 hTERT (control-NHA) are further transformed into tumorigenic astrocytoma cells by either overexpressing oncogenic H-Ras V12 (RAS-NHA) or introducing mutant IDH1 (IDH1mut-NHA), mimicking primary and secondary high-grade astrocytoma, respectively.

In this work we applied a quantitative MS-based strategy to characterize the proteomic and phosphoproteomic changes in HRAS and mutant IDH1 driven glioma cells. We report the driving roles of the MEK and PI3K signaling pathways in RAS-NHA cells, and describe previously unknown alterations in other pathways. Furthermore, we show that simultaneous MEK and PI3K inhibition reverses many but not all signaling changes driven by oncogenic HRAS in RAS-NHA cells. Additionally, we provide a quantitative view of major effects on histone PTM occupancies resulting from the overproduction of the oncometabolite 2-HG and its inhibition of chromatin remodeling enzymes in IDH1mut-NHA cells.

EXPERIMENTAL PROCEDURES

Cell Culture—Immortalized NHAs were a gift from R. O. Pieper (University of California San Francisco, CA). The creation of NHA expressing E6/E7 and hTERT (control-NHA), NHA expressing E6/E7, hTERT, and IDH1 mutant (IDH1mut-NHA), and NHA expressing E6/E7, hTERT, and H-Ras V12 (RAS-NHA) has been described previously (7, 30). IDH1mut- and RAS-NHA cells were cultivated in stable isotope labeling by amino acids in cell culture (SILAC) DMEM-H21 medium (Thermo Fisher Scientific, Waltham, MA) supplemented with 10% dialyzed FBS and 1% Penicillin Streptomycin containing heavy $^{13}\text{C}_6$ $^{15}\text{N}_2$ -lysine and $^{13}\text{C}_6$ $^{15}\text{N}_4$ -arginine (R10K8) or medium $^2\text{H}_4$ -lysine and $^{13}\text{C}_6$ arginine (R6K4) (Cambridge Isotope Laboratories, Andover, MA), respectively. Control-NHA cells were grown in SILAC DMEM-H21 medium supplemented with light lysine and arginine (R0K0). All cell lines were maintained at 37 °C and 5% CO_2 and collected after a minimum of 5 passages.

Immunoblotting—To prepare the lysates for immunoblot assays, cells were lysed in RIPA buffer (Thermo Fisher Scientific) supplemented with 1% phosphatase and protease inhibitors. Samples were rocked at 4 °C for 30 min and cleared by centrifugation at 10,000 rpm for 10 min at 4 °C. Protein levels were measured using the Pierce BCA protein assay (Thermo Fisher Scientific) by fluorescence spectrometry. Equal amount of protein extracts were incubated at 70 °C in LDS sample buffer and reducing agent (Invitrogen, Carlsbad, CA) for 10 min before being separated by SDS-polyacrylamide gel electrophoresis (NuPAGE, Invitrogen) and transferred to a nitrocellulose membrane. After blocking for 1 h at room temperature with bovin serum albumin, membranes were incubated with primary antibodies overnight at 4 °C. Anti-nestin (NES) antibody was obtained from Santa Cruz Biotechnology. Specific antigen-antibody interaction was detected with anti-mouse secondary antibodies labeled with horseradish peroxidase (HRP). Signal was revealed by SuperSignal West Pico Chemiluminescent HRP substrate (Thermo Fisher Scientific).

MEK and PI3K Inhibition—We treated heavy SILAC-labeled RAS-NHA cells with 2 μM of cobimetinib (GDC-0973, Genentech Inc.) and unlabeled RAS-NHA cells with DMSO for 4 h. Dual MEK and PI3K inhibition was carried out by the addition of 2 μM GDC-0973 and 5 μM pictilisib (GDC-0941, Genentech Inc.). After 4 h, cells were collected and stored at -80 °C. GDC-0973 and GDC-0941 drug concentrations were selected based on previously described conditions (31).

Sample Preparation and Titanium Dioxide (TiO₂) Phosphopeptide Enrichment—After full incorporation of SILAC labels, cell pellets were collected and stored at -80°C . Cell pellets were lysed in 8 M urea, 20 mM HEPES buffer by sonication and clarified by centrifugation at $16000 \times g$ for 10 min. Protein content was measured using Pierce BCA protein assay (Thermo Fisher Scientific) according to the manufacturer's protocol. Heavy and medium SILAC-labeled samples were combined with the unlabeled sample at 1:1:1 final protein content ratio. Proteins were reduced with dithiothreitol and alkylated with iodoacetamide prior to overnight tryptic in-solution digestion at 37°C . Tryptic peptides were further desalted with Sep-Pak C18 cartridges (Waters Corporation, Milford, MA). Peptides were enriched for phosphopeptides with an in-house packed TiO₂ column using an AKTA Purifier HPLC system (GE Healthcare, Chicago, IL) as described previously (32). Both, phosphopeptide and flowthrough (non-phosphorylated peptides) fractions were further fractionated by off-line high-pH reversed phase chromatography into 20 fractions as previously described (32). The fractions were desalted with Zip Tip C18 pipette tips (Millipore Corporation, Billerica, MA) before liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. This experiment was repeated in three biological replicates.

To assess the SILAC labeling efficiency, about 200 μg of proteins of heavy RAS-NHA and medium IDH1mut-NHA SILAC-labeled samples were separately reduced, alkylated, and digested before LC-MS/MS analysis as described above.

LC-MS/MS Analysis—Nanoflow LC-MS/MS analysis of tryptic peptides was conducted on an LTQ-Orbitrap Velos (Thermo Fisher Scientific) fitted with an EASY-Spray PepMap® RSLC C18, 3 μm , 100 \AA , 75 $\mu\text{m} \times 15$ cm column (Thermo Fisher Scientific). About 0.4 μg of desalted peptides were loaded and eluted over the course of 57 min from 2–27% solvent B (100% acetonitrile, 0.1% formic acid) and stepped up to 50% in 2 min. The mass spectrometer was operated in “top-6” data-dependent mode, collecting MS spectra in the Orbitrap mass analyzer (60000 resolution, 350–1500 m/z range) with an automatic gain control (AGC) target of 2E6 and a maximum ion injection time of 250 ms. Following higher-energy collisional dissociation (HCD), MS/MS spectra were collected in the Orbitrap (7500 resolution, 350–1500 m/z range) with an AGC target of 9E4 and a maximum ion injection time of 500 ms. Label free analysis was performed on an Orbitrap Fusion mass spectrometer in a “top-10” data-dependent mode, collecting MS spectra in the Orbitrap (120,000 resolution, 375–1600 m/z scan range) with an AGC target of 2E5 and a maximum ion injection time of 50 ms. After HCD fragmentation, MS/MS spectra were collected in the Orbitrap (30,000 resolution, 350–1400 m/z scan range) with an AGC target of 5E4 and a maximum injection time of 60 ms. Desalted peptides were loaded on an EASY-Spray PepMap® RSLC C18, 2 μm , 100 \AA , 75 $\mu\text{m} \times 15$ cm column (Thermo Fisher Scientific) and eluted over the course of 85 min with an acetonitrile gradient from 2–25% solvent B (98% acetonitrile, 0.1% formic acid) and stepped up to 40% in 2 min.

Experimental Design and Statistical Rationale—For the characterization of the proteomic and phosphoproteomic changes in HRAS and mutant IDH1 driven glioma cells, we applied a SILAC quantitative MS-based strategy. Biological triplicates of each triple-SILAC experiment provided p values for the statistical analyses. An additional label free quantification was measured to validate the proteomic changes observed in IDH1mut-NHA cells compared with control cells. Two technical replicates were measured for the analysis of histone PTM changes of IDH1mut- IDH1wt-, and control-NHA cells.

Data Analysis—Tandem mass spectra were searched against the UniProt_2015_07 human database (containing 68561 entries) using MaxQuant version 1.5.2.8 with a 1% false discovery rate (FDR) at the phosphosite, peptide, and protein level. Default settings with the following changes were applied: “phospho (STY)” were selected as

variable modifications for phosphopeptide searches, “re-quantify” and “match between runs” were enabled, and a minimum ratio count of 1 was selected, meaning that at least one labeled peptide triplet needs to be quantified to report a SILAC ratio. Carbamidomethylation of cysteine was selected as fixed modification and N-terminal protein acetylation and methionine oxidation as variable modifications and a maximum of two missed cleavages was selected with trypsin as protease. The mass tolerance for precursor ions and fragment ions were 20 ppm. Bioinformatics analysis was performed with Perseus version 1.5.1.6, Microsoft Excel, and R statistical software. Heavy/light and medium/light log₂ ratios were calculated for each detected protein and normalized by median-centering (i.e. setting the median log₂ ratio equal to zero). Average log₂ ratios from the biological repeats and the corresponding p values were visualized with volcano plots. p values were calculated based on a t test and a p value of < 0.05 and fold change of > 2 ($\log_2 > 1$) were chosen as the significance cutoff based on the distributions observed in volcano plots. All raw data, search parameters, and results are available on ProteomeXchange via the PRIDE database (PXD004945). We provide further information about all identified phosphopeptides in supplemental Tables S1–S3 and the annotated spectra can be visualized using the ProteinProspector MS-viewer with the following search keys: rqb1zvxx60 (triple SILAC experiment, supplemental Table S1), t3344bbxng (MEK inhibition experiment, supplemental Table S2), and w8fxjv0hoj (dual MEK and PI3K inhibition experiment, supplemental Table S3).

Gene Ontology, Molecular Signature Database (MSigDB) and Principal Component Analysis (PCA) Analysis—To identify enriched biological processes terms within our data set we used DAVID bioinformatics functional annotation tool (<http://david.abcc.ncicfcr.gov/>) (33). The significance of fold enrichment was calculated using a Benjamini-Hochberg adjusted p value of ≤ 0.01 . Gene set enrichment analysis (GSEA) was performed using gene set collections “h,” “c2,” “c3,” “c4,” “c5,” and “c6” from the MSigDB (34). We estimated the significance of fold enrichments for each gene set by hypergeometric testing in R. PCA was performed on the processed data as described previously (35).

Histone Purification and Mass Spectrometry Analysis by Parallel Reaction Monitoring (PRM)—H2A, H2B, H3, and H4 histones were purified with a histone purification kit (Active Motif, Carlsbad, CA) according to the manufacturer's protocol. Histone concentrations were measured using Direct Detect Infrared Spectrometer (Millipore). In total, 4 μg of unlabeled histones (extracted from IDH1mut-, IDH1wt-, and RAS-NHA cells) and heavy amino acid-labeled histones (extracted from PC9 cells, used as a common standard) were mixed in a 1:1 ratio. Histone digestion and MS analysis were performed as previously described (36). Briefly, histones were propionylated using deuterated propionic anhydride; the reaction was quenched by hydroxylamine followed by tryptic digestion overnight and phenyl-isocyanate (PIC) labeling. Histone peptides were then analyzed by capillary reversed phase ultra high-pressure liquid chromatography-electrospray ionization tandem MS on an Orbitrap Fusion mass spectrometer as described previously (37). Targeted product ion spectra were acquired by PRM based on a targeted inclusion list containing 72 transitions covering 100 peptide combinations and 71 distinct histone PTMs for histones H3 and H4. Data were further processed with Skyline (38) and the relative abundance of histone PTMs across the samples was normalized using the heavy SILAC-labeled internal standard (PC9 cells). Experiments were repeated twice (technical replicates) and a log₂ ratio cutoff of ≥ 0.2 was used to define significant changes.

RESULTS AND DISCUSSION

1) Quantitation of Proteomic and Phosphoproteomic Changes by SILAC-MS in RAS or Mutant IDH1 Driven Glioma Cells—To compare signaling alterations in RAS or mutant IDH1 driven

Signaling and Epigenetics in Astrocytoma

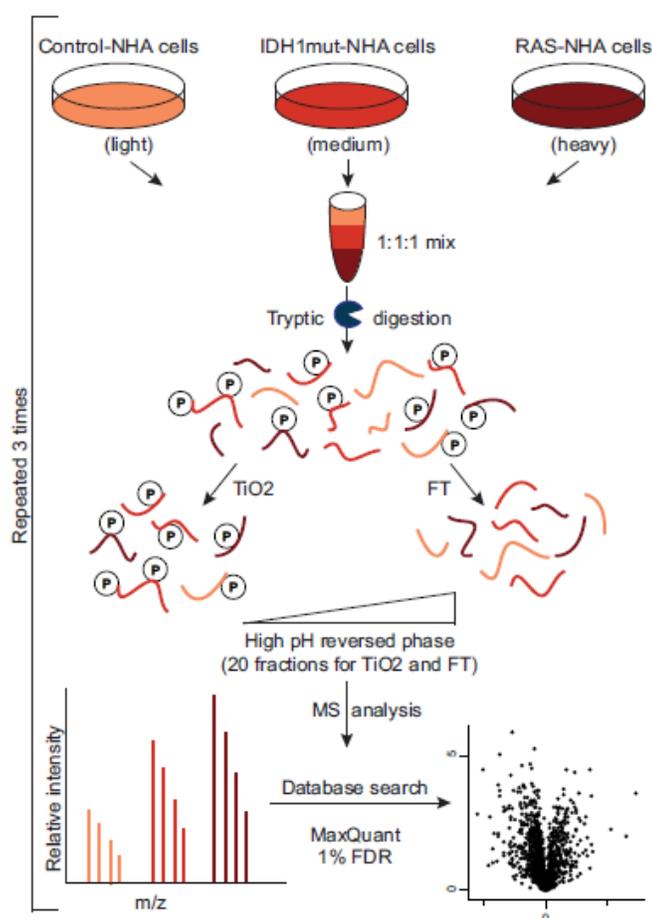


Fig. 1. Experimental design. IDH1mut-NHA and RAS-NHA cells were medium and heavy SILAC-labeled, respectively. Control-NHA cells remained unlabeled. Protein lysates were mixed in a 1:1:1 ratio, enzymatically digested, enriched for phosphopeptides by TiO₂, and both phosphopeptides and the flow through (FT) were fractionated by high-pH reversed phase chromatography into 20 fractions before analysis by LC-MS/MS. Data were processed using MaxQuant software at 1% FDR.

astrocytoma cell lines we determined proteome-wide phosphorylation and protein abundance changes in immortalized NHA cells in response to overexpression of HRAS or mutant IDH1 using SILAC based MS (39) (Fig. 1A). After full incorporation of SILAC labels as assessed by MS (98 and 96% labeling efficiency for RAS-NHA and IDH1mut-NHA cells, respectively (supplemental Table S4)), the phosphorylation and protein abundance changes of these two cell lines were compared with control-NHA cells. Overall, we identified 6942 phosphorylation sites with a 1% FDR at the protein, peptide, and phosphorylation site level. Among these, 4976 phosphorylation sites were mapped with high confidence (localization probability ≥ 0.75 (26)) (supplemental Table S5). A total of 2817 of these phosphosites were detected and quantified in at least two of the three biological replicates, and 1402 sites

were detected in all three replicates and all three cell types (Fig. 2A). The majority (77%) of the detected phosphopeptides were singly phosphorylated, whereas 19% were doubly and 4% triply phosphorylated. A comparison of relative phosphopeptide abundances (ratios to the control-NHA cells) between biological replicates showed high reproducibility, with an average Pearson correlation factor R of 0.77 (supplemental Figs. S3A and S4).

We detected a total of 4034 proteins at 1% FDR, of which 2800 proteins were identified in at least two biological replicates, and 1764 proteins were detected in all three biological replicates (Fig. 2B). Measured relative protein abundances in the biological replicates showed high degree of reproducibility with an average Pearson correlation factor R of 0.76 (supplemental Fig. S3B). MS validated the over-expression of HRAS

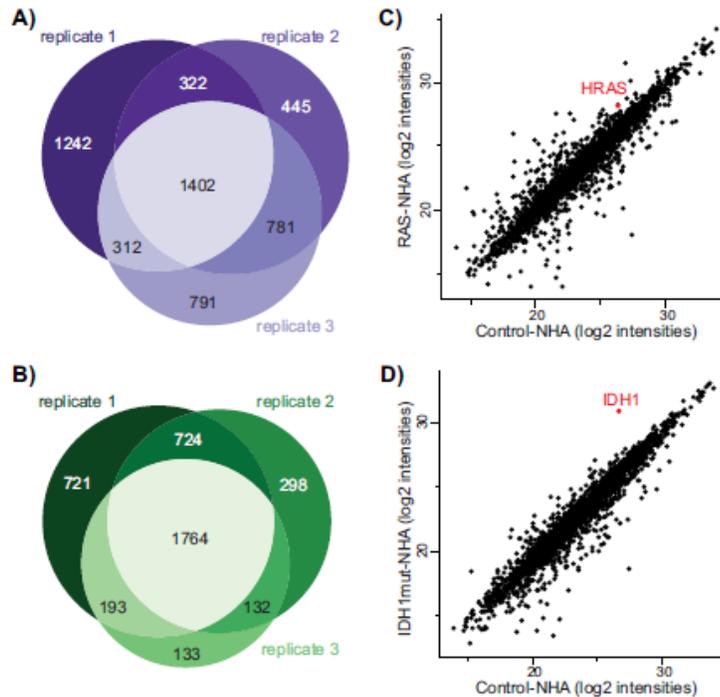


FIG. 2. **Detected proteomes and phosphoproteomes in three biological replicates.** The Venn diagrams indicate the number of (A) phosphosites (localization probability ≥ 0.75) and (B) proteins identified in each biological replicate. Scatterplots illustrate the comparison of protein intensities (log₂) in (C) RAS-NHA cells compared with control-NHA cells and (D) IDH1mut-NHA cells compared with control-NHA cells in one of three biological replicates.

at the protein level in RAS-NHA cells (Fig. 2C). Although previous studies confirmed the presence of the mutant form of IDH1 in IDH1mut-NHA by Western blot (29), we also observed significant over-expression of IDH1 at the protein level (Fig. 2D).

II) Characterization of the Phosphoproteome in RAS-NHA Cells—In order to detect signaling network alterations mediated by oncogenic HRAS in glioma, we quantitatively compared phosphorylation changes of RAS-NHA to those of NHA cells as the control. For all comparisons, we used a 2-fold change ($\log_2 \geq 1$) and p value of <0.05 cutoff to define significant up or downregulation. Overall, 18.6% of all identified phosphosites in RAS-NHA cells displayed significant changes in regulation (Fig. 3D). We found that 278 phosphorylation sites corresponding to 154 proteins showed upregulation, whereas 245 phosphorylation sites, mapping to 160 proteins, displayed downregulation (supplemental Table S6). In total, 672 phosphorylated proteins were also detected in the protein abundance measurements, allowing normalization of phosphorylation changes by the respective protein expression to identify instances of abundance-independent changes in phosphorylation (supplemental Table S7).

A) Affected Pathways Downstream of RAS—

Overexpression of HRAS Triggers Activation of the MAPK and PI3K Pathways—To identify the biological processes and pathways that are associated with altered phosphorylation after HRAS overexpression, we performed GSEA using MSigDB database of annotated gene sets (34). We found that upregulated phosphoproteins in RAS-NHA cells were significantly enriched for the MAPK ($p = 0.0005$), PI3K ($p = 0.009$), and MTOR ($p = 2e^{-9}$) pathways (Fig. 3A). Furthermore, sequences surrounding the regulated phosphosites showed a significant enrichment of canonical MAPK and PI3K substrate recognition motifs (Fig. 3B).

MAPK Pathway—Examining the correlation between phosphorylation and protein abundance revealed that a subset of upregulated phosphosites, including MAP2K2, MAPK1, and MAPK3 showed abundance independent upregulation (Fig. 3C). Phosphosites of MAPK1 (pT185 and pY187) are known MAP2K1/2 target sites (40) and displayed 3.4- and 6.5-fold abundance independent upregulation in our experiments, respectively. The dual phosphorylation of these two sites is known to promote the kinase activity of MAPK1 and is required for maximal activation (41). Similarly, phosphorylation of MAPK3 on pY204, which is another target site of MAP2K1/2 (42), revealed 6-fold higher abundance-independent

Signaling and Epigenetics in Astrocytoma

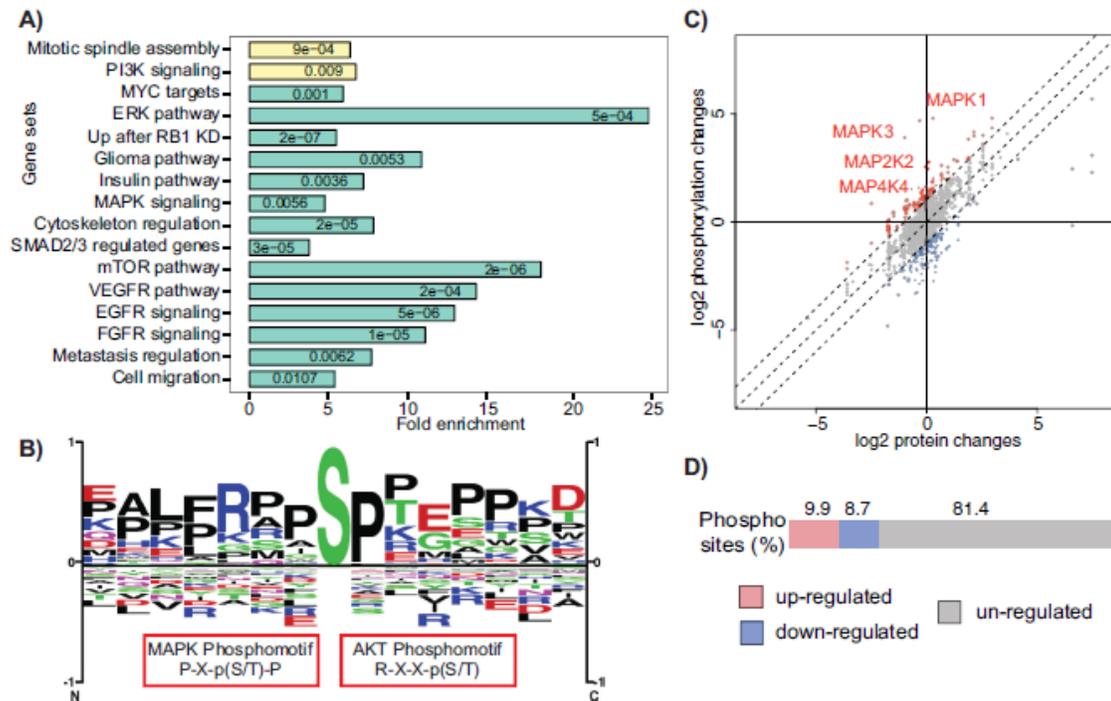


FIG. 3. MAPK and PI3K signaling pathways are significantly upregulated in RAS-NHA cells. Upregulated phosphosites are significantly enriched for (A) MAPK and PI3K pathways members based on GSEA analysis (p values are indicated for each gene set) and (B) for MAPK and PI3K phosphorylation motifs. C, A subset of at least 2-fold regulated phosphosites show abundance-independent changes in all 3 biological replicates. D, Proportion of regulated phosphosites.

ent upregulation (Fig. 4A and 4B). In addition to perturbed phosphorylation we found multiple changes at the protein abundance level that are consistent with activation of the MAPK pathway. For example, Sprouty protein 4 (SPRY4), whose expression is induced by the MAPK pathway (43, 44), showed a 10-fold upregulation at the protein level.

PI3K Members—We also found that several key members of the PI3K pathway showed regulation at the phosphorylation level, including AKT1S1, RPTOR, and RICTOR (Fig. 4A and 4B). For example, AKT1S1 phosphosites pT246 and pS212 displayed 3.1-fold and 1.9-fold abundance-independent upregulation, respectively. Both phosphosites have been described as activating and are target sites of AKT1 and MTOR, respectively (45, 46). Phosphorylation of AKT1S1 on residue pT246 promotes activation of the mTORC1. Similarly, phosphorylation of RPTOR on pS863 displayed a 2.6-fold upregulation; this site is also known to indicate mTORC1 activation (47). Activation of mTORC1 leads to the stimulation of its downstream targets including RPS6KB1 and subsequent activation of EIF4B, which triggers transcription of multiple proteins involved in cell proliferation and survival (48). Phosphorylation level of EIF4B (pS93) was identified in one of the three biological replicates and showed 2.2-fold abun-

dance independent upregulation compared with control cells. In contrast, other phosphosites on EIF4B, including pS238 and pS406, did not show significant changes. Interestingly, another member of the EIF family, EIF4E, showed a 2.5-fold or higher increase of phosphorylation on multiple sites in the C-terminal tail, including pT205, pS207, and pS209. The phosphorylation of EIF4E on pS209, known to be mediated by the EIF4G-associated kinases MNK1 and MNK2, promotes cap-dependent translation (49, 50). MNK1/2 are protein-serine/threonine kinases that are activated by MAPK, and inhibition of these two kinases results in decreased glioma formation in mice (51). Notably, pT205 and pS207 phosphosites on EIF4E have not been functionally characterized. Their significant upregulation points to a possible role in regulating the function of EIF4E.

Other regulated proteins in PI3K pathway in RAS-NHA cells included Niban (FAM129A). FAM129A regulates the phosphorylation of multiple proteins involved in translation regulation, including EIF2A, EIF4EBP1 and RPS6KB1. Consistent with overexpression of FAM129A in multiple cancer types (52–54), we observed a 2-fold upregulation of its abundance in RAS-NHA cells. Additionally, we detected a 2-fold upregulation of phosphorylation at pS602. This site is a known substrate of

Signaling and Epigenetics in Astrocytoma

to the changes in downstream effectors of RAS, we also found changes in phosphorylation of multiple proteins upstream of HRAS, pointing to feedback regulation mechanisms. For example, EGFR, FGFR4, EPHA2, and TGFBR2 receptors showed significant phosphorylation upregulation in RAS-NHA cells. EGFR phosphosites pT648, pS991, and pS1166 displayed 3.5-fold, 3.7-fold, and 1.6-fold upregulation, respectively. The pS991 and pS1166 sites are both thought to be negative regulators of EGFR (63, 64). EGFR pT648 has not been previously functionally characterized, but our results suggest that by analogy to pS991 and pS1166 this phosphosite may also be involved in feedback downregulation of EGFR. Additionally, pS251 and pS369 of ERFF1, which is a negative regulator of EGFR (65), were also significantly upregulated, raising the possibility that phosphorylation at these sites may promote the activity of ERFF1 as a negative regulator of EGFR.

Furthermore, we found that FGFR4 phosphorylation site pS573 was one of the most significantly upregulated (34-fold) sites in our experiment. Whether phosphorylation of pS573 on FGFR4 promotes its activation or inhibition, however, remains unclear. Interestingly, the receptor tyrosine kinases (RTKs) EPHA2 and EPHB2, which are both involved in cell adhesion and migration dependent signaling, showed distinct responses to RAS activation. Although EPHB2 didn't reveal regulation in RAS-NHA cells, EPHA2 showed significant upregulation at the protein and phosphorylation levels. The AKT-mediated EPHA2 phosphorylation on residue pS897 displayed a 4.9-fold upregulation.

C) Dual MAPK and PI3K Inhibition Yields a More Complete Reversal of the Phosphorylation Changes Driven by Oncogenic HRAS Compared with MEK Inhibition Only—We next wanted to know if the protein and phosphorylation regulation changes observed as a result of oncogenic HRAS overexpression in NHA cells could be reversed by inhibition of its downstream effects. Specifically, we sought to identify and compare the sets of regulatory changes reversible by either a MEK inhibitor alone or in combination with a PI3K inhibitor. For this purpose, we treated RAS-NHA cells with a highly selective small-molecule inhibitor of MEK1/2 (GDC-0973, Genentech Inc.) (11) and a potent class I PI3K inhibitor (GDC-0941, Genentech Inc.) (12) and measured global protein and phosphorylation changes (Fig. 5A).

MEK Inhibition—After MEK inhibition alone, we identified 95 and 170 phosphosites displaying at least 2-fold down or upregulation in RAS-NHA cells compared with DMSO treatment, respectively (supplemental Table S8 and supplemental Fig. S6A). Overall, we identified 115 (41.4%) of all significantly upregulated sites observed in the RAS-NHA versus control-NHA cells experiments. Although 15 (13.0%) of upregulated sites in RAS-NHA cells displayed at least 2-fold downregulation after MEK inhibition (reversible sites) (Fig. 5B), 25 (21.7%) were paradoxically activated upon MEK inhibition beyond that seen from HRAS overexpression (nonreversible sites) (Fig. 5C)

(supplemental Table S9). Reversible sites that showed at least 2-fold downregulation after MEK inhibition, included MAPK1 (pY187) and MAPK3 (pY204) as well as phosphoproteins involved in RAS protein signaling transduction (FAM129B and ARHGAP29). Other reversible sites included EPS8L2 (pS459), NES (pT338), PML (pS527) and TACC1 (pS276). On the contrary, we observed nonreversible sites such as AKT1S1 (pT246), which is consistent with the compensatory activation of PI3K pathway as a result of MEK inhibition observed in other systems (66). Interestingly, NES phosphorylation sites showed opposite regulation after MEK inhibition. Although pT338 was reversible, pS905 was activated upon MEK inhibition. HMGA1 (pS44), GREM1 (pS77), and IL6ST (pS667) were also categorized as non-reversible sites. In addition to the changes in downstream proteins, we found altered phosphorylation of multiple proteins upstream of MEK pointing to feedback regulation mechanisms. For example, NGFR pS313, which already showed 16-fold upregulation in RAS-NHA cells (compared with control-NHA cells) was upregulated 20-fold more after MEK inhibition alone, suggesting that this site might be involved in compensatory activation of NGFR following MEK inhibition.

Dual MEK and PI3K Inhibition—We next investigated the efficacy of dual MEK and PI3K inhibition in reversing signaling changes mediated by oncogenic HRAS in RAS-NHA cells. By applying quantitative MS after 4 h of treatment, we found 403 phosphosites displaying at least 2-fold downregulation after dual inhibition compared with DMSO treatment. Only a minority of phosphosites (129) showed more than 2-fold upregulation after dual inhibition (supplemental Table S10 and supplemental Fig. S6B).

In this dual inhibition experiment, we quantified 217 (78.1%) of significantly upregulated phosphosites identified in the RAS-NHA versus control-NHA cells experiment. Of these, 49 (22.6%) showed significant downregulation after dual inhibition (reversible sites) (Fig. 5D) and 6 (2.8%) displayed significant upregulation (nonreversible sites) (Fig. 5E) (supplemental Table S11).

Among the reversible phosphorylation sites, MAPK1 (pT185, pY187), MAPK3 (pY90), and AKT1S1 (pT246) showed the most substantial downregulation (up to 68-fold) after dual inhibition. Furthermore, downstream targets of the MAPK signaling cascade such as pS641, pS646, pS692 and pS696 of FAM129B, ARHGAP29 pS1029, and pS405 CTTN showed significant reversed phosphorylation patterns. Multiple downstream targets of PI3K, including BAG3 and EIF4E showed significant reversed phosphorylation patterns as well. Interestingly, pS946 and pT648 of the upstream regulator EGFR were also reversed. Similar to MEK inhibition only, NES phosphorylation sites showed distinct responses to dual inhibition. Although pT338 was significantly downregulated after MEK inhibition only and dual inhibition, pS1577 was one of the few nonreversible phosphorylation sites after dual inhibition. Other nonrevers-

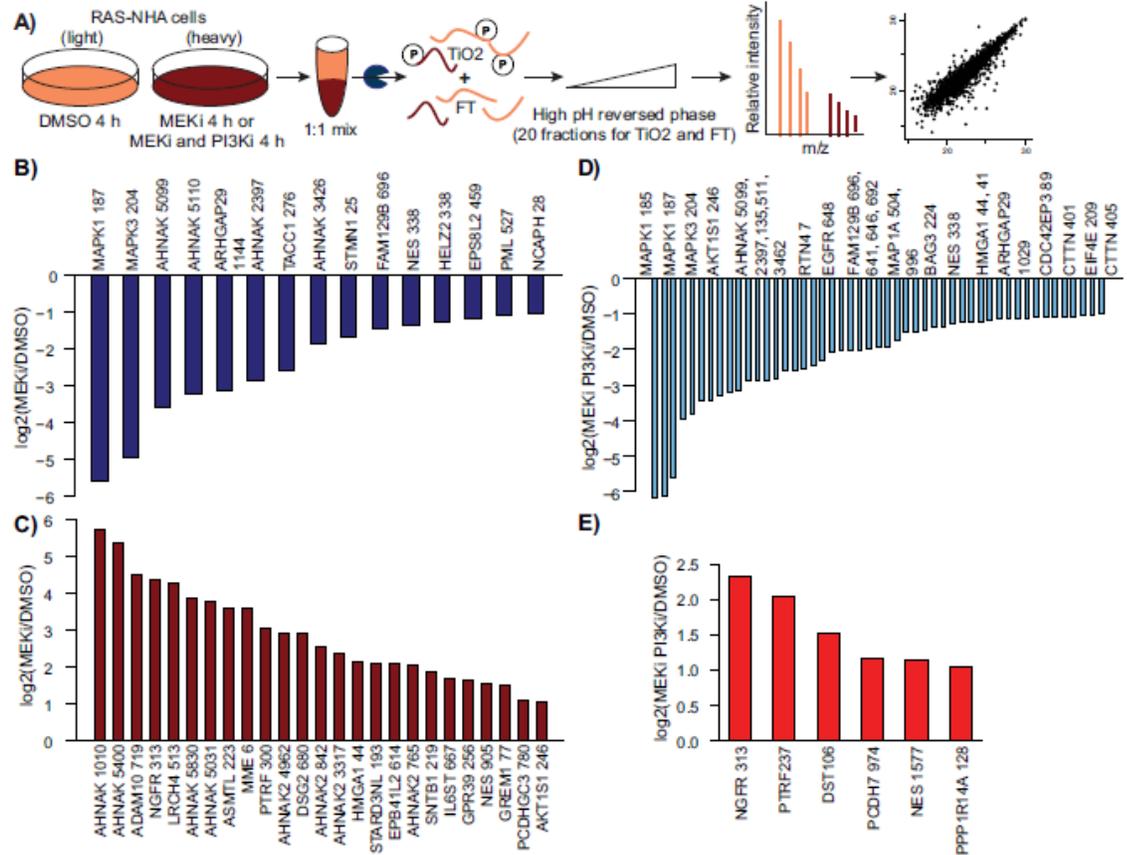


FIG. 5. MEK and PI3K inhibition reverses phosphorylation signaling patterns driven by oncogenic HRAS overexpression in RAS-NHA cells. A, Experimental overview. B, Down and (C) upregulated phosphosites after MEK inhibition that were also upregulated in RAS-NHA versus control-NHA cells. D, Down and (E) upregulated phosphosites after dual MEK and PI3K inhibition that were upregulated in RAS-NHA cells compared with control-NHA cells.

ible phosphosites included NGFR (pS313), PTRF (pS237), DST (pS106), PCDH7 (pS974), and PPP1R14A (pS126), suggesting that different pathways, other than PI3K and MAPK regulate these sites.

MEK Inhibition Only Versus Dual MEK and PI3K Inhibition—Comparing both experiments to each other revealed that several phosphorylation sites showed opposite or more substantial downregulation after MEK and PI3K inhibition compared with MEK inhibition alone. For example, pT246 of AKT1S1 showed 12-fold abundance-independent downregulation after dual inhibition, but 2-fold abundance-independent upregulation after MEK inhibition alone. RICTOR pS1302, which was 1.7-fold upregulated after MEK inhibition, showed no changes in phosphorylation after simultaneous inhibition. Several phosphosites of CDC42EP proteins and TGFB2, which showed abundance-independent increased phosphorylation after MEK inhibition displayed

downregulation after dual inhibition. Multiple phosphosites of IRS-1 and 2, which are known substrates of MAPK (67), showed 3.5-fold or higher downregulation after dual inhibition, but only moderate responses (< 1.5-fold) after MEK inhibition. We further observed that in addition to RAS and cell cycle associated proteins, phosphosites of focal adhesions and cell motion proteins were significantly downregulated after dual inhibition compared with MEK inhibition alone (supplemental Fig. S6C, S6D, and S6F).

Taken together, inhibition of MEK reversed only a minority of HRAS-driven phosphorylation patterns. In comparison, simultaneous inhibition of the MAPK and PI3K signaling pathways reversed many signaling changes driven by oncogenic HRAS with the exception of six phosphorylation sites that were not reversible. Dual inhibition was clearly superior in reversing oncogenic signaling in HRAS driven glioma cells compared with MEK inhibition alone.

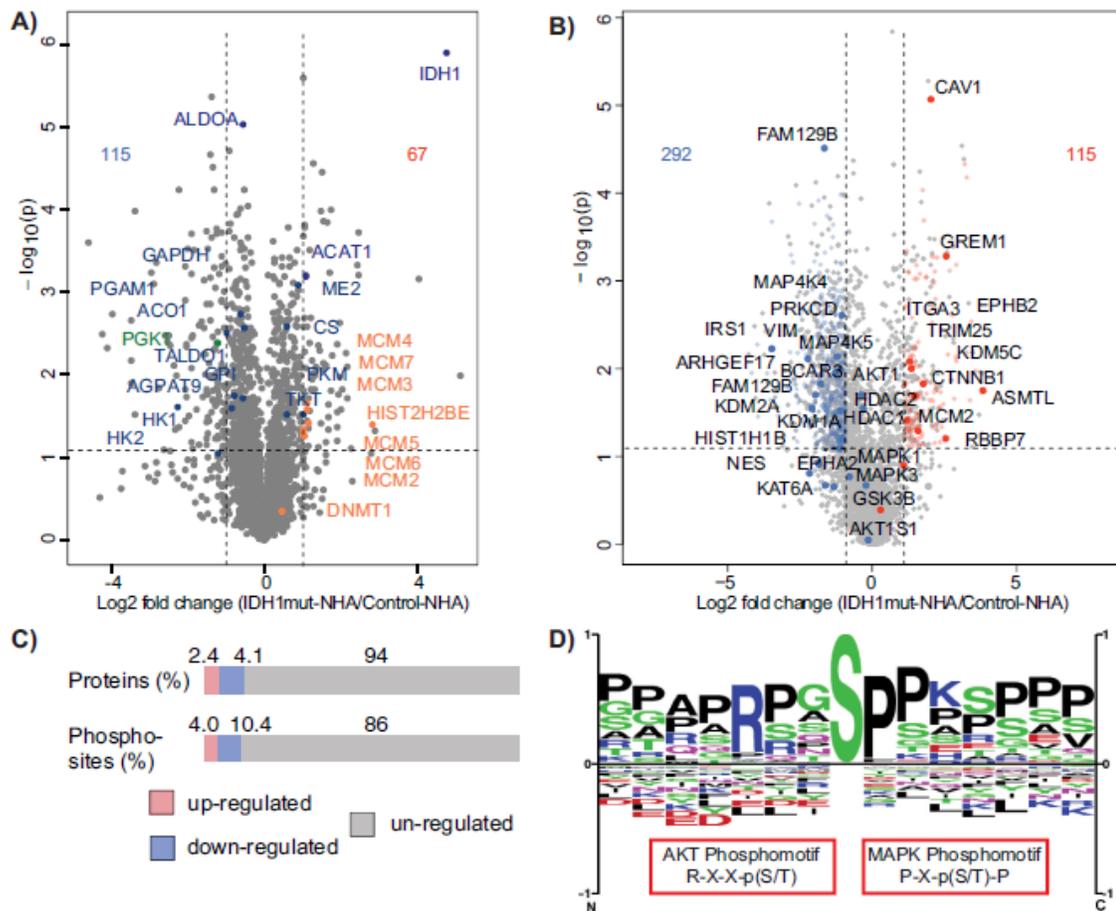


Fig. 6. Significant dysregulation of epigenetic and metabolic proteins in IDH1mut-NHA cells. The volcano plots illustrate the Log₂ fold changes in IDH1mut-NHA compared with control-NHA cells (x axis) at the (A) protein abundance or (B) phosphorylation level and *p* values estimated in biological replicates (y axis). C, Proportion of regulated (at least 2-fold) proteins and phosphosites in all 3 biological replicates. D, Downregulated phosphosites in IDH1mut-NHA cells are enriched for MAPK and PI3K motifs compared with control cells.

III) Oncogenic Activity of Mutant IDH1 Induces Global Protein Expression Changes and Perturbs Epigenetic Regulation in IDH1mut-NHA Cells—

Global Proteome and Phosphoproteome in IDH1mut-NHA Cells—IDH1 mutation is a selective marker of secondary glioblastoma, and given its pivotal role in the inhibition of histone demethylases and reorganization of DNA methylation, we assessed whether protein abundance changes can be linked to the histone PTM signature of IDH1mut-NHA cells.

Overall, we observed more down than upregulated proteins as well as phosphorylation sites in IDH1mut-NHA cells compared with control-NHA cells. In total, 115 proteins (4.1%) and 292 phosphosites (10.4%) showed at least 2-fold downregulation compared with 67 (2.4%) proteins and 115 (4.0%) phosphosites displaying significant upregulation (Fig. 6C and sup-

plemental Table S12). The observed protein abundance changes were concordant with prior transcriptomic profiling of IDH1mut-NHA cells that revealed a global downregulation of gene expression compared with wild type IDH1 (IDH1wt) NHA cells (22). Both proteomic and transcriptomic observations are consistent with the G-CIMP of IDH1 mutant gliomas, which induces global reorganization of the transcriptome (22, 68). Similar to other studies showing that mRNA expression levels are not fully reflective of protein abundances (23, 24), we found that relative protein intensities in IDH1mut-NHA cells did not correlate with relative gene expression when comparing our data with the microarray data (supplemental Fig. S7).

IDH1 Mutation Induces Epigenetic Alterations in IDH1mut-NHA Cells—Upregulated proteins in IDH1mut-NHA cells were

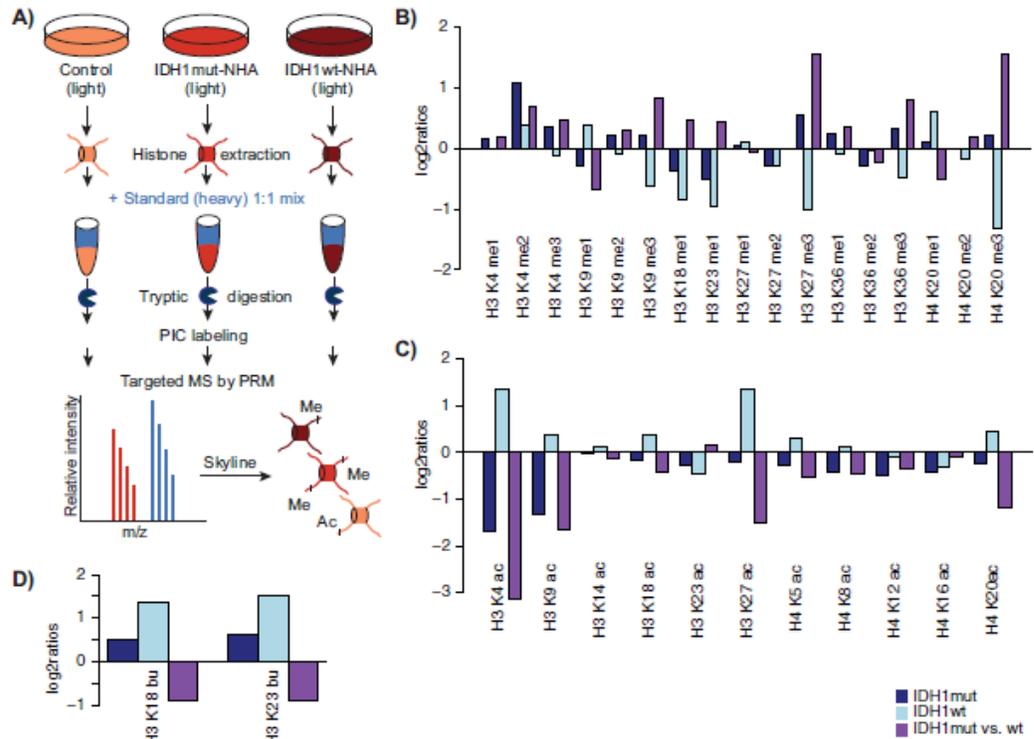


FIG. 7. Histone PTM occupancies in IDH1mut- and IDH1wt-NHA cells. A, Experimental design to assign histone PTM occupancies. The barplots illustrate histone (B) methylation, (C) acetylation, and (D) butyrylation changes (log₂ ratios) in IDH1mut- (dark blue) and IDH1wt-NHA (light blue) compared with control-NHA cells and IDH1mut- compared with IDH1wt-NHA cells (purple).

enriched for chromatin-associated proteins, such as DNA replicating licensing factors MCM2–7, DNMT1, and HELLS (highlighted in orange in Fig. 6A). In addition, multiple chromatin-associated proteins, including α -KG-dependent KDM2A, displayed significant phosphorylation alterations (Fig. 6B). Interestingly, the most significantly upregulated phosphorylation site in IDH1mut-NHA cells was localized on N-acetylserotonin O-methyltransferase-like protein ASMTL (pS223) protein. Additionally, phosphosites in lysine specific demethylases KDM1A (pS166) and KDM2A (pS740) showed significant downregulation, whereas pS301 of KDM5C displayed significant upregulation.

α -KG-dependent histone demethylases are major targets of IDH1 (20, 69). Most α -KG dependent histone modifying enzymes, however, did not show significant changes in protein abundance in our dataset, suggesting that functional inhibition of α -KG dependent enzymes by mutant IDH1 does not significantly affect their protein expression level.

Histone PTM Occupancies in IDH1mut-NHA Cells Reveal a Transcriptional Silencing State—We next sought to investigate the downstream targets of α -KG dependent histone demethylases. To this end, we compared histone PTM occu-

pancies of IDH1mut-NHA to control-NHA cells. We quantified site-specific histone PTMs by applying a hybrid chemical labeling method and MS analysis by PRM (Fig. 7A and supplemental Table S13).

Histone Methylation—Overall, we identified 29 different methylation marks on histone H3. Consistent with previous studies using antibody-based detection approaches (22, 70), multiple histone methylation sites showed significant increase in methylation in IDH1mut-NHA compared with control-NHA cells, including H3K9me3 (1.2-fold), H3K27me3 (1.5-fold), and H3K36me3 (1.3-fold). We further identified two other significantly upregulated tri-methylated marks on histone H3 (H3K4me3 and H4K20me3) (Fig. 7B). Although H3K9me3, H3K27me3, H4K20me3, and H3K36me3 are known to be associated with transcriptional silencing, H3K4me3 is linked to transcriptional activation (71). To control for possible effects of IDH1 protein overexpression, we also included IDH1wt-NHA cells in our histone PTM analysis. We found that IDH1wt-NHA cells presented opposite histone tri-methylation regulation compared with IDH1mut-NHA cells, as H3K9me3, H3K27me3, H3K36me3, and H4K20me3 displayed significant downregulation in IDH1wt-NHA cells. All three histone marks,

Signaling and Epigenetics in Astrocytoma

displayed 1.8-, 2.9-, 1.7-, and 2.9-fold increased upregulation in IDH1mut- compared with IDH1wt-NHA cells, respectively. These results are consistent with the function of wild type IDH1, which produces α -KG and subsequently activates α -KG-dependent histone demethylases. Mutant IDH1, however, inhibits α -KG-dependent histone demethylases, which results in increased histone tri-methylation.

In contrast to tri-methylated histone lysines, mono- and di-methylated lysines did not show significant regulation and appear not to be affected by mutant IDH1. Overall, the observed methylation patterns suggest a transcriptional repressive state in IDH1mut-NHA cells.

Histone Acetylation—While the majority of histone methylation marks have been linked to transcriptional repression, lysine acetylation is associated with transcriptional activation (72). We found that IDH1mut-NHA cells displayed significant downregulation of histone acetylation. In fact, all associated marks, including H3K4ac, H3K9ac, H3K23ac, H3K27ac, H4K5ac, H4K8ac, H4K12ac, H4K16ac, and H4K20ac, showed significant downregulation (Fig. 7C). H3K4ac and H3K9ac displayed the most significant downregulation (8.6-fold and 3.1-fold compared with IDH1wt-NHA cells, respectively). Both of these marks, if acetylated, are known to induce transcriptional activation (73, 74). Finally, H3K14ac and H3K18ac showed a trend for downregulation in IDH1mut-NHA cells. These results indicate that mutant IDH1 induces a global transcriptional state through reduced acetylation.

Histone Butyrylation—Among all identified histone PTMs, butyrylation was one of the most upregulated marks. We identified three histone butyrylation sites on histone H3 (H3K14, H3K18, and H3K23). The H3K18 and H3K23 butyrylation marks displayed 1.4-fold and 1.5-fold upregulation in IDH1mut-NHA cells compared with control cells, respectively (Fig. 7D). Histone butyrylation, however, was even more significantly upregulated in IDH1wt-NHA cells. In fact, histone butyrylation was 1.9-fold downregulated in IDH1mut- compared with IDH1wt-NHA cells. These observations suggest a regulatory link between IDH1 mutation and histone butyrylation.

In summary, our histone PTM analysis revealed a significant upregulation of histone lysine trimethylation and downregulation of histone acetylation and butyrylation in IDH1mut-NHA cells compared with IDH1wt-NHA cells, pointing to a global transcriptional repressive state mediated by mutant IDH1, consistent with the observed downregulated proteome in IDH1mut-NHA cells.

Mutant IDH1 Induces Metabolic Changes in IDH1mut-NHA Cells—In addition to epigenetic changes, we found that metabolic proteins (highlighted in blue in Fig. 6A and supplemental Fig. S8) regulating glycolysis, lipolysis, as well as oxidoreductase activity were significantly dysregulated in IDH1mut-NHA cells compared with control-NHA cells. Although metabolic proteins AGPAT9, ACO1, GAPDH, PGAM1, TALDO1, ALDOA, HK1, HK2, and GPI displayed downregulation,

ACAT1, ME2, CS, TKT, and PKM were significantly upregulated. Changes in glucose flux in IDH1mut-NHA cells were reflected by the downregulation of the AKT pathway (Fig. 6D) and other enzymes stimulating the glycolytic flux, such as hexokinases. These findings have also been observed in a label free analysis, comparing the proteomes of IDH1mut-NHA to control-NHA cells in a single-run mass spectrometry workflow (supplemental Fig. S8). These observations are concordant with the pivotal role of IDH1 within the TCA cycle (75).

IV) HRAS-overexpressing or IDH1-mutated Astrocytomas Are Driven by Different Oncogenic Pathways—We showed that both RAS-NHA and IDH1mut-NHA cells express dysregulated altered oncogenic pathways. Using PCA, both cell lines can be clearly differentiated from each other based on their proteomic and phosphoproteomic profiles (supplemental Fig. S9). The driving role of the MAPK and PI3K pathways in RAS-NHA cells was not reflected in IDH1mut-NHA cells. The surrounding sequences of downregulated phosphosites in IDH1mut-NHA cells significantly matched with canonical MAPK substrate recognition motifs (Fig. 6D). Concordantly, downstream targets of the MAPK signaling cascade such as pS641, pS646, pS692 and pS696 of FAM129B (57), which displayed significant upregulation in RAS-NHA cells, showed significant downregulation in IDH1mut-NHA cells. In addition, PI3K targets, such as AKT1S1 pT246 did not reveal any regulation in IDH1mut-NHA cells (Fig. 6B).

Other signaling pathways that showed different regulation in IDH1mut- compared with RAS-NHA cells include the Wnt pathway. IDH1mut-NHA cells displayed significant upregulation of β -catenin phosphosites and no upregulation of the AKT-mediated inactivating GSK3 β pS9 phosphorylation site, indicating an inactive or less active Wnt signaling pathway compared with RAS-NHA cells (supplemental Fig. S10). This finding is consistent with a recent report highlighting that IDH1 Arg132 mutation reduces cell proliferation by downregulating Wnt/ β -Catenin signaling in glioblastoma (76). Although the role of the Wnt pathway in human glioma remains unclear, the Wnt pathway has oncogenic (77, 78) and antioncogenic (79–81) activities in several cancer types.

Interestingly, one of the most downregulated (7-fold) proteins in IDH1mut-NHA cells was NES, which showed opposite regulation in RAS-NHA cells with a 10-fold upregulation at the protein level. We further showed by label free MS analysis and Western blot that NES was exclusively downregulated in IDH1mut-NHA cells when compared with IDH1wt-NHA, RAS-NHA, or control cells (supplemental Table S14 and supplemental Fig. S11). NES has been previously reported to be overexpressed in GBM (27), to correlate with the malignancy of glioblastoma (82), and its inhibition significantly reduced tumor growth (83).

At the cell membrane level, EPHA2 and EPHB2 showed distinct responses to RAS activation or IDH1 mutation in our experiments. Although EPHA2 showed significant upregulation in RAS-NHA cells, no significant changes were observed

in IDH1mut-NHA cells. EPHB2, however, revealed significant upregulation at the protein and phosphosite level exclusively in IDH1mut-NHA cells.

To assess the involvement of identified proteins in cancer we compared our phosphoproteomic and proteomic data to a catalogue of 150 tumor suppressor genes and 350 oncogenes listed in the cancer gene census (84). Overall, we quantitated the relative abundance of 33 tumor suppressor proteins (TSPs) and 102 oncoproteins in our proteomic or phosphoproteomic experiments. Among the TSPs, Neurofibromin (NF1) and NF2 showed the most significant downregulation of both abundance and phosphorylation in RAS-NHA cells but did not show any significant regulation in IDH1mut-NHA cells. At the abundance level, NF1 (identified in one of the three biological replicates) and NF2 displayed 9-fold and 2.8-fold downregulation, respectively. NF1 and NF2 have been shown to inhibit the activity of HRAS but not vice-versa (85–88). Loss of NF1 subsequently results in the activation of multiple pathways involved in gliomagenesis, including the MAPK and PI3K pathways (89). The regulatory role of NF2, however, remains poorly understood (90). Downregulated phosphorylation sites included pS864, pS2523 (2.6-fold), pS2543 (4.6-fold) on NF1 and pS13 (2.8-fold) on NF2 proteins. The downregulated phosphosites of NF1 have been previously reported, but their function is unknown. In particular, all three phosphorylation residues we uncovered on NF2 including pS10, pS12, and pS13 have not been characterized, pointing to previously unknown regulatory mechanisms.

Conclusion and Outlook—In the present study we used SILAC-based MS to characterize global phosphorylation and protein abundance changes in immortalized NHA cells expressing oncogenic HRAS or mutant IDH1. Although these two models are not fully reflective of primary and secondary GBM, they showed very distinct proteomic profiles consistent with fundamentally different biological programs driving tumorigenesis in primary and secondary GBM (Fig. 8).

In RAS-NHA cells, we found activation of MAPK and PI3K pathways. In addition to known changes in MAPK and PI3K pathways, we identified significant changes in previously uncharacterized phosphosites within these pathways implicating them as likely sites of regulation downstream of oncogenic HRAS signaling. These include EIF4E, FAM129A, FAM129B, PAK, and NF2.

MEK inhibition resulted in partial reversibility of the changes driven by oncogenic HRAS, whereas dual MAPK and PI3K inhibition yielded a more complete reversal. Some compensatory changes because of MEK inhibition as a single agent, including PI3K activation and upstream regulators TGFBR2 and EGFR were blocked by the dual inhibition. However, even dual MEK and PI3K inhibition did not reverse all signaling changes mediated by oncogenic HRAS. NES (pS1577), NGFR (pS313), PTRF (pS237), DST (pS106), PCDH7 (pS974), and PPP1R14A (pS128) were some of the few non-reversible phosphorylation sites after dual inhibition, suggesting that

previously unknown feedback mechanisms regulate these sites. Although dual inhibition of MEK and PI3K is clearly superior to single MEK inhibition and is currently in phase I clinical trials, based on these findings and multiple other studies, clinical efficacy remains to be determined (91).

In addition to the activation of MAPK and PI3K pathways in RAS-NHA cells, we identified changes in several other pathways. We observed a potential driving role for PAKs in glioma development, consistent with previous studies (92–94). We further found several interaction partners of the RAL proteins, which are commonly over-expressed in multiple cancer types, including glioblastoma, to be highly upregulated in RAS-NHA cells (62) (95, 96). In addition to regulation of the main signaling cascades downstream of RAS, we observed substantial phosphorylation and protein expression changes in other pathways such as the Wnt pathway, presumably triggered by interplay between the cascades. We also found regulated phosphorylation sites of multiple proteins upstream of RAS, pointing to feedback mechanisms. For instance, the FGF signaling pathway, in particular FGFR4, which showed massive upregulation at the phosphorylation level has been reported to stimulate the growth of GBM (97). We also identified activating phosphorylation sites on EPHA2, which have been associated with glioblastoma invasion, consistent with the malignant phenotype of RAS-NHA cells (98). EPHA2 and other Ephrin family receptors are particularly attractive for targeted therapy, because they are expressed at very low abundance or not expressed in healthy tissues but show high expression in most cancerous tissues (99).

Unlike the changes in RAS-NHA cells characterized by dysregulation of major signaling pathways, IDH1mut-NHA cells displayed a global downregulation of protein expression, detected at the level of both protein abundance and phosphorylation. The overall downregulation of protein expression is consistent with evidence of genome-scale transcriptional silencing identified previously (22). Protein expression changes were accompanied by changes in histone PTMs, pointing to transcriptional repression, which in turn results from mutant IDH1-mediated inhibition of histone demethylases. Our histone PTM analysis revealed a significant upregulation of tri-methylated histone lysines, which are known to be demethylated exclusively by α -KG-dependent histone demethylases. Therefore, our data suggest that α -KG dependent histone demethylases are inhibited in IDH1mut-NHA cells. This observation is consistent with previous reports that demonstrated the overproduction of 2-HG, acting as a competitive inhibitor of α -KG-dependent enzymes in IDH1mut-NHA cells (18). Increase in methylation was accompanied by a decrease in acetylation at essentially all measured sites providing further support for a state of global transcriptional repression mediated by mutant IDH1. We speculate that histone deacetylase (HDAC) inhibitors, including those currently in clinical trials for the treatment of GBM (100), may reverse the downregulation of the histone acetylome in IDH1mut-NHA

targeted MS-based demonstration of multiple histone PTM changes driven by mutant IDH1 in a cellular model of secondary glioblastoma. Based on these biological observations, we predict very different therapeutic interventions to be useful in primary versus secondary GBM. We provide a valuable resource of proteins altered in glioma, including potential biomarkers or therapeutic targets.

Acknowledgments—We thank R. O. Pieper and J. Mukherjee for helpful discussions and providing immortalized human astrocytes, T. Maile for help with the analysis of histone PTMs and R. Levin for helpful input.

* This work was supported by National Institutes of Health/National Institute of General Medical Sciences (NIH/NIGMS) grant 8P41GM103481 and the Adelson Medical Research Foundation. The Howard Hughes Medical Institute (HHMI) contributed to the purchase of the LTQ-Orbitrap Velos. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

§ This article contains supplemental material.

¶ Current address: Dept. of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany.

¶ To whom correspondence should be addressed: Pharmaceutical Chemistry Institution, University of California, San Francisco, 600 16th Street Genentech Hall N472, Box 2240, San Francisco, CA 94158-2517. Tel: 415-476 5641; Fax: 415-502 1655; E-mail: alb@cgl.ucsf.edu.

REFERENCES

- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. (2015) Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012: Globocan 2012. *Int. J. Cancer* **136**, E359–E386
- Furnari, F. B., Fenton, T., Bachoo, R. M., Mukasa, A., Stommel, J. M., Stegh, A., Hahn, W. C., Ligon, K. L., Louis, D. N., Brennan, C., Chin, L., DePinho, R. A., and Cavenee, W. K. (2007) Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev.* **21**, 2683–2710
- Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., Scheithauer, B. W., and Kleihues, P. (2007) The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **114**, 97–109
- Ostrom, Q. T., Gittleman, H., Liao, P., Rouse, C., Chen, Y., Dowling, J., Wolinsky, Y., Kruchko, C., and Barnholtz-Sloan, J. (2014) CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. *Neuro-Oncol.* **16**, iv1–iv63
- Ohgaki, H., and Kleihues, P. (2013) The Definition of Primary and Secondary Glioblastoma. *Clin. Cancer Res.* **19**, 764–772
- Brennan, C. W., Verhaak, R. G. W., McKenna, A., Campos, B., Nounshmehr, H., Salama, S. R., Zheng, S., Chakravarty, D., Sanborn, J. Z., Berman, S. H., Beroukhi, R., Bernard, B., Wu, C.-J., Genovese, G., Shmulevich, I., Barnholtz-Sloan, J., Zou, L., Vegesna, R., Shukla, S. A., Ciriello, G., Yung, W., Zhang, W., Sougnez, C., Mikkelson, T., Aldape, K., Bigner, D. D., Van Meir, E. G., Prados, M., Sloan, A., Black, K. L., Eschbacher, J., Finocchiaro, G., Friedman, W., Andrews, D. W., Guha, A., Iacocca, M., O'Neill, B. P., Foltz, G., Myers, J., Weisenberger, D. J., Penny, R., Kucherlapati, R., Perou, C. M., Hayes, D. N., Gibbs, R., Marra, M., Mills, G. B., Lander, E., Spellman, P., Wilson, R., Sander, C., Weinstein, J., Meyerson, M., Gabriel, S., Laird, P. W., Haussler, D., Getz, G., and Chin, L. (2013) The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477
- Sonoda, Y., Ozawa, T., Hirose, Y., Aldape, K. D., McMahon, M., Berger, M. S., and Pieper, R. O. (2001) Formation of intracranial tumors by genetically modified human astrocytes defines four pathways critical in the development of human anaplastic astrocytoma. *Cancer Res.* **61**, 4956–4960
- Feldkamp, M. M., Lala, P., Lau, N., Roncari, L., and Guha, A. (1999) Expression of activated epidermal growth factor receptors, Ras-guanosine triphosphate, and mitogen-activated protein kinase in human glioblastoma multiforme specimens. *Neurosurgery* **45**, 1442–1453
- Prabhu, A., Sarcar, B., Miller, C. R., Kim, S.-H., Nakano, I., Forsyth, P., and Chinnaiyan, P. (2015) Ras-mediated modulation of pyruvate dehydrogenase activity regulates mitochondrial reserve capacity and contributes to glioblastoma tumorigenesis. *Neuro-Oncol.* **17**, 1220–1230
- Guha, A., Feldkamp, M. M., Lau, N., Boss, G., and Pawson, A. (1997) Proliferation of human malignant astrocytomas is dependent on Ras activation. *Oncogene* **15**, 2755–2765
- Rice, K. D., Aay, N., Anand, N. K., Blazey, C. M., Bowles, O. J., Bussenius, J., Costanzo, S., Curtis, J. K., Defina, S. C., Dubenko, L., Engst, S., Joshi, A. A., Kennedy, A. R., Kim, A. I., Koltun, E. S., Loughheed, J. C., Manalo, J.-C. L., Martini, J.-F., Nuss, J. M., Peto, C. J., Tsang, T. H., Yu, P., and Johnston, S. (2012) Novel Carboxamide-Based Allosteric MEK Inhibitors: Discovery and Optimization Efforts toward XL518 (GDC-0973). *ACS Med. Chem. Lett.* **3**, 416–421
- Folkes, A. J., Ahmadi, K., Alderton, W. K., Alix, S., Baker, S. J., Box, G., Chuckowree, I. S., Clarke, P. A., Depledge, P., Eccles, S. A., Friedman, L. S., Hayes, A., Hancox, T. C., Kugendradas, A., Lensun, L., Moore, P., Olivero, A. G., Pang, J., Patel, S., Pergl-Wilson, G. H., Raynaud, F. I., Robson, A., Saghir, N., Salphati, L., Sohal, S., Ulfisch, M. H., Valenti, M., Wallweber, H. J. A., Wan, N. C., Wiesmann, C., Workman, P., Zhyvoloup, A., Zvelebil, M. J., and Shuttleworth, S. J. (2008) The identification of 2-(1H-indazol-4-yl)-6-(4-methanesulfonyl-piperazin-1-ylmethyl)-4-morpholin-4-yl-thieno[3,2-d]pyrimidine (GDC-0941) as a potent, selective, orally bioavailable inhibitor of class I PI3K kinase for the treatment of cancer. *J. Med. Chem.* **51**, 5522–5532
- Wu, S., Wang, S., Zheng, S., Verhaak, R., Koul, D., and Yung, W. K. A. (2016) MSK1-Mediated β -catenin phosphorylation confers resistance to PI3K/mTOR inhibitors in glioblastoma. *Mol. Cancer Ther.* **15**, 1656–1668
- El Meskini, R., Iacovelli, A. J., Kulaga, A., Gumprecht, M., Martin, P. L., Baran, M., Householder, D. B., Van Dyke, T., and Weaver Ohler, Z. (2015) A preclinical orthotopic model for glioblastoma recapitulates key features of human tumors and demonstrates sensitivity to a combination of MEK and PI3K pathway inhibitors. *Dis. Model. Mech.* **8**, 45–56
- Parsons, D. W., Jones, S., Zhang, X., Lin, J. C.-H., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.-M., Gallia, G. L., Olivi, A., McLendon, R., Rasheed, B. A., Keir, S., Nikolskaya, T., Nikolsky, Y., Busam, D. A., Tekleab, H., Diaz, L. A., Hartigan, J., Smith, D. R., Strausberg, R. L., Marie, S. K. N., Shinjo, S. M. O., Yan, H., Riggins, G. J., Bigner, D. D., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V. E., and Kinzler, K. W. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807
- Cohen, A., Holmen, S., and Colman, H. (2013) IDH1 and IDH2 mutations in gliomas. *Curr. Neurol. Neurosci. Rep.* **13**, 345
- Yan, H., Parsons, D. W., Jin, G., McLendon, R., Rasheed, B. A., Yuan, W., Kos, I., Batinić-Haberle, I., Jones, S., Riggins, G. J., Friedman, H., Friedman, A., Reardon, D., Herndon, J., Kinzler, K. W., Velculescu, V. E., Vogelstein, B., and Bigner, D. D. (2009) IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773
- Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., Fantin, V. R., Jang, H. G., Jin, S., Keenan, M. C., Marks, K. M., Prins, R. M., Ward, P. S., Yen, K. E., Liu, L. M., Rabinowitz, J. D., Cantley, L. C., Thompson, C. B., Vander Heiden, M. G., and Su, S. M. (2010) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **465**, 966
- Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.-H., Ito, S., Yang, C., Wang, P., Xiao, M.-T., Liu, L., Jiang, W., Liu, J., Zhang, J., Wang, B., Frye, S., Zhang, Y., Xu, Y., Lei, Q., Guan, K.-L., Zhao, S., and Xiong, Y. (2011) Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30
- Chowdhury, R., Yeoh, K. K., Tian, Y.-M., Hillinghaus, L., Bagg, E. A., Rose, N. R., Leung, I. K. H., Li, X. S., Woon, E. C. Y., Yang, M., McDonough, M. A., King, O. N., Clifton, I. J., Klose, R. J., Claridge, T. D. W., Ratcliffe, P. J., Schofield, C. J., and Kawamura, A. (2011) The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12**, 463–469
- Rose, N. R., McDonough, M. A., King, O. N. F., Kawamura, A., and Schofield, C. J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem. Soc. Rev.* **40**, 4364–4397

22. Turcan, S., Rohle, D., Goenka, A., Walsh, L. A., Fang, F., Yilmaz, E., Campos, C., Fabius, A. W. M., Lu, C., Ward, P. S., Thompson, C. B., Kaufman, A., Guryanova, O., Levine, R., Heguy, A., Viale, A., Morris, L. G. T., Huse, J. T., Mellinghoff, I. K., and Chan, T. A. (2012) IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479–483

23. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548

24. Maier, T., Göell, M., and Serrano, L. (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583**, 3966–3973

25. Johnson, H., and White, F. M. (2014) Quantitative analysis of signaling networks across differentially embedded tumors highlights interpatient heterogeneity in human glioblastoma. *J. Proteome Res.* **13**, 4581–4593

26. Nilsson, C. L., Dillon, R., Devakumar, A., Shi, S. D.-H., Greig, M., Rogers, J. C., Krastins, B., Rosenblatt, M., Kilmer, G., Major, M., Kaboord, B. J., Sarracino, D., Rezaei, T., Prakash, A., Lopez, M., Ji, Y., Priebe, W., Lang, F. F., Colman, H., and Conrad, C. A. (2010) Quantitative phosphoproteomic analysis of the STAT3/IL-6/HIF1 α signaling network: an initial study in GSC11 glioblastoma stem cells. *J. Proteome Res.* **9**, 430–443

27. Heroux, M. S., Chesnik, M. A., Halligan, B. D., Al-Gizawi, M., Connelly, J. M., Mueller, W. M., Rand, S. D., Cochran, E. J., LaViolette, P. S., Malkin, M. G., Schmainda, K. M., and Mirza, S. P. (2014) Comprehensive characterization of glioblastoma tumor tissues for biomarker identification using mass spectrometry-based label-free quantitative proteomics. *Physiol. Genomics* **46**, 467–481

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

29. Ohba, S., Mukherjee, J., See, W. L., and Pieper, R. O. (2014) Mutant IDH1-driven cellular transformation increases RAD51-mediated homologous recombination and temozolomide resistance. *Cancer Res.* **74**, 4836–4844

30. Sonoda, Y., Ozawa, T., Aldape, K. D., Deen, D. F., Berger, M. S., and Pieper, R. O. (2011) Akt pathway activation converts anaplastic astrocytoma to glioblastoma multiforme in a human astrocyte model of glioma. *Cancer Res.* **61**, 6674–6678

31. Hoeflich, K. P., Merchant, M., Orr, C., Chan, J., Otter, D. D., Berry, L., Kasman, I., Koeppen, H., Rice, K., Yang, N.-Y., Engst, S., Johnston, S., Friedman, L. S., and Belvin, M. (2012) Intermittent administration of MEK inhibitor GDC-0973 plus PI3K inhibitor GDC-0941 triggers robust apoptosis and tumor growth inhibition. *Cancer Res.* **72**, 210–219

32. Sos, M. L., Levin, R. S., Gordan, J. D., Oses-Prieto, J. A., Webber, J. T., Salt, M., Hann, B., Burlingame, A. L., McCormick, F., Bandyopadhyay, S., and Shokat, K. M. (2014) Oncogene mimicry as a mechanism of primary resistance to BRAF inhibitors. *Cell Rep.* **8**, 1037–1048

33. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57

34. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550

35. Deeb, S. J., Tyanova, S., Hummel, M., Schmidt-Supprian, M., Cox, J., and Mann, M. (2015) Machine learning-based classification of diffuse large B-cell lymphoma patients by their protein expression profiles. *Mol. Cell. Proteomics MCP* **14**, 2947–2960

36. Maile, T. M., Izael-Tomasevic, A., Cheung, T., Guler, G. D., Tindell, C., Masselot, A., Liang, J., Zhao, F., Trojer, P., Classon, M., and Arnott, D. (2015) Mass spectrometric quantification of histone post-translational modifications by a hybrid chemical labeling method. *Mol. Cell. Proteomics* **14**, 1148–1158

37. Vinogradova, M., Gehling, V. S., Gustafson, A., Arora, S., Tindell, C. A., Wilson, C., Williamson, K. E., Guler, G. D., Gangurde, P., Manieri, W., Busby, J., Flynn, E. M., Lan, F., Kim, H., Odate, S., Cochran, A. G., Liu, Y., Wongchenko, M., Yang, Y., Cheung, T. K., Maile, T. M., Lau, T., Costa, M., Hegde, G. V., Jackson, E., Pitti, R., Arnott, D., Bailey, C., Bellon, S., Cummings, R. T., Albrecht, B. K., Harmange, J.-C., Kiefer, J. R., Trojer, P., and Classon, M. (2016) An inhibitor of KDM5 demethylase reduces survival of drug-tolerant cancer cells. *Nat. Chem. Biol.* **12**, 531–538

38. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968

39. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386

40. Wang, W., Chen, J. X., Liao, R., Deng, Q., Zhou, J. J., Huang, S., and Sun, P. (2002) Sequential activation of the MEK-extracellular signal-regulated kinase and MKK3/6-p38 mitogen-activated protein kinase pathways mediates oncogenic ras-induced premature senescence. *Mol. Cell. Biol.* **22**, 3389–3403

41. Payne, D. M., Rossomando, A. J., Martino, P., Erickson, A. K., Her, J. H., Shabanowitz, J., Hunt, D. F., Weber, M. J., and Sturgill, T. W. (1991) Identification of the regulatory phosphorylation sites in pp42/mitogen-activated protein kinase (MAP kinase). *EMBO J.* **10**, 885–892

42. Butch, E. R., and Guan, K.-L. (1996) Characterization of ERK1 Activation site mutants and the effect on recognition by MEK1 and MEK2. *J. Biol. Chem.* **271**, 4230–4235

43. Kim, H. J., and Bar-Sagi, D. (2004) Modulation of signalling by Sprouty: a developing story. *Nat. Rev. Mol. Cell Biol.* **5**, 441–450

44. Ozaki, K., Kadomoto, R., Asato, K., Tanimura, S., Itoh, N., and Kohno, M. (2001) ERK pathway positively regulates the expression of sprouty genes. *Biochem. Biophys. Res. Commun.* **285**, 1084–1088

45. Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnäd, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S., and Mann, M. (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* **3**, ra3

46. Wang, L., Harris, T. E., and Lawrence, J. C. (2008) Regulation of Prolinrich Akt Substrate of 40 kDa (PRAS40) Function by mammalian target of rapamycin complex 1 (mTORC1)-mediated phosphorylation. *J. Biol. Chem.* **283**, 15619–15627

47. Foster, K. G., Acosta-Jaquez, H. A., Romeo, Y., Ekim, B., Soliman, G. A., Carriere, A., Roux, P. P., Ballif, B. A., andingar, D. C. (2010) Regulation of mTOR complex 1 (mTORC1) by raptor Ser863 and multisite phosphorylation. *J. Biol. Chem.* **285**, 80–94

48. Shahbazian, D., Parsyan, A., Petroulakis, E., Topisirovic, I., Martineau, Y., Gibbs, B. F., Svitkin, Y., and Sonenberg, N. (2010) Control of cell survival and proliferation by mammalian eukaryotic initiation factor 4B. *Mol. Cell. Biol.* **30**, 1478–1485

49. Raught, B., andingar, A.-C. (1999) eIF4E activity is regulated at multiple levels. *Int. J. Biochem. Cell Biol.* **31**, 43–57

50. Silvera, D., Formenti, S. C., and Schneider, R. J. (2010) Translational control in cancer. *Nat. Rev. Cancer* **10**, 254–266

51. Ueda, T., Sasaki, M., Elia, A. J., Chio, I. I. C., Hamada, K., Fukunaga, R., and Mak, T. W. (2010) Combined deficiency for MAP kinase-interacting kinase 1 and 2 (Mnk1 and Mnk2) delays tumor development. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13984–13990

52. Matsumoto, F., Fujii, H., Abe, M., Kajino, K., Kobayashi, T., Matsumoto, T., Ikeda, K., and Hino, O. (2006) A novel tumor marker, Niban, is expressed in subsets of thyroid tumors and Hashimoto's thyroiditis. *Hum. Pathol.* **37**, 1592–1600

53. Kannangai, R., Diehl, A. M., Sicklick, J., Rojkind, M., Thomas, D., and Torbenson, M. (2005) Hepatic angiomyolipoma and hepatic stellate cells share a similar gene expression profile. *Hum. Pathol.* **36**, 341–347

54. Ito, S., Fujii, H., Matsumoto, T., Abe, M., Ikeda, K., and Hino, O. (2010) Frequent expression of Niban in head and neck squamous cell carcinoma and squamous dysplasia. *Head Neck* **32**, 96–103

55. Ji, H., Ding, Z., Hawke, D., Xing, D., Jiang, B.-H., Mills, G. B., and Lu, Z. (2012) AKT-dependent phosphorylation of Niban regulates nucleophosmin- and MDM2-mediated p53 stability and cell apoptosis. *EMBO Rep.* **13**, 554–560

56. Chen, S., Evans, H. G., and Evans, D. R. (2011) FAM129B/MINERVA, a novel adherens junction-associated protein, suppresses apoptosis in HeLa cells. *J. Biol. Chem.* **286**, 10201–10209

57. Old, W. M., Shabb, J. B., Houel, S., Wang, H., Coutts, K. L., Yen, C., Litman, E. S., Croy, C. H., Meyer-Arendt, K., Miranda, J. G., Brown, R. A., Witze, E. S., Schweppe, R. E., Resing, K. A., and Ahn, N. G. (2009)

- Functional proteomics identifies targets of phosphorylation by B-Raf signaling in melanoma. *Mol. Cell* **34**, 115–131
58. Hsu, P. P., Kang, S. A., Rameseder, J., Zhang, Y., Ottina, K. A., Lim, D., Peterson, T. R., Choi, Y., Gray, N. S., Yaffe, M. B., Marto, J. A., and Sabatini, D. M. (2011) The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* **332**, 1317–1322
 59. Wang, L., Harris, T. E., and Lawrence, J. C. (2008) Regulation of Proline-rich Akt Substrate of 40 kDa (PRAS40) Function by Mammalian Target of Rapamycin Complex 1 (mTORC1)-mediated Phosphorylation. *J. Biol. Chem.* **283**, 15619–15627
 60. Zenke, F. T., Krendel, M., DerMardirossian, C., King, C. C., Bohl, B. P., and Bokoch, G. M. (2004) p21-activated kinase 1 phosphorylates and regulates 14–3–3 binding to GEF-H1, a microtubule-localized Rho exchange factor. *J. Biol. Chem.* **279**, 18392–18400
 61. Kissil, J. L., Wilker, E. W., Johnson, K. C., Eckman, M. S., Yaffe, M. B., and Jacks, T. (2003) Merlin, the product of the NF2 tumor suppressor gene, is an inhibitor of the p21-activated kinase, Pak1. *Mol. Cell* **12**, 841–849
 62. Kashatus, D. F. (2013) Ral GTPases in tumorigenesis: emerging from the shadows. *Exp. Cell Res.* **319**, 2337–2342
 63. Tong, J., Taylor, P., Peterman, S. M., Prakash, A., and Moran, M. F. (2009) Epidermal growth factor receptor phosphorylation sites Ser991 and Tyr998 are implicated in the regulation of receptor endocytosis and phosphorylations at Ser1039 and Thr1041. *Mol. Cell. Proteomics MCP* **8**, 2131–2144
 64. Assidqi, B. F., Tan, K. Y., Toy, W., Chan, S. P., Chong, P. K., and Lim, Y. P. (2012) EGFR S1166 phosphorylation induced by a combination of EGF and Gefitinib has a potentially negative impact on lung cancer cell growth. *J. Proteome Res.* **11**, 4110–4119
 65. Ying, H., Zheng, H., Scott, K., Wiedemeyer, R., Yan, H., Lim, C., Huang, J., Dhakal, S., Ivanova, E., Xiao, Y., Zhang, H., Hu, J., Stormel, J. M., Lee, M. A., Chen, A.-J., Paik, J.-H., Segatto, O., Brennan, C., Elferink, L. A., Wang, Y. A., Chin, L., and DePinho, R. A. (2010) Mig-6 controls EGFR trafficking and suppresses gliomagenesis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6912–6917
 66. Gioelli, D., Wunderlich, W., Sebolt-Leopold, J., Bekiranov, S., Wulfschlegel, J. D., Petricoin, E. F., Conaway, M., and Weber, M. J. (2011) Compensatory pathways induced by MEK inhibition are effective drug targets for combination therapy against castration-resistant prostate cancer. *Am. Assoc. Cancer Res.* **10**, 1581–1590
 67. Fritsche, L., Neukamm, S. S., Lehmann, R., Kremmer, E., Hennige, A. M., Hunder-Gugel, A., Schenk, M., Haring, H.-U., Schleicher, E. D., and Weigert, C. (2011) Insulin-induced serine phosphorylation of IRS-2 via ERK1/2 and mTOR: studies on the function of Ser675 and Ser907. *Am. J. Physiol. - Endocrinol. Metab.* **300**, E824–E836
 68. Nobusawa, S., Watanabe, T., Kleihues, P., and Ohgaki, H. (2009) IDH1 Mutations as molecular signature and predictive factor of secondary glioblastomas. *Clin. Cancer Res.* **15**, 6002–6007
 69. Yang, H., Ye, D., Guan, K.-L., and Xiong, Y. (2012) IDH1 and IDH2 mutations in tumorigenesis: mechanistic insights and clinical perspectives. *Clin. Cancer Res.* **18**, 5562–5571
 70. Lu, C., Ward, P. S., Kapoor, G. S., Rohle, D., Turcan, S., Abdel-Wahab, O., Edwards, C. R., Khanin, R., Figueroa, M. E., Melnick, A., Wellen, K. E., O'Rourke, D. M., Berger, S. L., Chan, T. A., Levine, R. L., Mellinghoff, I. K., and Thompson, C. B. (2012) IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478
 71. Chantalat, S., Depaux, A., Héry, P., Barral, S., Thuret, J.-Y., Dimitrov, S., and Gérard, M. (2011) Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome Res.* **21**, 1426–1437
 72. Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124
 73. Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddeback, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M. Q., and Zhao, K. (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903
 74. Agaloti, T., Chen, G., and Thanos, D. (2002) Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **111**, 381–392
 75. Fujii, T., Khawaja, M. R., DiNardo, C. D., Atkins, J. T., and Janku, F. Targeting isocitrate dehydrogenase (IDH) in cancer. *Discov. Med.* **21**, 373–380
 76. Cui, D., Ren, J., Shi, J., Feng, L., Wang, K., Zeng, T., Jin, Y., and Gao, L. (2016) R132H mutation in IDH1 gene reduces proliferation, cell survival and invasion of human glioma by downregulating Wnt/ β -catenin signaling. *Int. J. Biochem. Cell Biol.* **73**, 72–81
 77. Kamino, M., Kishida, M., Kibe, T., Ikoma, K., Iijima, M., Hirano, H., Tokudome, M., Chen, L., Koriyama, C., Yamada, K., Arita, K., and Kishida, S. (2011) Wnt-5a signaling is correlated with infiltrative activity in human glioma by inducing cellular migration and MMP-2. *Cancer Sci.* **102**, 540–548
 78. Shojima, K., Sato, A., Hanaki, H., Tsujimoto, I., Nakamura, M., Hattori, K., Sato, Y., Dohi, K., Hirata, M., Yamamoto, H., and Kikuchi, A. (2015) Wnt5a promotes cancer cell invasion and proliferation by receptor-mediated endocytosis-dependent and -independent mechanisms, respectively. *Sci. Rep.* **5**, 8042
 79. Thiele, S., Göbel, A., Rachner, T. D., Fuessel, S., Froehner, M., Muders, M. H., Baretton, G. B., Bernhardt, R., Jakob, F., Jakob, C. C., Bombhäuser, M., Rauner, M., and Hofbauer, L. C. (2015) WNT5A has anti-prostate cancer effects in vitro and reduces tumor growth in the skeleton in vivo. *J. Bone Miner. Res.* **30**, 471–480
 80. Jönsson, M., Dejmeek, J., Bendahl, P.-O., and Andersson, T. (2002) Loss of Wnt-5a protein is associated with early relapse in invasive ductal breast carcinomas. *Cancer Res.* **62**, 409–416
 81. Zhang, Y., Du, J., Zheng, J., Liu, J., Xu, R., Shen, T., Zhu, Y., Chang, J., Wang, H., Zhang, Z., Meng, F., Wang, Y., Chen, Y., Xu, Y., and Gu, L. (2015) EGF-reduced Wnt5a transcription induces epithelial-mesenchymal transition via Arf6-ERK signaling in gastric cancer cells. *Oncotarget* **6**, 7244–7261
 82. Kitai, R., Horita, R., Sato, K., Yoshida, K., Arishima, H., Higashino, Y., Hashimoto, N., Takeuchi, H., Kubota, T., and Kikuta, K.-I. (2010) Nestin expression in astrocytic tumors delineates tumor infiltration. *Brain Tumor Pathol.* **27**, 17–21
 83. Lu, W. J., Lan, F., He, Q., Lee, A., Tang, C. Z., Dong, L., Lan, B., Ma, X., Wu, J. C., and Shen, L. (2011) Inducible expression of stem cell associated intermediate filament nestin reveals an important role in glioblastoma carcinogenesis. *Int. J. Cancer J. Int. Cancer* **128**, 343–351
 84. Forbes, S. A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J. W., Futreal, P. A., and Stratton, M. R. (2008) The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet. Genet. Editor. Board Jonathan Haines AI* CHAPTER, Unit-10.11
 85. Hirokawa, Y., Tikoo, A., Huynh, J., Utermark, T., Hanemann, C. O., Giovannini, M., Xiao, G.-H., Testa, J. R., Wood, J., and Maruta, H. (2004) A clue to the therapy of neurofibromatosis type 2: NF2/merlin is a PAK1 inhibitor. *Cancer J. Sudbury Mass* **10**, 20–26
 86. Trovó-Marqui, A., and Tajara, E. (2006) Neurofibromin: a general outlook. *Clin. Genet.* **70**, 1–13
 87. Johnson, M. R., DeClue, J. E., Felzmann, S., Vass, W. C., Xu, G., White, R., and Lowy, D. R. (1994) Neurofibromin can inhibit Ras-dependent growth by a mechanism independent of its GTPase-accelerating function. *Mol. Cell. Biol.* **14**, 641–645
 88. Gutmann, D. H. (2001) The neurofibromatoses: when less is more. *Hum. Mol. Genet.* **10**, 747–755
 89. Lau, N., Feldkamp, M. M., Roncarì, L., Loehr, A. H., Shannon, P., Gutmann, D. H., and Guha, A. (2000) Loss of neurofibromin is associated with activation of RAS/MAPK and PI3-K/AKT signaling in a neurofibromatosis 1 astrocytoma. *J. Neuropathol. Exp. Neurol.* **59**, 759–767
 90. Petrilli, A. M., and Fernández-Valle, C. (2015) Role of Merlin/NF2 inactivation in tumor biology. *Oncogene*
 91. Britten, C. D. (2013) PI3K and MEK inhibitor combinations: examining the evidence in selected tumor types. *Cancer Chemother. Pharmacol.* **71**, 1395–1409
 92. Aoki, H., Yokoyama, T., Fujiwara, K., Tari, A. M., Sawaya, R., Suki, D., Hess, K. R., Aldape, K. D., Kondo, S., Kumar, R., and Kondo, Y. (2007) Phosphorylated Pak1 level in the cytoplasm correlates with shorter survival time in patients with glioblastoma. *Clin. Cancer Res.* **13**, 6603–6609
 93. Kumar, R., Gururaj, A. E., and Barnes, C. J. (2006) p21-activated kinases in cancer. *Nat. Rev. Cancer* **6**, 459–471

Signaling and Epigenetics in Astrocytoma

94. Kesanakurti, D., Chetty, C., Rajasekhar Maddirela, D., Gujrati, M., and Rao, J. S. (2012) Functional cooperativity by direct interaction between PAK4 and MMP-2 in the regulation of anoikis resistance, migration and invasion in glioma. *Cell Death Dis.* **3**, e445
95. Di, J., Huang, H., Qu, D., Tang, J., Cao, W., Lu, Z., Cheng, Q., Yang, J., Bai, J., Zhang, Y., and Zheng, J. (2015) Rap2B promotes proliferation, migration, and invasion of human breast cancer through calcium-related ERK1/2 signaling pathway. *Sci. Rep.* **5**
96. Gutierrez-Erlandsson, S., Herrero-Vidal, P., Fernandez-Alfara, M., Hernandez-Garcia, S., Gonzalo-Flores, S., Mudarra-Rubio, A., Fresno, M., and Cubelos, B. (2013) R-RAS2 overexpression in tumors of the human central nervous system. *Mol. Cancer* **12**, 127
97. Loilome, W., Joshi, A. D., Rhys C. M. J. ap, Piccirillo, S., Angelo, V. L., Gallia, G. L., and Riggins, G. J. (2009) Glioblastoma cell growth is suppressed by disruption of fibroblast growth factor pathway signaling. *J. Neurooncol.* **94**, 359–366
98. Miao, H., Gale, N. W., Guo, H., Qian, J., Petty, A., Kaspar, J., Murphy, A. J., Valerzuela, D. M., Yancopoulos, G., Hambardzumyan, D., Lathia, J. D., Rich, J. N., Lee, J., and Wang, B. (2015) EphA2 promotes infiltrative invasion of glioma stem cells in vivo through crosstalk with Akt and regulates stem properties. *Oncogene* **34**, 558–567
99. Day, B. W., Stringer, B. W., and Boyd, A. W. (2014) Eph receptors as therapeutic targets in glioblastoma. *Br. J. Cancer* **111**, 1255–1261
100. Lee, P., Murphy, B., Miller, R., Menon, V., Banik, N. L., Giglio, P., Lindhorst, S. M., Varma, A. K., Vandergrift, W. A., Patel, S. J., and Das, A. (2015) Mechanisms and clinical significance of histone deacetylase inhibitors: epigenetic glioblastoma therapy. *Anticancer Res.* **35**, 615–625
101. Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S. C., Falck, J. R., Peng, J., Gu, W., and Zhao, Y. (2007) Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol. Cell. Proteomics* **6**, 812–819
102. Vogelauer, M., Krall, A. S., McBrien, M. A., Li, J.-Y., and Kurdistani, S. K. (2012) Stimulation of histone deacetylase activity by metabolites of intermediary metabolism. *J. Biol. Chem.* **287**, 32006–32016
103. Eliuk, S. M., Maltby, D., Panning, B., and Burlingame, A. L. (2010) High resolution electron transfer dissociation studies of unfractionated intact histones from murine embryonic stem cells using on-line capillary LC separation. *Mol. Cell. Proteomics MCP* **9**, 824–837
104. Young, N. L., DiMaggio, P. A., Plazas-Mayorca, M. D., Baliban, R. C., Floudas, C. A., and Garcia, B. A. (2009) High throughput characterization of combinatorial histone codes. *Mol. Cell. Proteomics* **8**, 2266–2284

Article 4: Phosphoproteome analysis of the MAPK pathway reveals previously undetected feedback mechanisms²¹⁰

Proteomics

Florian Gnad¹, Sophia Doll², Kyung Song³, Matthew P. Stokes⁵, John Moffat⁴, Bonnie Liu³, David Arnott², Jeffrey Wallin³, Lori S. Friedman³, Georgia Hatzivassiliou³ and Marcia Belvin³

¹ Department of Bioinformatics and Computational Biology, Genentech Inc., South San Francisco, CA, USA

² Department of Protein Chemistry, Genentech Inc., South San Francisco, CA, USA

³ Department of Translational Oncology, Genentech Inc., South San Francisco, CA, USA

⁴ Department of Biochemical Pharmacology, Genentech Inc., South San Francisco, CA, USA

⁵ Cell Signaling Technology Inc., Danvers, MA, USA

The Ras-Raf-MEK-ERK pathway is commonly upregulated in cancer. The establishment of targeted MAPK pathway therapies, however, has proven difficult as any cancers treated with MEK inhibitors rapidly develop resistance. To decipher the underlying cause, it is essential to decrypt the phosphorylation network spanned by the MAPK core axis. In this publication, we examined the spectrum of phosphorylation signaling downstream of the key nodes of the Ras-Raf-MEK-ERK pathway. We employed the label-free based PTMScan method from Cell Signaling Technology, and quantified changes in the levels of phosphorylation sites in colon cancer cell line HCT116 cells treated with MEK inhibitor cobimetinib (GDC-0973) or the ERK inhibitor G-824.

DATASET BRIEF

Phosphoproteome analysis of the MAPK pathway reveals previously undetected feedback mechanisms

Florian Gnad¹, Sophia Doll², Kyung Song³, Matthew P. Stokes⁵, John Moffat⁴, Bonnie Liu³, David Arnott², Jeffrey Wallin³, Lori S. Friedman³, Georgia Hatzivassiliou³ and Marcia Belvin³

¹ Department of Bioinformatics and Computational Biology, Genentech Inc., South San Francisco, CA, USA

² Department of Protein Chemistry, Genentech Inc., South San Francisco, CA, USA

³ Department of Translational Oncology, Genentech Inc., South San Francisco, CA, USA

⁴ Department of Biochemical Pharmacology, Genentech Inc., South San Francisco, CA, USA

⁵ Cell Signaling Technology Inc., Danvers, MA, USA

The RAS-RAF-MEK-ERK (MAPK) pathway is prevalently perturbed in cancer. Recent large-scale sequencing initiatives profiled thousands of tumors providing insight into alterations at the DNA and RNA levels. These efforts confirmed that key nodes of the MAPK pathway, in particular KRAS and BRAF, are among the most frequently altered proteins in cancer. The establishment of targeted therapies, however, has proven difficult. To decipher the underlying challenges, it is essential to decrypt the phosphorylation network spanned by the MAPK core axis. Using mass spectrometry we identified 2241 phosphorylation sites on 1020 proteins, and measured their responses to inhibition of MEK or ERK. Multiple phosphorylation patterns revealed previously undetected feedback, as upstream signaling nodes, including receptor kinases, showed changes at the phosphorylation level. We provide a dataset rich in potential therapeutic targets downstream of the MAPK cascade. By integrating TCGA (The Cancer Genome Atlas) data, we highlight some downstream phosphoproteins that are frequently altered in cancer. All MS data have been deposited in the ProteomeXchange with identifier PXD003908 (<http://proteomecentral.proteomexchange.org/dataset/PXD003908>).

Received: February 29, 2016

Revised: May 9, 2016

Accepted: June 5, 2016

Keywords:

Cell Biology / ERK / MAPK / Mass spectrometry / MEK / Phosphoproteomics / Signaling



Additional supporting information may be found in the online version of this article at the publisher's web-site

Many cancer types are characterized by defects in the MAPK (RAS-RAF-MEK-ERK) pathway leading to uncontrolled growth and proliferation. Prevalent perturbation of the MAPK pathway is most evident in thyroid carcinoma, where BRAF, acting as effector kinase of RAS at the apex of the MAPK pathway, showed mutations in 240 (60%) of 399 tumors [1]. The associated missense mutation of BRAF (V600E) is the most frequent hot spot mutation in TCGA (<http://cbioportal.org>) [2, 3] (Supporting Information Fig. 1). Besides BRAF, KRAS is the most frequently altered signal-

ing node of the MAPK pathway (Supporting Information Fig. 2), showing recurrent activating mutations on position 12 (G12D/V/C/A/S). In colorectal carcinoma, for example, KRAS revealed somatic mutations in 94 (42.0%) of 224 tumors [4]. Interestingly, 31 colorectal tumors revealed both PIK3CA and KRAS mutations, suggesting tendency toward co-occurrence ($p < 0.001$). Furthermore, recurrent mutations in NRAS ranked in the top ten of oncogenic mutation hot spots in TCGA.

While it is difficult to target KRAS, numerous small molecule inhibitors targeting RAF or other nodes of the MAPK core axis, including MEK or ERK, have been developed for potential cancer treatments. The effect of targeting a single node of the MAPK phosphorylation network, however, is commonly extenuated by (predominantly unknown)

Correspondence: Dr. Florian Gnad, Department of Bioinformatics and Computational Biology, Genentech Inc., South San Francisco, CA 94080, USA

E-mail: gnad.florian@gen.com

feedback mechanisms [5], accentuating the rationale for simultaneous inhibition of multiple pathway members. For example, the selective MEK inhibitor cobimetinib [6] has been approved for the treatment of melanoma in combination with BRAF inhibition.

Dual inhibition of the MAPK pathway has shown clinical efficacy, and several substrates of the pathway are well characterized. Many nodes and connections of this complex signaling network, however, are still unknown. Information processing within the signaling web is primarily mediated by changes in post-translational protein modifications, in particular by phosphorylation. Studying the dynamics of signal transduction networks, for example, in response to EGF stimulation [7], revealed that different phosphorylation sites of the same protein commonly show distinct kinetics. Hence, an appropriate model of a signaling pathway is actually not a network of proteins but a network of interconnected phosphorylation sites. Modifying specific nodes in this network, by altering the activities of associated protein kinases for example, may significantly perturb the entire signaling cascade and lead to uncontrolled proliferation.

The few studies that investigated MAPK phosphorylation signaling on a large scale demonstrated the ability of mass spectrometry-based proteomics to identify hundreds of phosphorylation sites and quantify perturbed signaling [8]. To further investigate the MAPK-regulated signaling cascade, we quantified 2241 phosphorylation sites from 1020 proteins in response to direct or indirect inhibition of ERK using label-free mass spectrometry.

To this end, we applied the PTMScan® method [9, 10] as employed for the characterization of the PI3K-regulated phosphoproteome [11] (Supporting Information Methods), and measured phosphorylation levels in the HCT116 colon cancer cell line in response to treatment with the MEK inhibitor cobimetinib (GDC-0973) or the ERK inhibitor G-824 [12]. Notably, ERK is the main substrate of MEK, and we therefore postulate that phosphorylation changes observed after both inhibitions are indeed attributed to decreased ERK activity. HCT116 cells contain oncogenic KRAS (G13D) and PIK3CA (H1047R) mutations, representative of the significant co-occurrence of both mutations in colorectal tumors.

Western blot profiling of protein lysates with 17 phospho-motif antibodies (Kinome-View® western blotting service, Cell Signaling Technology [13]) revealed the most apparent changes of phosphorylation after inhibition of ERK activity when using the PKA, AKT, phosphotyrosine, and MAPK phospho-motif antibodies. Following phosphopeptide enrichment of each sample by these antibodies, mass spectrometry analysis uncovered 2241 phosphorylation sites from 1020 proteins (Fig. 1, Supporting Information Tables 1 and 2). The majority (75%) of the 2158 detected peptides were singly phosphorylated, while 16 and 9% showed double or higher multiplicity of phosphorylation, respectively. In most multiply phosphorylated peptides, only one of the identified phosphosites matched with the associated antibody motif, indicating that this was the target of the antibody selection (Fig. 1).

We further found that different phospho-motif enrichments were complementary. Overall, 88.7% of the AKT-enriched phosphopeptides were unique, as were 99.6% of the pY-, and 98.9% of the MAPK-enriched phosphopeptides. In comparison, only 58% of the PKA phospho-motif enriched peptides were unique, because many of the associated peptides were also detected after AKT phospho-motif enrichment. Notably, the AKT phospho-motif antibody yielded the highest identification rate, and the associated sequence motif (R-X-X-p[S/T]) is not as specific as the PKA phospho-motif (R-R-X-X-p[S/T]). Taken together, the distribution of phosphorylated serines, threonines, and tyrosines was 49, 23, and 28%, respectively. Because of the enrichment with the phosphotyrosine antibody, the relative occurrence of tyrosine phosphorylation was an order of magnitude higher than the usually observed frequency of around 2% in metal ion-based enrichment strategies [7, 14].

Mass spectrometry and Western blot profiling made evident that the inhibition of ERK mainly leads to decreased phosphorylation (Fig. 2A and Supporting Information Fig. 3). Phosphorylated peptides that showed 3-fold or higher changes were defined as regulated and manually reviewed by inspecting corresponding extracted ion chromatograms (Supporting Information Table 1). We selected this cutoff, because 85–90% of the phosphopeptide intensity changes fell within the \pm 3-fold change interval. Overall, 221 (10.2%) peptides, corresponding to 135 proteins, showed decreased phosphorylation in response to MEK inhibition. Similarly, 288 (13.3%) peptides, corresponding to 156 proteins, displayed decreased phosphorylation after ERK inhibition. Most proteins with decreased phosphorylation after MEK inhibition were also affected by ERK inhibition (70.4%) (Fig. 2A). Thus, quantitative results were not only consistent between technical replicates (Supporting Information Fig. 4), but also between MEK and ERK inhibition ($p = 2.2 \times 10^{-16}$). This significant correlation confirmed our postulation that MEK inhibition reflects indirect inhibition of ERK activity. Western blot analysis confirmed downregulation of ERK activity, as the associated marker (S6 S235/S236) showed decreased phosphorylation in response to both inhibitors (Supporting Information Fig. 5). For further analyses, we defined phosphopeptides that responded to the inhibition of MEK or ERK, as 'MAPK regulated'. The merged phosphoproteome contained 346 down- and 197 upregulated phosphopeptides, mapping to 194 and 132 proteins, respectively. Similar to PI3K signaling [11], Gene Set Enrichment Analysis (GSEA) revealed that phosphoproteins downregulated in response to inhibition of the MAPK pathway were involved in the cell cycle ($p_{\text{CSEA}} = 7.3 \times 10^{-5}$), associated with signal transduction of multiple growth factors ($p_{\text{CSEA}} < 0.001$), and commonly deregulated in various cancer types ($p_{\text{CSEA}} < 0.001$) (Supporting Information Table 3). Overall, GSEA made evident that the PI3K- and MAPK-regulated phosphoproteomes showed multiple commonalities at the functional level (despite the investigation in different cell lines), consistent with the concerted roles of the two pathways as the cell's primary

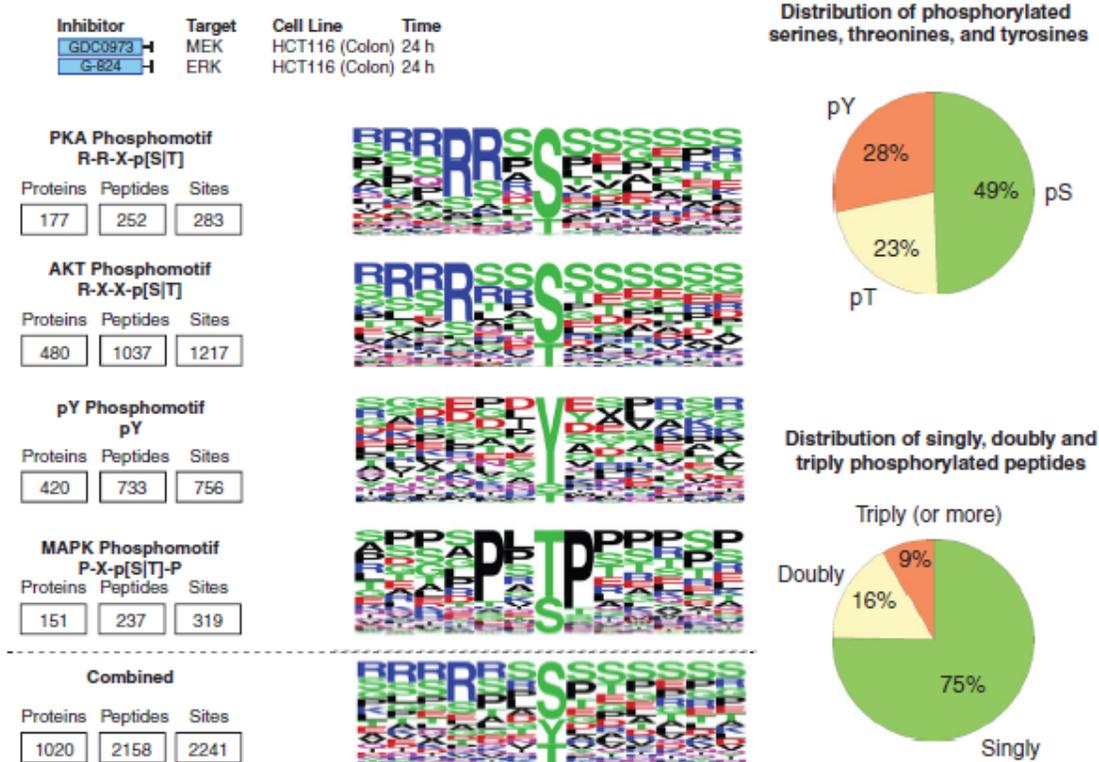


Figure 1. Identification of the MAPK-regulated phosphoproteome. Using label-free mass spectrometry, phosphorylation changes after MEK or ERK inhibition were measured in HCT116 cells. Four phospho-motif antibodies (PKA, AKT, pY, and MAPK) were employed for phosphopeptide enrichment, and the corresponding numbers of identified phosphoproteins, peptides and sites are illustrated for each enrichment experiment (left panel). Associated ‘sequence logos’ (middle panel) reflect the relative frequencies of amino acid residues proximate to detected phosphosites. The relative proportions of identified phosphorylated serines, threonines, and tyrosines, and the distribution of peptides that contained one or more phosphosites are reflected by pie charts (right panel).

mechanisms for controlling cell survival and proliferation. Strikingly, the PI3K- and MAPK-regulated phosphoproteomes also showed a substantial overlap at the substrate level, though the PI3K and MAPK experiments were conducted in different colorectal cell lines. When using a less conservative cutoff (2-fold) to observe similar trends, we found that 163 and 67 of the 435 commonly identified phosphoproteins showed down- or upregulation after both PI3K and MAPK inhibition, respectively. Such redundant phosphorylation patterns indicate crosstalk with other pathways. For example, both pathways are known to interplay with PKA signaling [15, 16], and we found PKA substrate sites, including Ser435 of cell cycle checkpoint regulator ATR [17], to be substantially downregulated after inhibition of both pathways (Supporting Information Table 4). Other common changes, such as decreased phosphorylation of Rictor, TSC2, and EIF4B, can be traced back to the convergence of the two pathways. Notably, phosphorylation changes may occur on different sites, when comparing the two studies. Supporting Information Fig. 6 illustrates phosphorylation changes of selected proteins in response to PI3K or MAPK inhibition.

Strikingly, multiple changes in response to ERK inhibition, increased phosphorylation levels in particular, indicated posttranslational- or transcription-mediated feedback mechanisms. For example, several proteins that act at the apex of the MAPK pathway, including B-Raf, showed regulated phosphorylation after inhibition of MEK or ERK. Interestingly, feedback of B-Raf has been associated with phosphorylation of the C-terminus [18], while we detected regulated phosphorylation on Thr440 and Ser465 (B-Raf contains 766 amino acids) (Supporting Information Table 2). Furthermore, multiple transmembrane receptors, including ERBB3, MET, IGF2R, NRP1, and EPHA1, showed increased phosphorylation after inhibition of the MAPK pathway (Fig. 2C). Phosphorylation of these receptors that control the activity of several pathways may be attributed to feedback-mediated crosstalk. For example, activation of ERBB3 enhances PI3K/AKT signaling and is transcriptionally induced after MAPK inhibition [19], consistent with higher levels of its phosphorylated form (Supporting Information Table 2). Activation of ERBB3 has been further linked to loss of phosphorylation of EGFR and ERBB2 [20], and we found phosphorylation levels of both receptors to be

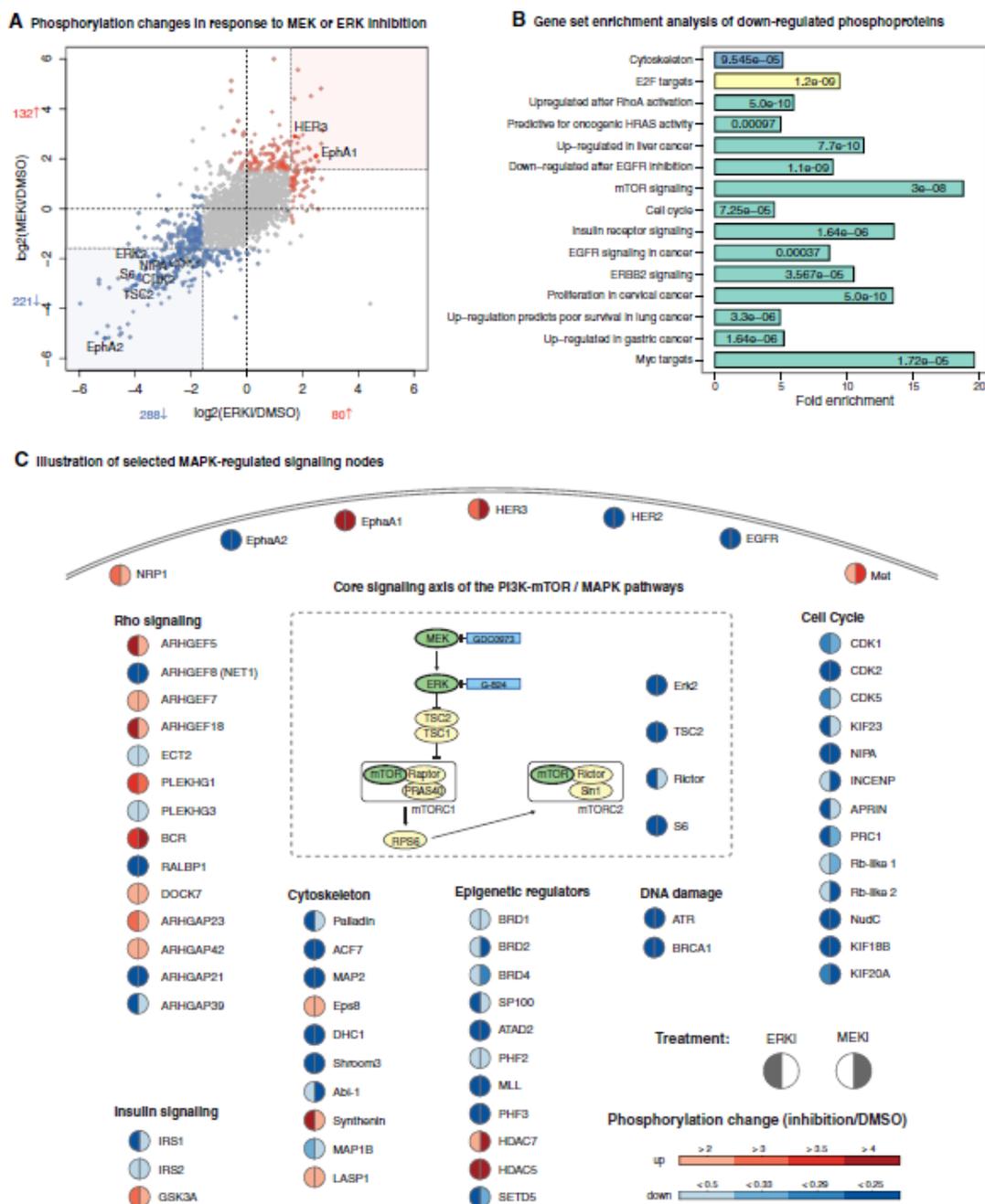


Figure 2. Functional characterization of the MAPK-regulated phosphoproteome. (A) The scatter plot displays phosphorylation changes (log₂-scaled) after MEK (y-axis) or ERK (x-axis) inhibition in HCT116 cells, when compared to DMSO treatment. Each dot reflects intensity changes of a given phosphopeptide, and the corresponding protein names are displayed for a selected set of peptides. Peptides that showed 3-fold or higher increased (red) or decreased (blue) phosphorylation changes are highlighted. Phosphopeptides that showed lower than 3-fold changes or converse responses to MEK or ERK inhibition are presented in gray. (B) Using a 3-fold cutoff to determine phosphopeptides downregulated after MAPK inhibition, gene sets from the Molecular Signatures Database (MSigDB) were tested for enrichment (hypergeometric test). Bars reflect the associated fold enrichments for significantly enriched gene sets, and corresponding p-values are listed. Colors illustrate the correspondence of gene sets to MSigDB collections. (C) Selected phosphoproteins and their responses (red: increased; blue: decreased) to inhibition of MAPK activity are shown.

A Most significantly over-expressed genes encoding identified phosphoproteins

	BLCA	BRCA	COAD	KICH	KIRC	LUAD	LUSC	STAD	THCA	UCEC
TPX2	2.94	3.57	2	2.84	2.28	3.61	4.53	0.43	2.22	4.59
KIF18B	2.88	3.45	1.8	2.51	2.67	3.71	4.61	0.24	2.19	4.57
KIF20A	2.04	3.43	1.72	2.47	2.48	3.17	3.56	0.97	1.81	4.28
GTSE1	2.36	3.04	1.4	2.91	2.33	3.09	3.86	0.04	1.77	4.21
ANLN	0.8	3.29	2.1	2	1.98	3.74	4.1	0.95	2.08	3.63
DTL	1.85	3.06	1.3	1.62	2.13	2.01	2.75	1.14	2	3.25
CDK1	2.37	3.14	1.6	0.4	0.66	2.47	3.41	0.17	1.72	3.18
PRC1	1.48	2.59	1.25	1.39	1.29	2.53	3.55	0.4	1.69	2.75
GPRIN1	2.8	2.81	0.83	0.79	2.04	2.67	2.58	2.31	0.23	2.08
E2F8	1.49	2.59	0.7	5.21	2.89	3.14	3.2	0.39	0.85	3.04
KIF23	1.88	2.6	1.45	0.58	0.71	2.61	3.6	0.52	2.08	2.58
ECT2	1.74	1.77	1.79	0.72	0.07	2.14	3.13	0.19	2.3	2.76
LMNB1	1.57	2.41	0.81	0.34	1.35	1.88	2.02	0.29	1.99	2.98
ATAD2	1.45	1.98	1.64	1.28	0.76	1.83	2.27	0.87	2.06	1.59
MARCKSL1	1.1	1.49	0.76	0.16	0.77	1.87	1.49	0.44	0.77	1.04
SLC7A11	1.14	2.07	2.51	3.89	1.5	2.96	4.39	0.02	0.25	1.48
LMNB2	1	1.19	1.08	0.85	0.22	0.78	1.79	0.07	1.02	1.19
RANBP1	1.06	0.81	0.66	0.38	0.38	0.93	1.89	0.13	0.68	0.58
CDK5	1.06	1.48	0.94	0.51	0.48	0.74	0.82	0.49	0.21	1.42
AZ11	1.23	0.98	0.96	0.98	0.81	0.86	0.9	0.47	0.8	1.14
NASP	0.7	0.58	0.57	0.41	0.74	0.78	0.95	0.1	0.84	0.65
HGS	0.66	0.65	0.51	0.09	0.42	0.62	0.61	0.6	0.11	0.68

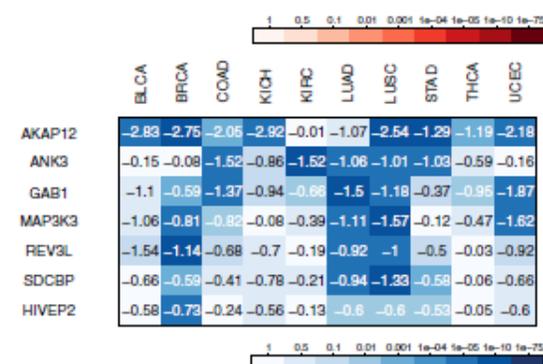
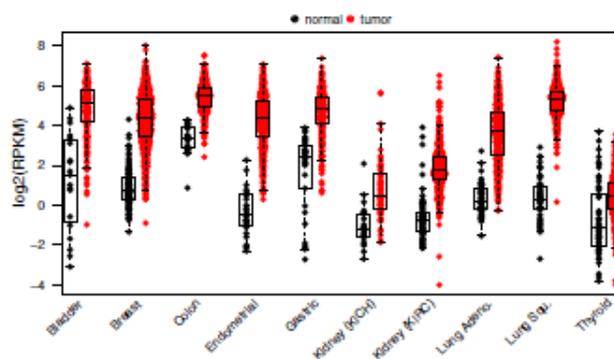
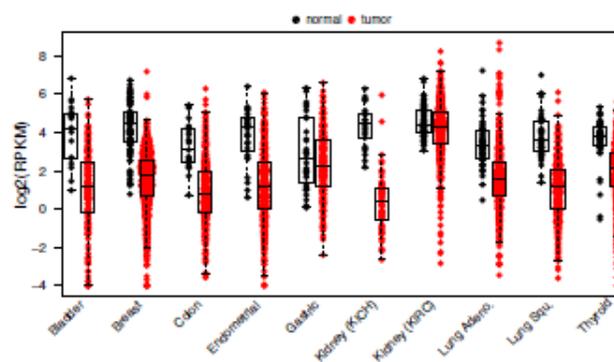

B Expression profile of TPX2

C Expression profile of AKAP12


Figure 3. Gene expression analysis of thousands of TCGA tumors to uncover potential targets among MAPK-regulated phosphoproteins. (A) For each TCGA cancer type (BLCA: bladder, BRCA: breast, COAD: colorectal, KICH: kidney chromophobe, KIRC: kidney renal clear cell, LUAD: lung adenocarcinoma, LUSC: lung squamous, STAD: gastric, THCA: thyroid, UCEC: uterine) differential expression analyses between tumors and healthy tissues were performed, and the most significantly over- or under-expressed genes, which we identified to encode MAPK-regulated phosphoproteins, are listed. Colors reflect corresponding *p*-values of the differential expression analysis, while numbers reflect mean \log_2 -fold changes of expression in tumors compared to healthy tissues. Genes are ranked by combined *p*-values (using Fisher's probability test). Box plots show gene expression levels of TPX2 (B) and AKAP12 (C), displaying significantly increased or decreased expression profiles in cancer, respectively. Each dot reflects the gene expression in a patient (black: healthy tissue, red: tumor), measured in RPKM units.

downregulated after MEK and ERK inhibition. Analogously, increased phosphorylation of EPH receptor A1 (EPHA1) has been associated with feedback loops triggered by AKT inhibition [21]. The converse link between phosphorylation of EPHA1 and inhibition of MEK or ERK, however, has not been previously demonstrated. Interestingly, phosphorylation of EPHA2 (S897), previously linked to AKT activity [22], showed decreased phosphorylation in response to inhibition of MAPK activity, consistent with Western blotting (Supporting Information Fig. 7).

Similarly, phosphorylation of Neuropilin-1 (NRP1) and the hepatocyte growth factor receptor encoded by MET—both enhancers of PI3K signaling—have not been previously linked to MAPK activities. Taken together, we identified

known phosphorylation patterns and discovered previously undetected phosphoregulatory events that can be attributed to feedback-control and crosstalk. These findings confirm the requirement of targeting multiple nodes for efficient inhibition of MAPK activity.

However, dual inhibition to overcome redundancies and feedback are rarely well tolerated. In addition, some drivers are not targetable (e.g. KRAS), which underscores the need to identify downstream nodes apart from the core signaling axes as potential alternative targets. To this end, we analyzed expression data from 4406 TCGA tumors from ten cancer types (Supporting Information Methods). We applied differential expression analyses between tumors and healthy tissues for each cancer type to uncover consistently upregulated

phosphoproteins. Albeit speculative, the associated phosphorylation sites may be also elevated in cancer because of increased expression of their harboring proteins. We found that 22 of those phosphoproteins, which revealed regulation after inhibition of ERK, showed consistently higher gene expression levels across different cancer types (fold change > 0, $p_{\text{fish}} < 0.001$) (Fig. 3A). Spindle assembly factor TPX2 had the most significant expression profile in cancer (Fig. 3B), and showed minimum 3.6-fold decrease of phosphorylation after indirect (MEK) as well direct inhibition of ERK. The functions of the associated phosphosites (T369 and T499), localized in the importin binding domain of TPX2 [23], have not been previously characterized.

Analogously, we further sought to uncover phosphorylation signals that may be frequently reduced in cancer because of decreased expression of their harboring proteins. Seven phosphoproteins showed regulation in our experiments and displayed significantly low expression levels in tumors (fold change < 0, $p_{\text{fish}} < 0.001$) (Fig. 3A). The A-kinase anchor protein 12 (AKAP12) showed the most significant downregulation in cancer (Fig. 3C), and featured 3-fold or higher increases of phosphorylation after ERK inhibition on multiple sites. AKAP12 mediates the subcellular compartmentation of PKA and PKC, but the roles of its phosphorylation sites have not been described to our knowledge.

Finally, we also analyzed mutation data from 4130 TCGA tumors from 15 cancer types (Supporting Information Methods), and looked for somatic point mutations of phosphorylated residues to any amino acid other than serine, threonine, tyrosine, or glutamic acid, implying loss of phosphorylation signal. However, we did not detect any phosphorylation site that was mutated in three or more tumors. Taken together, we found that multiple ERK-regulated phosphoproteins showed striking alteration profiles in cancer. Since their roles in oncogenic signaling are not well characterized, they provide an interesting set of candidates for future studies.

In conclusion, reviewing the mutational landscape of thousands of tumors from TCGA made clear that the perturbation of the MAPK pathway is mainly attributed to mutations of its upstream gatekeepers RAS and RAF. However, RAS cannot be directly targeted and blocking the downstream core kinase axis is challenging because of feedback mechanisms. To shed light on the spectrum of downstream phosphorylation signaling, we blocked ERK activities in a direct (ERK inhibition) and indirect manner (MEK inhibition), inducing both decreased and increased phosphorylation. Notably, for phosphopeptide enrichment prior to mass spectrometry, we selected antibodies that revealed the most substantial regulation of phosphorylation based on western blotting. The application of other phospho-motif antibodies may uncover additional changes, but the fraction of regulated phosphoproteins would be smaller.

Similar to the effect of blocking PI3K activities, many affected phosphoproteins were involved in the cell cycle or proliferation. Interestingly, we observed a large overlap of proteins that showed changes after the inhibition of both the

PI3K and MAPK pathway, even though the experiments were conducted in different colorectal cell lines.

The ERK-regulated phosphoproteome comprised multiple nodes that act upstream of the MAPK core axis, including transmembrane receptors. These phosphorylation patterns made clear that the MAPK signaling cascade is characterized by multiple feedback mechanisms, making targeted therapies difficult. Inhibitor combinations are predicted to overcome redundancies between pathways or feedback mechanisms. However, the establishment of dual inhibition in the clinic has proven difficult. By integrating TCGA data we found many MAPK-regulated phosphoproteins that show consistently decreased or increased expression in cancer. While the impact of perturbed expression levels (in the absence of mutations) in cancer is difficult to decipher (“dark matter”) [24], we postulate that some of the dysregulated MAPK-responders may present potential alternative therapeutic targets.

Taken together we provide a unique resource of MAPK-regulated phosphosites. The identification of multiple previously undetected upstream responding proteins illustrates the difficulty of targeting MAPK. We uploaded the data to the post-translational modification database PHOSIDA (<http://www.phosida.com>) and included additional replicates using different dosages or drug resistant derivatives, which further validate our observations. Furthermore, we deposited raw data in PRIDE proteomeXchange [25] (project accession PXD003908).

We thank Attila Csordas for support of ProteomeXchange. We also thank Gerard Manning for helpful discussions.

F.G. carried out the data analysis and drafted the manuscript. S.D., D.A., and J.W. helped in the interpretation and validation of the data. K.S., J.M., and B.L. prepared the samples for mass spectrometry analysis. M.S. supervised the mass spectrometry experiments. G.H., L.F., and M.B. initiated the project.

The authors have declared the following potential conflicts of interest: Matthew P. Stokes is employed by Cell Signaling Technology, Inc. and all other authors were employed by Genentech, Inc. during the time the study was done.

References

- [1] Agrawal, N., Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 2014, 159, 676–690.
- [2] Cerami, E., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012, 2, 401–404.
- [3] Gao, J., Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 2013, 6, p1.
- [4] Cancer Genome Atlas Network, Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012, 487, 330–337.

- [5] Mendoza, M. C., J. Blenis, E. E., Er, The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem. Sci.* 2011, **36**, 320–328.
- [6] Rice, K. D., Novel carboxamide-based allosteric MEK inhibitors: discovery and optimization efforts toward XL518 (GDC-0973). *ACS Med. Chem. Lett.* 2012, **3**, 416–421.
- [7] Olsen, J. V., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, **127**, 635–648.
- [8] Kirkpatrick, D. S., Phosphoproteomic characterization of DNA damage response in melanoma cells following MEK/PI3K dual inhibition. *Proc. Natl. Acad. Sci. U. S. A.* 2013, **110**, 19426–19431.
- [9] Stokes, M. P., PTMScan direct: identification and quantification of peptides from critical signaling proteins by immunoaffinity enrichment coupled with LC-MS/MS. *Mol. Cell. Proteomics MCP* 2012, **11**, 187–201.
- [10] Stokes, M. P., Quantitative profiling of DNA damage and apoptotic pathways in UV damaged cells using PTMScan direct. *Int. J. Mol. Sci.* 2012, **14**, 286–307.
- [11] Gnad, F., Wallin, J., Edgar, K., Doll, S. et al., Quantitative phosphoproteomic analysis of the PI3K-regulated signaling network. *Proteomics* 2016, **16**, 1992–1997.
- [12] Aronov, A. M., Structure-guided design of potent and selective pyrimidylpyrrole inhibitors of extracellular signal-regulated kinase (ERK) using conformational control. *J. Med. Chem.* 2009, **52**, 6362–6368.
- [13] Zhang, H., Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs. *J. Biol. Chem.* 2002, **277**, 39379–39387.
- [14] Trinidad, J. C., Specht, C. G., Thalhammer, A., Schoepfer, R. et al., Comprehensive identification of phosphorylation sites in postsynaptic density preparations. *Mol. Cell. Proteomics* 2006, **5**, 914–922.
- [15] Chen, Y. J., Interplay of PI3K and cAMP/PKA signaling, and rapamycin-hypersensitivity in TGFβ1 enhancement of FSH-stimulated steroidogenesis in rat ovarian granulosa cells. *J. Endocrinol.* 2007, **192**, 405–419.
- [16] Roberson, E. D., The mitogen-activated protein kinase cascade couples PKA and PKC to cAMP response element binding protein phosphorylation in area CA1 of hippocampus. *J. Neurosci. Off. J. Soc. Neurosci.* 1999, **19**, 4337–4348.
- [17] Jarrett, S. G., PKA-mediated phosphorylation of ATR promotes recruitment of XPA to UV-induced DNA damage. *Mol. Cell.* 2014, **54**, 999–1011.
- [18] Brummer, T., Naegele, H., Reth, M., Misawa, Y., Identification of novel ERK-mediated feedback phosphorylation sites at the C-terminus of B-Raf. *Oncogene* 2003, **22**, 8823–8834.
- [19] Sun, C., Intrinsic resistance to MEK inhibition in KRAS mutant lung and colon cancer through transcriptional induction of ERBB3. *Cell Rep.* 2014, **7**, 86–93.
- [20] Turke, A. B., MEK inhibition leads to PI3K/AKT activation by relieving a negative feedback on ERBB receptors. *Cancer Res.* 2012, **72**, 3228–3237.
- [21] Chandarlapaty, S., AKT inhibition relieves feedback suppression of receptor tyrosine kinase expression and activity. *Cancer Cell.* 2011, **19**, 58–71.
- [22] Miao, H., EphA2 mediates ligand-dependent inhibition and ligand-independent promotion of cell migration and invasion via a reciprocal regulatory loop with Akt. *Cancer Cell.* 2009, **16**, 9–20.
- [23] Schatz, C. A., Importin α-regulated nucleation of microtubules by TPX2. *EMBO J.* 2003, **22**, 2060–2070.
- [24] Vogelstein, B., Cancer genome landscapes. *Science* 2013, **339**, 1546–1558.
- [25] Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A. et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 2014, **32**, 223–226.

Article 5: Quantitative phosphoproteomic analysis of the PI3K-regulated signaling network²¹¹

Proteomics

Florian Gnad¹, Jeffrey Wallin², Kyle Edgar², Sophia Doll³, David Arnott³, Liliane Robillard², Donald S. Kirkpatrick³, Matthew P. Stokes⁴, Ulka Vijapurkar², Georgia Hatzivassiliou², Lori S. Friedman² and Marcia Belvin²

¹ Department of Bioinformatics and Computational Biology, Genentech Inc., South San Francisco, CA, USA ² Department of Translational Oncology, Genentech Inc., South San Francisco, CA, USA

³ Department of Protein Chemistry, Genentech Inc., South San Francisco, CA, USA

⁴ Cell Signaling Technology, Inc., Danvers, MA, USA

Similar to the MAPK pathway, the PI3K pathway is one of the most frequently activated signaling nodes in cancer. Only a few studies have attempted to explore the spectrum of phosphorylation signaling downstream of this kinase cascade. Such investigations, however, are imperative to understand the mechanisms responsible for oncogenic phenotypes. By applying mass spectrometry-based phosphoproteomics, we studied the disturbed phosphoproteome after activation or inhibition of PIK3CA using isogenic knock-ins and a series of inhibitors, including pictilisib (GDC-0941) and tselisib (GDC-0032).

We uncovered phosphorylation changes in a wide variety of proteins involved in cell growth and proliferation, for most of which this was not previously known. Multiple phosphoproteome patterns revealed previously undetected feedback, convergence and crosstalk between cancer pathways, accentuating the rationale for dual pathway inhibition. We provide a dataset rich in potential therapeutic targets downstream of the two most important signaling cascades in cancer.

DATASET BRIEF

Quantitative phosphoproteomic analysis of the PI3K-regulated signaling network

Florian Gnad¹, Jeffrey Wallin², Kyle Edgar², Sophia Doll³, David Arnott³, Liliane Robillard², Donald S. Kirkpatrick³, Matthew P. Stokes⁴, Ulka Vijapurkar², Georgia Hatzivassiliou², Lori S. Friedman² and Marcia Belvin²

¹ Department of Bioinformatics and Computational Biology, Genentech Inc., South San Francisco, CA, USA

² Department of Translational Oncology, Genentech Inc., South San Francisco, CA, USA

³ Department of Protein Chemistry, Genentech Inc., South San Francisco, CA, USA

⁴ Cell Signaling Technology, Inc., Danvers, MA, USA

The PI3K pathway is commonly activated in cancer. Only a few studies have attempted to explore the spectrum of phosphorylation signaling downstream of the PI3K cascade. Such insight, however, is imperative to understand the mechanisms responsible for oncogenic phenotypes. By applying MS-based phosphoproteomics, we mapped 2509 phosphorylation sites on 1096 proteins, and quantified their responses to activation or inhibition of PIK3CA using isogenic knock-in derivatives and a series of targeted inhibitors. We uncovered phosphorylation changes in a wide variety of proteins involved in cell growth and proliferation, many of which have not been previously associated with PI3K signaling. A significant update of the posttranslational modification database PHOSIDA (<http://www.phosida.com>) allows efficient use of the data. All MS data have been deposited in the ProteomeXchange with identifier PXD003899 (<http://proteomecentral.proteomexchange.org/dataset/PXD003899>).

Received: February 29, 2016

Revised: May 8, 2016

Accepted: June 7, 2016

Keywords:

Cell Biology / PHOSIDA / Phosphoproteomics / PI3K / Signaling



Additional supporting information may be found in the online version of this article at the publisher's web-site

The PI3K (PI3K-AKT-mTOR) pathway is commonly deregulated in human cancer. Recent large-scale genomic profiling of tumors through the TCGA consortium made the significance of the pathway very apparent on a global scale. The molecular characterization of thousands of carcinomas uncovered pivotal regulators, including PIK3CA and PTEN, among the most frequently mutated genes in cancer. In colorectal carcinoma, for example, PIK3CA showed somatic mutations in 45 (20.1%) of 224 tumors [1]. Furthermore, oncogenic mutations in the helical (E545K) and kinase (H1047R) domains of PIK3CA are among the most frequent hot spot mutations in TCGA (<http://cbiportal.org>) [2, 3] (Supporting Information Fig. 1). Small molecule inhibitors targeting the PI3K pathway have therefore been intensely pursued as po-

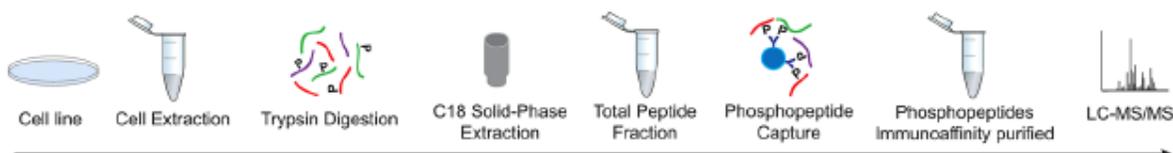
tential treatments for different types of cancers, and several of them have recently advanced into the clinic [4]. Here, we utilized a series of targeted PI3K inhibitors, including pictilisib (GDC-0941) and tasisib (GDC-0032). Pictilisib is a highly specific, ATP-competitive, small-molecule class I pan-PI3K inhibitor [5], and has demonstrated significant antitumor activity in a wide array of cancer models in preclinical studies [6]. Similarly, tasisib is a selective inhibitor of PI3K (highly selective for mutant PIK3CA), and is currently in clinical phase II and III trials for the treatment of lung and breast cancer, respectively [7].

Targeted therapies that block signaling through the PI3K pathway have shown clinical efficacy in several tumor types. The underlying signaling cascades, however, remain poorly understood, as only a few studies have attempted to explore the global spectrum of phosphorylation signaling downstream of the PI3K pathway [8, 9]. Here, we performed a label-free quantitative analysis of PI3K phosphorylation signaling using isogenic knock-in derivatives as well as targeted

Correspondence: Dr. Florian Gnad, Department of Bioinformatics and Computational Biology, Genentech Inc., South San Francisco, CA 94080, USA

E-mail: gnad.florian@gene.com

A Experimental design



B Depth of the identified phosphoproteome

PI3K inhibition in SW48 cells

Inhibitor	Target	Cell Line	Time	Mutants
GDC0941-H	PI3K	SW48 (Colon)	4 h	WT, E545K, H1047R
GDC0932-H	PI3K	SW48 (Colon)	4 h	WT, E545K, H1047R
PI3K-AD-H	PI3K	SW48 (Colon)	4 h	WT, E545K, H1047R

AKT Phosphomotif R-X-X-p[S/T]

Proteins	Peptides	Sites
694	1567	1917



ATM/R Phosphomotif p[S/T]-Q

Proteins	Peptides	Sites
311	500	645



PKC Phosphomotif [K/R]-X-pS-X-[K/R]

Proteins	Peptides	Sites
263	394	460

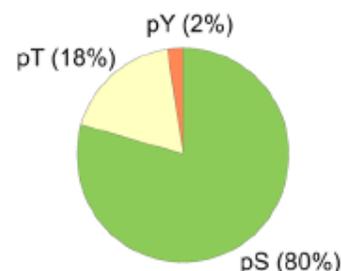


Combined

Proteins	Peptides	Sites
1096	2405	2509



Distribution of phosphorylated serines, threonines, and tyrosines



Distribution of singly, doubly and triply phosphorylated peptides

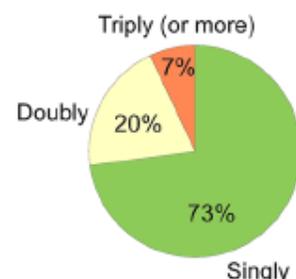


Figure 1. Experimental workflow and depth of the detected phosphoproteome. (A) To quantify phosphorylation changes in response to altered activities of PI3K, the label-free PTMscan[®] method was applied (Cell Signaling Technology). Briefly, for each condition proteins were extracted, digested, and separated from nonpeptide material. Phosphopeptides were isolated in independent immunoaffinity purifications using phosphorylation motif antibodies, and identified and quantified by MS. (B) For each experiment, the number of identified phosphoproteins and corresponding peptides and sites are shown. Phosphopeptides were quantified across conditions within one experiment. The associated “sequence motif logos” illustrate the amino acid occurrences proximal to identified phosphosites. The font height reflects the relative amino acid frequency at a given position. Pie charts show the distributions of phosphorylated serines, threonines, and tyrosine, and singly to multiply phosphorylated peptides.

inhibitors, providing a unique resource for future studies to characterize novel targets and biomarkers of oncogenic PI3K signaling.

To compare phosphorylation levels between cells with wild-type, oncogenic, or inhibited PI3K activities, we applied the label-free PTMscan[®] method [10, 11] as previously described [12] (Fig. 1A, Supporting Information Methods). In brief, views of cellular phosphorylation were obtained for each sample by Western blot profiling of protein lysates with 17 phospho-motif antibodies (Kinome-View[®] Western blotting service, Cell Signaling Technology [13]). AKT, ATM/ATR, and PKC phospho-motif antibodies yielded most

substantial regulation of phosphorylation in response to modulated PI3K activities based on this Western blot screen (Supporting Information Fig. 2), and were therefore used for phosphopeptide enrichment prior to MS analysis. Notably, applied phospho-motif antibodies have a broad immunoreactivity for phosphorylation and are not limited to direct substrates of specific kinases. MS analysis uncovered 2509 phosphorylation sites mapping to 1096 proteins across all experiments (Fig. 1B, Supporting Information Tables 1 and 2). For treatment versus control experiments we applied a threefold cutoff to define up- or downregulation. Regulated phosphorylated peptides were manually reviewed by

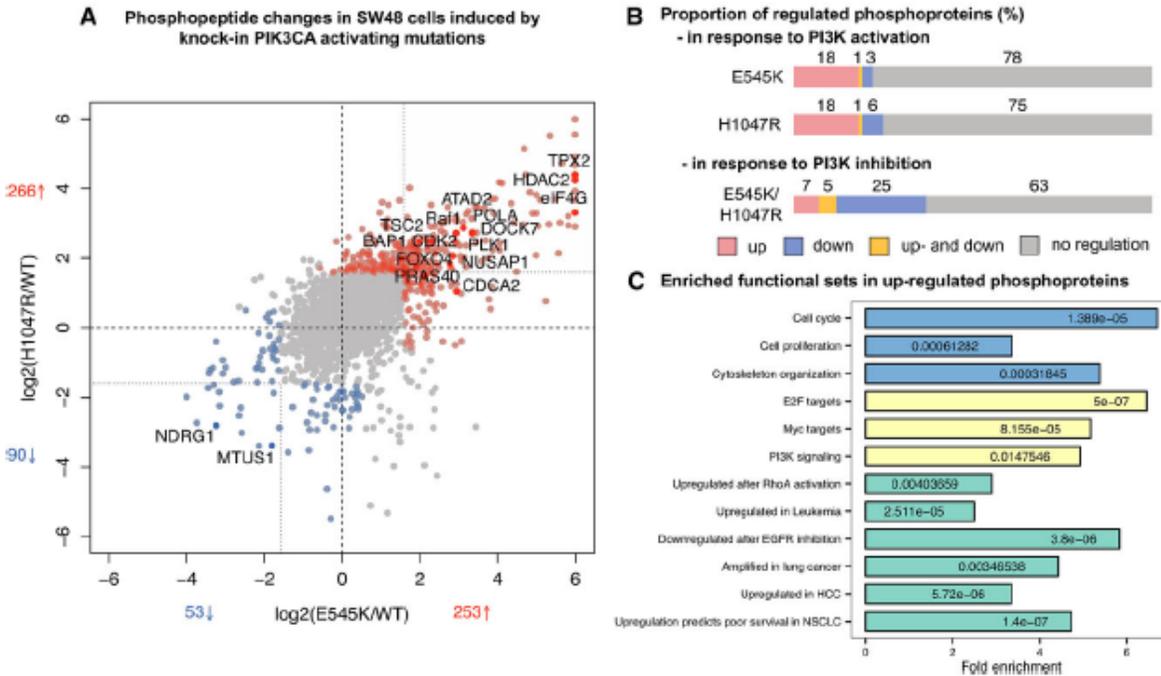


Figure 2. Identification of the PI3K-regulated phosphoproteome. (A) Phosphopeptide abundances were compared between the SW48 cell line and isogenic derivatives carrying oncogenic mutations in either the helical (E545K) or kinase (H1047R) domain of PIK3CA. Red and blue dots represent phosphopeptides that showed threefold up- or downregulation in at least one mutant cell line, respectively (and no opposite response between the mutant cell lines). (B) A threefold cutoff was used to estimate the number of peptides and corresponding proteins that showed phosphorylation changes in response to PIK3CA activation or inhibition by a series of small molecule inhibitors including pictilisib, taselisib, and PI3Ki-A/D. (C) Hypergeometric tests were performed to identify gene sets that were significantly enriched in genes that encode PI3K-regulated phosphoproteins (Supporting Information Table 3). The x-axis reflects the fold enrichment of the number of observed to expected genes that are associated with given categories. Corresponding adjusted p-values are shown within bars. Colors reflect different Molecular Signatures Database gene set collections.

inspecting corresponding extracted ion chromatograms and are highlighted in bold in Supporting Information Table 1.

To analyze aberrant PIK3CA signaling triggered by oncogenic mutations in the helical (E545K) or kinase (H1047R) domain, we quantified changes in the levels of phosphorylation sites between isogenic SW48 PIK3CA wild-type and mutant cells. Oncogenic RAS, RAF, or PTEN mutations are absent in the SW48 wild-type cell line, and the previously identified PIK3CA mutation (G914R) [14] is considered nonactivating. Thus, the SW48 cell line has been classified as PIK3CA wild-type (nononcogenic) [15], serving as an ideal model for the introduction of oncogenic PIK3CA mutations and direct activation of the pathway. MS revealed that the introduction of a single oncogenic amino acid change in PIK3CA led to substantial global elevation of protein phosphorylation levels, as 253 (10.5%) of the 2405 identified phosphopeptides mapping to 192 (17.6%) of 1096 proteins were upregulated in PIK3CA E545K mutant cells (Fig. 2A and B, Supporting Information Tables 1 and 2). Similarly, 266 (11.1%) phosphopeptides from 199 (18.2%) proteins were upregulated in H1047R mutant cells. Overall, phosphoproteomic

changes were concordant between the two PIK3CA mutant cell lines ($p = 2.2 \times 10^{-16}$ based on linear modeling; $p = 8.35 \times 10^{-26}$ based on Chi-Square test on number of regulated phosphopeptides). In addition, quantitative results were consistent between technical replicates in all experiments, as phosphopeptide intensities between duplicates showed strong positive correlations with an average correlation coefficient (R^2) of 0.95 (Supporting Information Fig. 3).

We confirmed the global elevation of phosphorylation by Western blotting (Supporting Information Fig. 2). Furthermore, using phosphosite-specific antibodies we validated elevated levels of PI3K activity markers, including pAKT (S473), pS6 (S235/S236), PRAS40 (T246), pFoxO3a (T32), pFoxO1 (T24), and pFoxO4 (T28), by Western blotting (Supporting Information Fig. 4). Increased phosphorylation of additional PI3K activity markers is identified by MS, including AS160 (T642), BAD (S118), CCT2 (S260), DNAJC5 (S10), EphA2 (S897), FOXO3A (S253), FOXO4 (T32), GSK3A (S21), GSK3B (S9), MST2 (T384), PFKFB2 (S466), RAF1 (S259), and TSC2 (S939).

To better understand the phosphorylation-mediated mechanisms downstream of the PI3K-AKT-mTOR core axis, we performed gene set enrichment analyses (GSEA) of genes encoding proteins with elevated phosphorylation in the PIK3CA mutant cell lines (Supporting Information Methods). We included gene set collections from the Molecular Signatures Database (MSigDB) [16] associated with certain biological processes, pathways, cancer types, or behaviors after treatment (gene set collections “h,” “c2,” “c3,” “c4,” and “c5”). Consistent with the role of the PI3K pathway, activation of PIK3CA significantly affected phosphoproteins involved in cell cycle ($p_{\text{CSEA}} = 1.39 \times 10^{-5}$) or proliferation ($p_{\text{CSEA}} = 6.12 \times 10^{-4}$) (Fig. 2C; Supporting Information Table 3). Cell-cycle regulators CDK2, PLK1, FOXO4, POLA, HEC1, NOLC1, NUSAP1, SUN2, and TPX2, for example, showed significant upregulation of phosphorylation in both mutant cell lines. Only few of the associated phosphosites have been previously functionally characterized. For example, the microtubule-associated protein TPX2 displayed strong (>21-fold) upregulation of phosphorylation in both PIK3CA mutant cell lines on Ser121. This is a target site of Aurora kinase A (AURKA) [17], and shows cell-cycle dependent occupancy [18], but its function is unknown. The drastic response of TPX2 to enhanced PI3K activity points to an “on or off” phosphorylation-mediated mechanism that may be triggered by CDK1-mediated activation of AURKA in both mutant cell lines.

In addition to cell-cycle proteins, GSEA revealed that multiple regulators of the cytoskeleton ($p_{\text{CSEA}} = 3.18 \times 10^{-4}$), including NUSAP1, plectstrin 2, piccolo, Fascin 1, HOOK3, and KIF23, had increased phosphorylation. These cytoskeleton proteins and many other responding phosphoproteins are commonly deregulated across different cancer types ($p_{\text{CSEA}} < 0.01$). The mitotic spindle associated protein NUSAP1, for example, is consistently overexpressed in cancer, and we identified one previously uncharacterized upregulated phosphosite (Ser296). A significant proportion ($p_{\text{CSEA}} = 4.27 \times 10^{-6}$) of genes encoding affected phosphoproteins is even part of a gene set, whose expression levels can be used to predict survival in cancer [19].

Moreover, target genes of transcription factors MYC ($p_{\text{CSEA}} = 8.2 \times 10^{-5}$) and E2F ($p_{\text{CSEA}} = 5.0 \times 10^{-7}$), which are both known components of the downstream PI3K-regulated proliferative machinery, were also significantly overrepresented. Furthermore, several epigenetic regulators, including DNMT3B, EZH1, HDAC2, ATAD2, and SMARCA4, displayed modulated phosphorylation in PIK3CA mutant cells, which may be associated with the reported link between PI3K/AKT signaling and epigenetic alterations [20]. Notably, epigenetic regulators are not represented as a gene set in Molecular Signatures Database. We therefore identified them by comparing our data with a previously defined set of 187 epigenetic regulators [21]. The putative deregulation of the proliferative machinery as well as the modification of some epigenetic regulators at the posttranslational level led us to the hypothesis that some phospho-

rylation changes may result from altered gene expression of the harboring phosphoproteins. To this end, we carried out a microarray analysis of the SW48 cell line and its isogenic counterparts (Supporting Information Methods). Strikingly, we found that the vast majority of phosphorylation changes could not be explained by alteration in gene expression (Supporting Information Fig. 5). These results suggest that perturbed phosphorylation patterns were mainly attributed to changes in phosphorylation instead of altered gene expression.

While mutational activation of PI3K mainly triggered elevation of phosphorylation, we also saw decreased phosphorylation (Supporting Information Tables 1 and 2). The majority of biological processes associated with these phosphoproteins were related to cytoskeleton organization ($p_{\text{CSEA}} = 2.84 \times 10^{-4}$). NDRG1, MTUS1, K8, PGAM1, OXR1, LIM, BAT2D1, SNX9, and C19orf21 had decreased phosphorylation in both mutant cell lines. Previously, microtubule-associated tumor suppressors NDRG1 and MTUS1 also showed decreased phosphorylation after KRAS activation [12]. NDRG1 revealed a hyperphosphorylated region, comprising 11 target phosphosites of SGK1 and GSK3 β [22] between positions 330 and 367, and we validated the decrease of one of these (S330) by Western blotting (Supporting Information Fig. 6). GSK3 β is known to be inactivated by AKT via phosphorylation on Ser9 [23], and we found this site to be upregulated in PIK3CA mutant cells. Accordingly, all NDRG1 phosphosites were decreased ranging from 1.7-fold (Ser367) to 16-fold (Ser330). While this phosphorylation cluster has not been functionally characterized, it may serve as biomarker for oncogenic activity. Similarly, phosphosites of MTUS1 and most of the other mapped phosphosites have either not been previously identified or functionally characterized.

To reverse the effect of PI3K activation and get additional evidence for PI3K-mediated phosphorylation changes, we treated the knock-in cell lines with targeted PI3K inhibitors pictilisib (GDC-0941), taselisib (GDC-0032), or the p110 α/δ selective tool compound inhibitor PI3Ki-A/D [24]. These inhibitor treatments partially reversed the impact of the two oncogenic PI3K nodes on the phosphorylation network. Overall, 168 (45.5%) of the phosphopeptides that had increased phosphorylation after mutational PI3K activation were downregulated (minimum twofold) in PIK3CA E545K or H1047R mutant cells in response to at least one of the three PI3K inhibitors. Most phosphorylation changes observed were consistent between different inhibitor treatments (Supporting Information Fig. 7). Downregulated phosphoproteins were even more significantly enriched for known nodes of PI3K ($p_{\text{CSEA}} = 4.2 \times 10^{-8}$), AKT ($p_{\text{CSEA}} = 8.4 \times 10^{-7}$), and growth factor receptor signaling, including EGFR signaling ($p_{\text{CSEA}} = 3.8 \times 10^{-6}$), validating PI3K activity-driven phosphorylation events versus long-term secondary effects (Supporting Information Fig. 8 and Supporting Information Table 3). We validated the reverse of global phosphorylation and the decrease of associated markers of PI3K activity by Western blotting (Supporting Information Figs. 2 and 4).

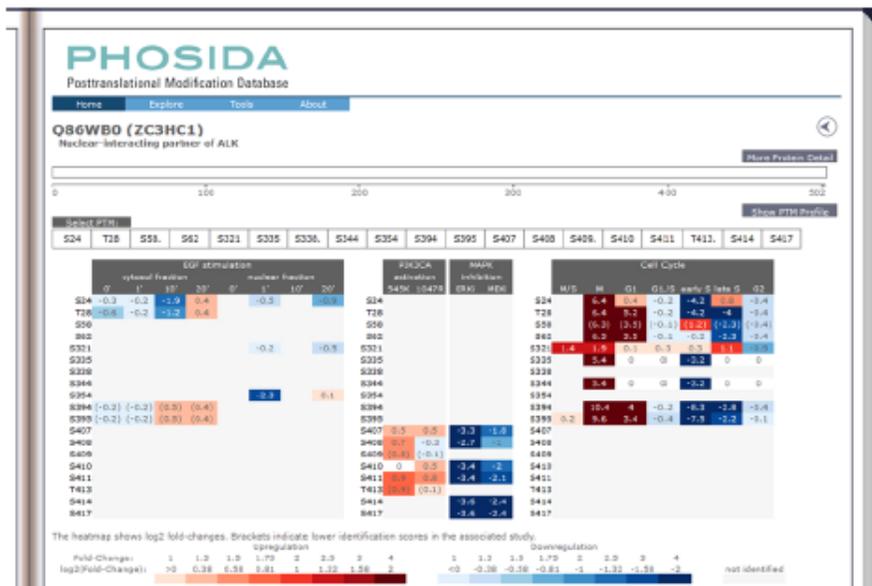


Figure 3. The Posttranslational Modification Database PHOSIDA. By remapping proteomic data to a common database (Uniprot), PI3K-mediated phosphorylation events can be compared with other treatments and conditions, turning PHOSIDA into a unique web environment for quantitative posttranslational modification data.

Phosphoproteins responding to pictilisib, taselisib, or PI3K-A/D present a valuable resource for further investigation into the impact of individual small molecule inhibitors. The provision of phosphoproteomic responses to small molecule compounds that are currently in clinical trials is unique to our knowledge.

To enable efficient use of these data, we updated the posttranslational modification database PHOSIDA (<http://www.phosida.com>) (Fig. 3) [25, 26] and deposited raw data in PRIDE proteomeXchange [27] (project accession PXD003899). To merge our dataset with other phosphoproteomic datasets, we remapped the phosphopeptides of all associated studies to Uniprot version 2016/4. This enables users to compare the response of PI3K-regulated phosphosites with other treatments (e.g., EGF stimulation [28]) or biological processes (e.g. the cell cycle [18, 29]). Notably, PHOSIDA reports quantitative results at the site level, and provides additional information including matching kinase motifs and secondary structure.

In summary, using two isogenic mutant models and a panel of different inhibitors allowed the identification of PI3K-mediated phosphorylation events by different kinase modulations (activation versus inhibition). Western blotting further validated our observations. The impact of the key pathway regulators on the entire signaling cascade became evident by the introduction of a single oncogenic mutation in PI3KCA, which triggered substantial global elevation of phosphorylation in hundreds of signaling nodes. Significant proportions of affected phosphoproteins were involved in the cell cycle or proliferation, consistent with increased proliferation of these mutant cell lines. It is possible that some phosphorylation changes resulted from altered protein expression. However, microarray analysis made clear

that most responding phosphoproteins did not show substantial gene expression changes. Under the assumption that gene expression levels reflect protein intensities to some degree, we therefore conclude that observed phosphorylation changes were mainly caused by differences in phosphorylation site occupancies. This hypothesis is further supported by reversal of boosted phosphorylation levels by short-term inhibitor treatments. Strikingly, only a fraction of the phosphoproteins that responded to inhibition of the core signaling axes have been described as downstream nodes. We postulate that several phosphosites were regulated as a result from the interplay with other pathways, while others are direct targets of the heart of the pathway. Taken together, we believe that our findings help to broaden our understanding of PI3K signaling, and we provide a valuable resource for future studies to characterize novel targets and biomarkers of oncogenic signaling. For efficient use of the data, we updated the posttranslational modification database PHOSIDA.

We thank Gerard Manning for helpful discussions. We also thank Attila Csordas for support of ProteomeXchange. F.G. carried out the data analysis and drafted the manuscript. J.W. and K.E. conducted the PI3K experiment. M.S. supervised the MS experiments. S.D., L.R., U.V., D.K., D.A., and G.H. helped in the interpretation and validation of the data. LF and MB initiated the project.

The authors have declared the following potential conflicts of interest: Matthew P. Stokes is employed by Cell Signaling Technology, Inc. and all other authors were employed by Genentech, Inc. during the time the study was done.

References

- [1] Cancer Genome Atlas Network, Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012, 487, 330–337.
- [2] Cerami, E., The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012, 2, 401–404.
- [3] Gao, J., Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 2013, 6, p11.
- [4] Shuttleworth, S., Progress in the preclinical discovery and clinical development of class I and dual class I/IV phosphoinositide 3-kinase (PI3K) inhibitors. *Curr. Med. Chem.* 2011, 18, 2686–2714.
- [5] Folkes, A. J., The identification of 2-(1H-indazol-4-yl)-6-(4-methanesulfonyl-piperazin-1-ylmethyl)-4-morpholin-4-yl-thieno[3,2-d]pyrimidine (GDC-0941) as a potent, selective, orally bioavailable inhibitor of class I PI3 kinase for the treatment of cancer. *J. Med. Chem.* 2008, 51, 5522–5532.
- [6] Raynaud, F. I., Biological properties of potent inhibitors of class I phosphatidylinositol 3-kinases: from PI-103 through PI-540, PI-620 to the oral agent GDC-0941. *Mol. Cancer Ther.* 2009, 8, 1725–1738.
- [7] Ndubaku, C. O., Discovery of 2-(3-[2-(1-isopropyl-3-methyl-1H-1,2,4-triazol-5-yl)-5,6-dihydrobenzo[f]imidazo[1,2-d][1,4]oxazepin-9-yl]-1H-pyrazol-1-yl)-2-methylpropanamide (GDC-0032): a β -sparing phosphoinositide 3-kinase inhibitor with high unbound exposure and robust in vivo antitumor activity. *J. Med. Chem.* 2013, 56, 4597–4610.
- [8] Hsu, P. P., The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* 2011, 332, 1317–1322.
- [9] Wu, X., Activation of diverse signalling pathways by oncogenic PIK3CA mutations. *Nat. Commun.* 2014, 5, 4961. doi:10.1038/ncomms5961.
- [10] Stokes, M. P., PTMScan Direct: identification and quantification of peptides from critical signaling proteins by immunoaffinity enrichment coupled with LC-MS/MS. *Mol. Cell. Proteomics* 2012, 11, 187–201.
- [11] Stokes, M. P., Quantitative profiling of DNA damage and apoptotic pathways in UV damaged cells using PTMScan Direct. *Int. J. Mol. Sci.* 2012, 14, 286–307.
- [12] Gnad, F., Systems-wide analysis of K-Ras, Cdc42, and PAK4 signaling by quantitative phosphoproteomics. *Mol. Cell. Proteomics* 2013, 12, 2070–2080.
- [13] Zhang, H., Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs. *J. Biol. Chem.* 2002, 277, 39379–39387.
- [14] Barretina, J., The Cancer Cell Line Encyclopedia enables predictive modeling of anticancer drug sensitivity. *Nature* 2012, 483, 603–607.
- [15] Ahmed, D., Epigenetic and genetic features of 24 colon cancer cell lines. *Oncogenesis* 2013, 2, e71.
- [16] Subramanian, A., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* 2005, 102, 15545–15550.
- [17] Nousiainen, M., Sillj , H. H. W., Sauer, G., Nigg, E. A. et al., Phosphoproteome analysis of the human mitotic spindle. *Proc. Natl. Acad. Sci.* 2006, 103, 5391–5396.
- [18] Olsen, J. V., Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* 2010, 3, ra3.
- [19] Shedden, K., Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat. Med.* 2008, 14, 822–827.
- [20] Zuo, T., Epigenetic silencing mediated through activated PI3K/AKT signaling in breast cancer. *Cancer Res.* 2011, 71, 1752–1762.
- [21] Gnad, F., Doll, S., Manning, G., Arnott, D. et al., Bioinformatics analysis of thousands of TCGA tumors to determine the involvement of epigenetic regulators in human cancer. *BMC Genomics* 2015, 16, S5.
- [22] Murray, J. T., Exploitation of KESTREL to identify NDRG family members as physiological substrates for SGK1 and GSK3. *Biochem. J.* 2004, 384, 477–488.
- [23] Dai, F., Human serum and glucocorticoid-inducible kinase-like kinase (SGKL) phosphorylates glycogen synthase kinase 3 beta (GSK-3 β) at serine-9 through direct interaction. *Biochem. Biophys. Res. Commun.* 2002, 293, 1191–1196.
- [24] Edgar, K. A., Isoform-specific phosphoinositide 3-kinase inhibitors exert distinct effects in solid tumors. *Cancer Res.* 2010, 70, 1164–1172.
- [25] Gnad, F., PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* 2007, 8, R250.
- [26] Gnad, F., Gunawardena, J., Mann, M., PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.* 2011, 39, D253–260.
- [27] Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A. et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 2014, 32, 223–226.
- [28] Olsen J. V., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006, 127, 635–648.
- [29] Daub, H., Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol. Cell* 2008, 31, 438–448.

Article 6: Region and cell-type resolved quantitative proteomic map of the human heart and its application to atrial fibrillation²¹²

Nature Communications

Sophia Doll^{1,2}, Martina Dreßen³, Philipp Geyer^{1,2}, Dan Itzhak¹, Christian Braun⁴, Stefanie Doppler³, Florian Meier¹, Marcus-Andre Deutsch^{3,5}, Harald Lahm³, Rüdiger Lange^{3,5}, Markus Krane^{*,3,5}, Matthias Mann^{*,1,2}

¹ Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany

² Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

³ Department of Cardiovascular Surgery, German Heart Center Munich at the Technische Universität München, Munich, Germany

⁴ Forensic Institute, Ludwig-Maximilians-University, Munich, Germany

⁵ DZHK (German Center for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany

The heart is a central organ. It beats approximately 100,000 times a day, precisely controlled by the interplay between electrical and mechanical fields. At the anatomical level, the heart is composed of four cavities, two septa, four valves, and six main vessels, which act in concert to ensure proper filling, ejection, contraction, and overall pump function. At the cellular level, the human heart is composed of four major cell types.

While this manuscript does not focus on cancer proteomics, here we combined state-of-the-art technologies, such as in StageTip sample preparation, ‘loss-less’ nano fractionation, and new MS measurement methods for the first time and apply them to map the human heart proteome. Starting with dissected heart samples from trauma victims, we performed a tour de force proteomics analysis, which resulted in the deepest heart proteome reported to date. Cumulatively, we quantified more than 11,000 proteins in the heart regions and (for the non-cardiomyocytes) in a cell-type

specific manner. Apart from extensive biological and bioinformatic analysis of the data, we demonstrate the usefulness of this resource by applying it to atrial fibrillation, which revealed distinct mitochondrial dysfunction patterns, opening up for a potential future molecular sub-classification. My study was mentioned in the Max Planck Institute press release and reached an altmetric score of 161 (meaning that my article was in the 98th percentile (ranked 2,277th of the 186,438 tracked articles of a similar age in all journals) in only a few weeks after publication in Nature Communications.

ARTICLE

DOI: 10.1038/s41467-017-01747-2

OPEN

Region and cell-type resolved quantitative proteomic map of the human heart

Sophia Doll^{1,2}, Martina Dreßen³, Philipp E. Geyer^{1,2}, Daniel N. Itzhak¹, Christian Braun⁴, Stefanie A. Doppler³, Florian Meier¹, Marcus-Andre Deutsch^{3,5}, Harald Lahm³, Rüdiger Lange^{3,5}, Markus Krane^{3,5} & Matthias Mann^{1,2}

The heart is a central human organ and its diseases are the leading cause of death worldwide, but an in-depth knowledge of the identity and quantity of its constituent proteins is still lacking. Here, we determine the healthy human heart proteome by measuring 16 anatomical regions and three major cardiac cell types by high-resolution mass spectrometry-based proteomics. From low microgram sample amounts, we quantify over 10,700 proteins in this high dynamic range tissue. We combine copy numbers per cell with protein organellar assignments to build a model of the heart proteome at the subcellular level. Analysis of cardiac fibroblasts identifies cellular receptors as potential cell surface markers. Application of our heart map to atrial fibrillation reveals individually distinct mitochondrial dysfunctions. The heart map is available at maxqb.biochem.mpg.de as a resource for future analyses of normal heart function and disease.

¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried 82152, Germany. ²Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen 2200, Denmark. ³Department of Cardiovascular Surgery, German Heart Center Munich at the Technische Universität München, Munich 80636, Germany. ⁴Forensic Institute, Ludwig-Maximilians-University, Munich 80336, Germany. ⁵DZHK (German Center for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich 80802, Germany. Correspondence and requests for materials should be addressed to M.K. (email: krane@dhm.mhn.de) or to M.M. (email: mmann@biochem.mpg.de)

The human heart beats more than two billion times in an average life span and each contraction is precisely controlled by an intricate interplay between electrical signals and mechanical forces. At the anatomical level, it is composed of four cavities, four valves, large arteries, and veins, which act in concert to achieve proper filling, ejection, contraction, and overall pump function. The heart's own blood supply is ensured by two coronary arteries. The human heart is composed of four major cell types—cardiac fibroblasts (CFs), cardiomyocytes, smooth muscle cells (SMCs), and endothelial cells (ECs)¹. Their proportion with respect to number and volume, however, remains controversial. CFs are mesenchymal cells, which produce the extracellular matrix (ECM) scaffold of the heart and are thought to constitute more than half of all heart cells². Cardiomyocytes are estimated to provide about 30% of the total cell number but account for over 70% of the total cardiac mass because of their large volume. In contrast, SMCs, which support the vascular system, and ECs, which form the interior lining of the heart, blood vessels, and cardiac valves, are generally believed to be much less abundant. However, these estimates have been challenged and a recent report claims that ECs are the largest cellular population within the heart³.

In common with other muscle tissues, the heart is dominated by a small number of proteins involved in the contractile apparatus. It employs tissue-specific isoforms such as cardiac troponins, which are used in the diagnosis of myocardial infarction. From a physiological and pathophysiological perspective, it would be desirable to gain deeper insights into the molecular characteristics of the heart at the spatial and cellular levels. In particular, characterization of the healthy state of the human heart would be an important starting point to investigate heart disease, which—despite major progress remains the leading cause of death in developed countries and is rapidly increasing in developing ones⁴.

Relatively little is known about the protein composition of the different regions and cell types of the heart. Previous studies have focused on defining differences between specific regions of the heart, or single-diseased heart compartments⁵, or from nonhuman, or subcellular material^{6,7}. Phosphoproteomic studies have also been applied for the analysis of mammalian hearts^{8,9}. Moreover, other studies use transcriptomic approaches^{10–12}, which is an imperfect proxy for protein levels and their dynamics. However, proteins are the driving forces of the cellular machinery and they are involved in the control of virtually all physiology

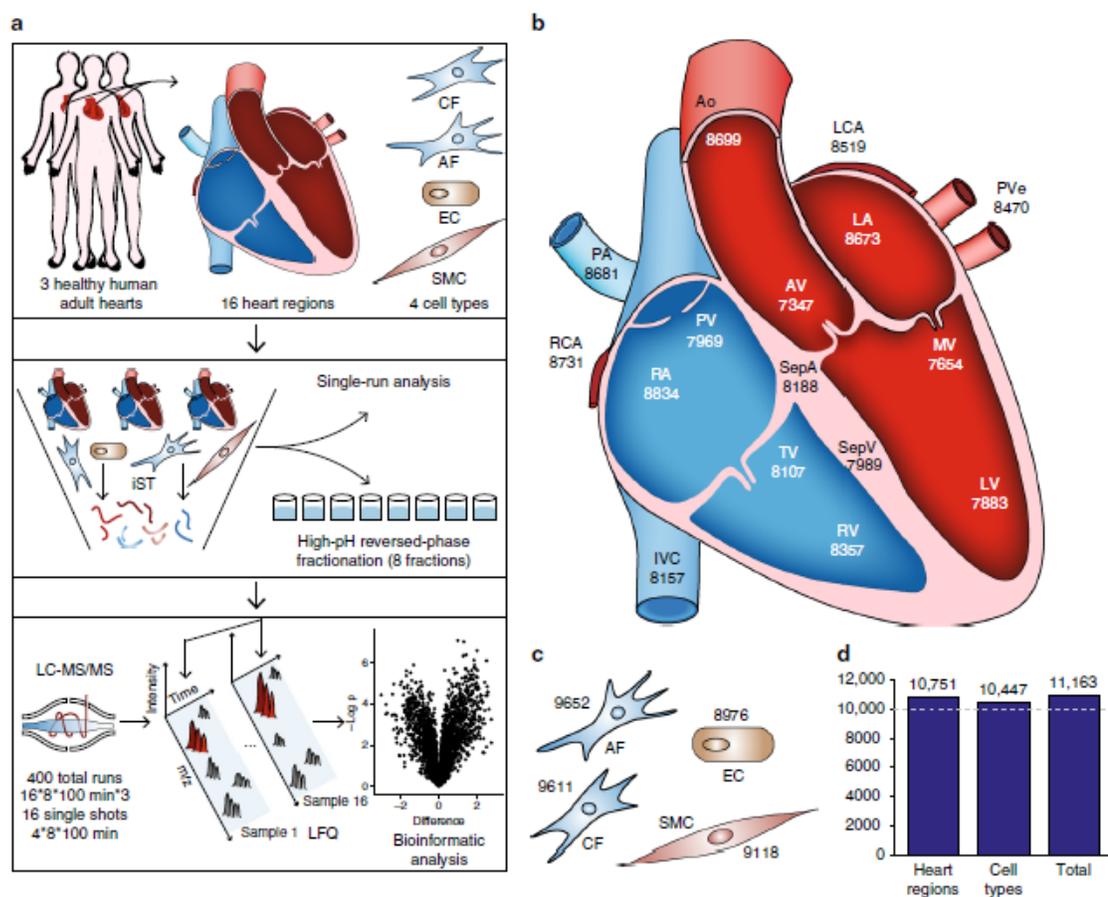


Fig. 1 The quantitative landscape of the human heart proteome. **a** Experimental design, including the source of material (upper panel), in-depth vs. single-run analyses (middle panel), and schematic depiction of the analytical workflow (lower panel). **b** Graphical illustration of the human heart showing the total number of quantified proteins in each region. **c** Quantified proteins in three cardiac cell types and adipose fibroblasts. **d** Bar plot of the total number of quantified proteins in all heart regions, cell types, and the entire data set

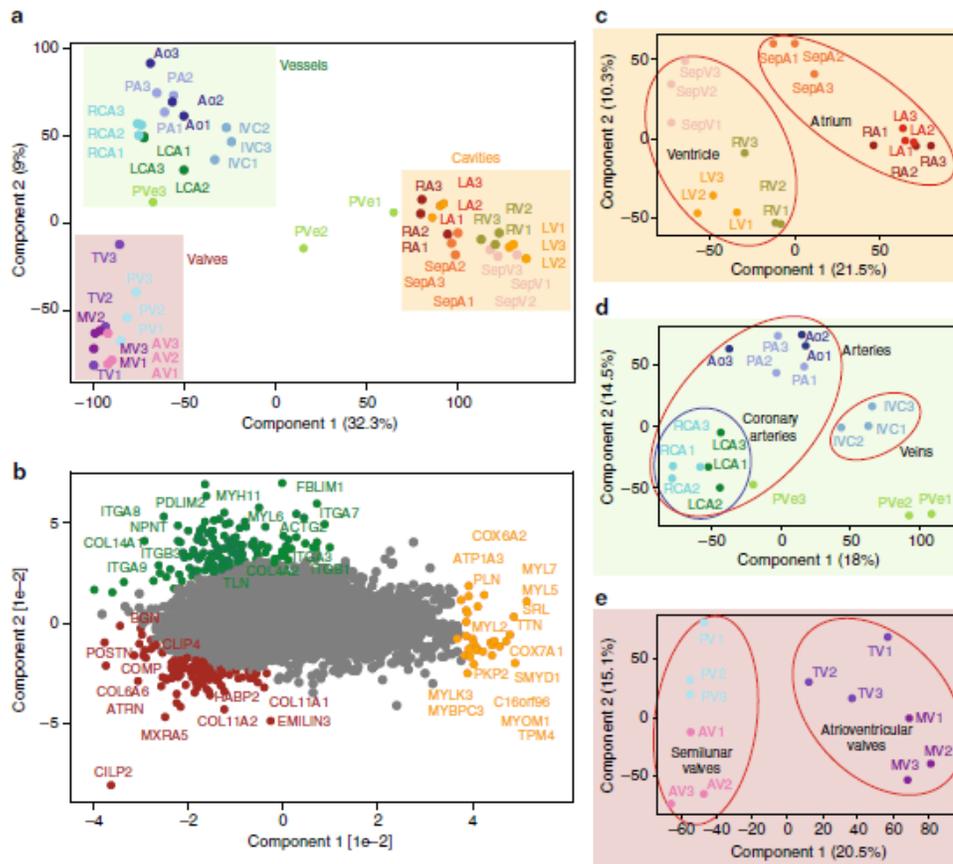


Fig. 2 Principal component analysis (PCA) of the 16 heart regions based on their proteomic expression profiles. **a** The proteomes of the cavities (RA, LA, RV, LV, SepA, and SepV), vessels (Ao, PA, RCA, LCA, IVC, and PVe), and valves (TV, MV, AV, and PV) depicted by replicate number (individuals 1, 2, and 3). The first and second component segregate the heart areas and account for 32.3 and 9% of the variability, respectively. **b** Proteins driving the segregation between the ventricular and atrial part. **c** Cavities segregate into the ventricular and atrial part, **d** vessels into coronary arteries (RCA, LCA) and outgoing vessels (Ao, PA), and **e** valves into ventricular (MV, TV) and semilunar valves (AV, PV)

events. The high dynamic range of the muscle proteome presents a formidable challenge to the comprehensive analysis of the heart at the level of expressed proteins. This is because very abundant proteins make it difficult to detect low abundant regulatory proteins in the same sample. The majority of studies only identified a few thousand proteins, and there is a paucity of studies of the human, nondiseased heart, because of the difficulty in obtaining the relevant tissue.

A global protein expression “footprint” of the healthy heart can be used as a reference library to compare against footprints of malfunctioning hearts in the search for biomarkers, therapeutic targets, or disease signatures. Recent advances in MS-based proteomics technology now allow the identification of very deep proteomes^{13,14}. Our group has already established proteomics maps of the mouse liver and brain^{15,16} and analyzed skeletal muscle in considerable depth and sensitivity^{17,18}. Here, we set out to generate a spatial and cell-type-resolved proteomic map of the healthy human heart. To this end, we measured 16 regions of three human hearts, as well as primary cell types. We employed high-sensitivity sample preparation, peptide fractionation, and an advanced label-free LC-MS workflow to quantify a total of more than 11,000 proteins. Our results establish proteomic differences

between heart regions, suggest functional differences, and pinpoint potential cell-type markers. To illustrate the usefulness of the heart proteomic map, we apply it to define molecular changes in patients suffering from atrial fibrillation (AFib).

Results

Establishing a proteomic map of the human heart. Three adult hearts were obtained from male trauma victims aged 21–47 years with no apparent adverse heart condition (Supplementary Table 1). We selected a total of 16 anatomically defined regions from each heart for MS analysis (Fig. 1a, b): the atrial and ventricular septa (SepA and SepV) separating the atria and ventricles, respectively; the right atrium (RA) and right ventricle (RV) connected via the tricuspid valve (TV); the left atrium (LA) and left ventricle (LV) linked via the mitral valve (MV); the right and left ventricles connected to the pulmonary artery (PA) and aorta (Ao) via the pulmonary and aortic valves (PV and AV); the inferior vena cava (IVC) collecting deoxygenated blood; the pulmonary vein (PVe) carrying oxygenated blood; and the main right and left coronary arteries (RCA and LCA) supplying the heart with oxygen-rich blood. In addition, we isolated CFs, ECs,

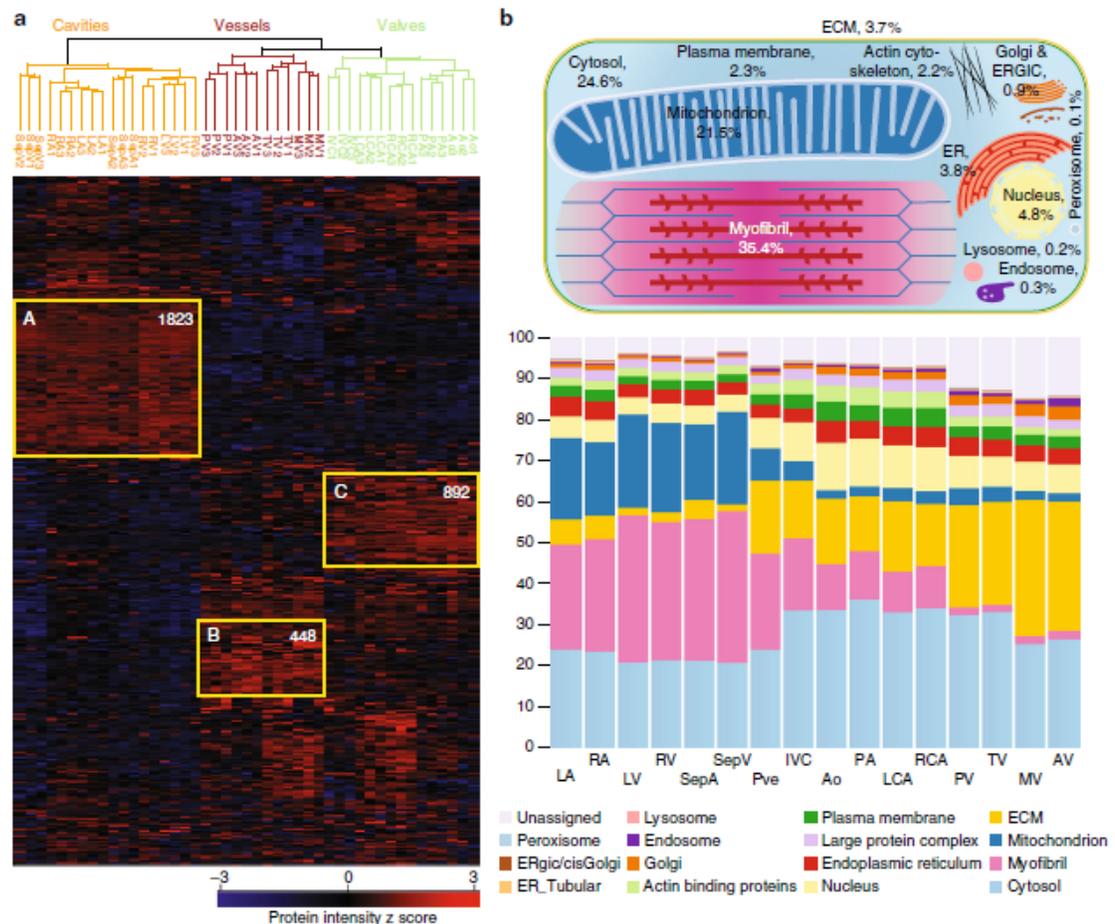


Fig. 3 Proteins differentially expressed across the different heart areas. **a** Heat map of z-scored protein abundances (LFQ intensities) of the differentially expressed proteins (ANOVA, FDR < 0.05) after unsupervised hierarchical clustering reveals proteins significantly upregulated in the cavities, vessels, or valves (highlighted in yellow: A, B, and C). **b** The upper panel shows a schematic of an average heart cavity cell, where organelles are sized according to their contribution to total protein mass. Percentages are taken from the median of all cavities and scaled to account for unassigned proteins. The lower panel shows the contribution of each organelle to cellular protein mass, as a percentage of the total, in each heart region

and SMCs from patients undergoing cardiovascular surgery (Fig. 1c).

After tissue homogenization in liquid nitrogen, we performed all sample preparation using the “in-StageTip (iST) method” (see “Methods” section), reducing sample contamination, loss, preparation time, and increasing quantification accuracy¹⁹. The recently described “loss-less” nano-fractionator enabled efficient fractionation of a total of only 30 µg of peptides into eight fractions, of which a third of each fraction was loaded in the subsequent LC-MS step²⁰. The resulting 400 samples were analyzed with a state-of-the-art label-free workflow on a quadrupole-Orbitrap mass spectrometer (Fig. 1a).

Analysis in the MaxQuant environment using a false-discovery rate (FDR) of less than 1% at the peptide and protein levels²¹, identified a total of 181,814 sequence-unique peptides. These assembled into 11,236 protein groups. Many high-abundance proteins had very high sequence coverage—such as 100% for myosin regulatory light chain 2 (MYL7)—whereas median coverage of all proteins was ~38%. The MaxLFQ algorithm²²

quantified 11,163 proteins, 10,751 in the 16 heart regions, and 10,447 in the noncardiomyocyte cell types, including AFs (Fig. 1d). Proteomic depth was high in all regions, including the four cardiac valves, in which we identified a mean of about 7800 proteins despite the fact that it mainly consists of ECM. To put this number in perspective, reanalysis of the “human draft proteome” heart data²³ with the settings used here revealed that our study identified more than three times as many proteins, most of which were of low abundance (Supplementary Fig. 1). For further analysis, we only considered a subset of 8908 proteins with quantitative values in all biological triplicates of at least one heart region.

Signal intensities for the quantified proteins spanned more than six orders of magnitude, while only six proteins—myosin 7 (MYH7), titin (TTN), cardiac muscle-specific actin (ACTC1), alpha-actinin-2 (ACTN2), and hemoglobin (HBA1 and HBB) represented 25% of the total protein molecules in cavities, with similar values in vessels and valves (Supplementary Fig. 2a–f). A large amount of hemoglobin remained, despite extensive washing

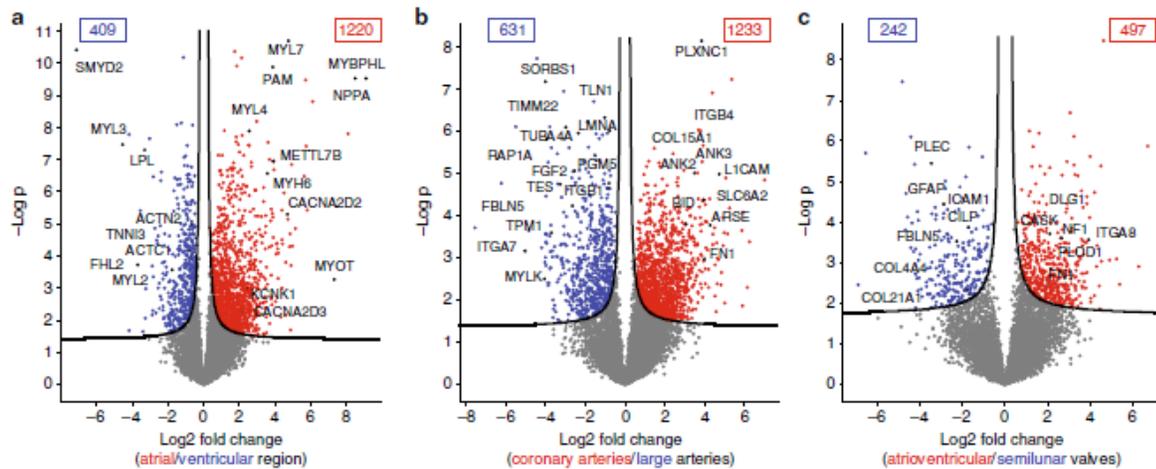


Fig. 4 Proteins differentially expressed in human heart regions. Volcano plots of the p values vs. the \log_2 protein abundance differences between regions, with proteins outside the significance lines colored in red or blue (FDR < 0.05). p values are calculated from the data of three healthy hearts. **a** Ventricular (LV, RV, and SepV) compared to the atrial (LA, RA, and SepA) regions, **b** coronary arteries (LCA, RCA) compared to arteries (Ao, PA), and **c** semilunar (AV, PV) compared to atrioventricular (TV, MV) valves

of the samples with PBS, since hearts of trauma victims cannot be perfused. Due to efficient peptide fractionation, our measurement covered regulatory proteins such as transcription factors GATA4, GATA6, TBX20, TBX3, and TBX5 controlling cardiac-specific gene expression (Supplementary Fig. 3).

To assess quantitative reproducibility, we analyzed several samples in technical triplicates. Pearson correlation coefficients (0.97–0.99) were on par with, or exceeded the values previously achieved in cell line systems²⁴ (Supplementary Fig. 4). Likewise, we observed high correlation values between biological replicates; ranging from 0.83 (PVe) to 0.95 (LA) (Supplementary Fig. 5). As these values incorporate any differences due to postmortem sample treatment, we conclude that our results from three individuals can likely be generalized to the adult male population at large. Raw data and MaxQuant results are provided online and the human cardiac proteome resource is available in our online database MaxQB²⁵ (see below).

Comparative analysis between anatomical areas of the heart.

For an overall assessment of proteomics similarities and differences of the 16 heart regions, we employed principal component analysis (PCA). Cavities, vessels, and valves clearly clustered separately with the samples from different individuals tightly grouped together (Fig. 2a). The only exception was the PVe from patient 1 and 2, which clustered closer to the atrium than the other vessels. This is readily explained by the difficulty of resecting PVe without contamination from LA; thus, PVe samples were excluded from the subsequent analyses.

The segregation of the three groups was mainly driven by MYL7, MYL5, cytochrome c oxidase subunit 7A1 (COX7A1), sarcalumenin (SRL) and TTN (highlighted in red, driving segregation of the cavities), and collagen proteins, such as COL4A2, COL14A1, and integrins, including ITGA7, ITGA8, and ITGB1 (highlighted in blue, segregating vessels), as well as biglycan (BGN), COL11A1, COL11A2, and COL6A6 (highlighted in purple, segregating valves) (Fig. 2b). As these proteins reflect known biological differences between the cardiac cavities and the vessels and valves that are rich in ECM components, they serve as positive controls of our proteomic analysis. Furthermore, the PCA analysis highlighted several interesting candidates, such as

cartilage intermediate layer protein 2 (CILP2) and MXRA5 (valves), nephronectin (NPNT), a functional ligand of ITGA8 and ITGB1 (vessels), and uncharacterized proteins such as C16orf96 (cavities).

Each of the three main clusters exhibited further subgroupings. Heart cavities were divided into atrial (RA, LA, and SepA) and ventricular (RV, LV, and SepV) parts and within them atrial and ventricular septa were separated from atria and ventricles, respectively. Furthermore, there was a moderate but clear distinction of the left and right side of the heart (Fig. 2c). The vessel group subdivided into large arteries (Ao and PA) and large veins (IVC) (Fig. 2d). Within the arteries, the RCA and LCA formed a subcluster, demonstrating differences between coronary and large arteries at the proteomic level. Finally, both atrioventricular valves (MV and TV) clustered together, whereas semilunar valves (AV and PV) formed a separate group (Fig. 2e). The main drivers of the PCA separation are highlighted in Supplementary Fig. 6.

For a functional view of the proteomic differences in the human heart, we performed unsupervised hierarchical clustering of the 6807 proteins with statistically different expression across the heart regions (FDR < 0.05) (Supplementary Data 1). This again clustered individuals in all but one case (RV of one individual), followed by cavities, vessels, and valves with their subdivisions (Fig. 3a). The heat map shows one major cluster of highly and coexpressed proteins for each of the three anatomical areas. Gene ontology and GSEA²⁶ revealed that proteins in cluster A (high expression in the cavities) were enriched ($p < 10^{-12}$) in terms of cardiac muscle contraction, Z disc, and sarcomere organization compared to clusters B and C (high expression in the vessels and valves, respectively). The terms mitochondria and respiratory electron transport chain were also enriched ($p < 10^{-55}$) in this cluster, concordant with the large number of mitochondria to ensure sufficient amounts of ATP for continuous muscle contraction (Supplementary Data 2). Thus, our proteomic data provide a global protein expression basis for the functional specialization of cardiac muscle tissue.

To provide insights into the organelle sizes in the heart proteome at a quantitative level, we used the proteomic ruler approach to estimate copy numbers per cell²⁷ together with

subcellular localization annotations^{28,29}. We calculated an approximate total protein content of 1 ng per diploid nucleus and found that a heart cavity muscle cell has an approximate volume of 5 pL per nucleus (note that about 30% of all cardiomyocytes have two or more nuclei²). These values are roughly double that of mouse heart muscle cells³⁰. Estimated protein copy numbers per diploid nucleus and protein concentration³¹ across our samples ranged from ~10 to 10⁹ and <0.1 nM to 200 μM, respectively (Supplementary Data 3). We found that mitochondria constituted 21% of protein mass in the cavities (Fig. 3b, “Methods” section, and Supplementary Data 3). This compares to 7% of mitochondrial protein mass in HeLa cells²⁸ and 3% in valves and vessels, demonstrating the immense aerobic respiration in cardiac muscle cells localized in the atrial and ventricular part of the heart at a quantitative level.

To further mine our quantitative and in-depth proteome resource, we used volcano plots³² to compare expression differences within the three anatomical areas of the heart. We specifically focused on proteins that were in the top 75% in abundance and only identified in one of the two regions that we compared (Supplementary Data 4).

The atrial vs. ventricular proteome. Although they are both heart muscles, the main role of atria is to collect and transfer pulmonary and systemic blood, whereas ventricles need to pump the blood throughout the entire body. Consistent with these different functions, we found drastic differences in their proteomes, with 1220 (13.7%) proteins showing significantly higher expression in the atria and 409 (4.6%) displaying higher expression in the ventricles (Fig. 4a). As expected, mitochondrial proteins were more abundant in ventricles ($p < 10^{-150}$ by GSEA analysis, Supplementary Data 5). Cardiomyocytes use fatty acids as their main energy source and ventricular myocytes have higher energy demands due to the greater force of contraction. Accordingly, lipid metabolic processes were overrepresented in the ventricular region ($p < 7 \times 10^{-26}$), exemplified by the ~tenfold increased expression of lipoprotein lipase (LPL). Likewise, it was enriched in muscle contraction ($p < 9 \times 10^{-27}$), due to increased cardiomyocyte size, as estimated by the proteomic ruler (“Methods” section). The known markers for ventricles vs. atria, such as MYL2, MYL3, and LPL were clearly recovered as such, and our data set contains many additional ones, including the lysine methyltransferase SMYD2, which is thought to have a role in myocyte function³³. Several interesting candidates were only identified in the ventricles but not in the atria, such as the probable histone demethylase JMJD1C, the ubiquitin ligase TRIM38, the tumor suppressor RASSF8, and the uncharacterized KIAA1324L protein. These proteins have not been associated with ventricular functions before and suggest starting points for exploring their role in heart physiology. Proteins previously reported as atrium specific^{34–36}, including myosin 6 (MYH6), peptidyl-glycine alpha-amidating monooxygenase (PAM), and natriuretic peptides A (NPPA) displayed ten to several hundred-fold higher abundance in the atrium. Interestingly, these proteins are only highly expressed in the ventricular regions under pathological conditions—for example, cardiac hypertrophy leads to elevated PAM levels in ventricles³⁷. We found potassium ion channels predominantly in the atrial part, such as KCNK1, which induces background currents³⁸, and calcium-dependent ion channels, including CACNA2D2 and 3 (>fourfold), as well as gap junction GJA5 (>ninefold), reflecting the presence of the sinus and AV node, which generate the electrical impulse for heart contraction. Others, such as CACNA1C, which play an important role in excitation–contraction coupling in the heart, were equally

expressed in the atria and ventricles (see also Supplementary Fig. 7).

The high protein sequence coverage encouraged us to investigate isoform-specific expression patterns of sarcomeric proteins (Supplementary Data 6). These isoforms are of particular importance because their altered expression has been associated with diverse cardiac dysfunctions³⁹ and because troponins are routinely used biomarkers for myocardial infarction⁴⁰. The myosin isoform families MYH6, MYH7, MYL2-7, and MYL8 localized highly specifically to atria or ventricles, confirming the regional specificity of our data set despite high sequence identity, for instance, over 80% between MYHs⁴¹.

Myocardial infarction remains one of the largest causes of death and although rapid ELISA tests against cardiac troponins TNNT2 and TNNI3 play a crucial role, further improvements in diagnosis would be of great clinical benefit^{42,43}. Here, we found that TNNT2 was more abundant (>twofold) than TNNI3 in all cavities and that their expression largely correlated across regions (Pearson correlation: 0.99). Interestingly, the cardiac isoform myosin-binding protein C3 (MYBPC3), which participates in stabilizing sarcomere structures, displayed a strikingly similar protein expression profile to cardiac troponins and was similarly abundant to TNNT2 (about 1.5-fold higher abundant) (Supplementary Fig. 8). Moreover, it can be detected by antibody- and MS-based approaches in human plasma after myocardial injury^{44,45}, showing that it can be used as an useful additional parameter to monitor myocardial infarct.

Finally, while the RV and LV did not show any significantly altered protein expression, the protein myotilin, which stabilizes thin filaments during muscle contraction was much more abundant in LA compared to RA (>100-fold; Supplementary Fig. 9).

Large vs. small arteries proteomes. Coronary arteries are relatively small as they supply the heart itself with blood; however, their malfunction is responsible for the high prevalence of coronary artery diseases (CADs), affecting more than 16 million individuals in the United States alone⁴⁶. Overall, 1233 (13.8%) of quantified proteins were significantly more abundant in coronary arteries (RCA and LCA) and 631 (7.1%) in large arteries (Ao and PA) (Fig. 4b). Proteins involved in mitochondrial functions, collagen proteins, and integrins, such as COL15A1 and ITGB4 were highly enriched in the coronary vs. large arteries. Fibronectin (FN1) showed a 12-fold increase and has been previously associated with CADs⁴⁷, although it would need to be investigated whether or not it is a specific marker for CAD patients. Arylsulfatase E (ARSE) is a constituent of artery walls where it regulates the composition of cartilage, and we found it to be >20-fold more abundant in coronary arteries.

Large arteries showed significant ($p < 3 \times 10^{-42}$) enrichment for cytoskeleton proteins and proteins involved in cell junction, consistent with the higher structural demands on them. For instance, fibulin 5 (FBLN5) was six- to tenfold higher expressed in PA and Ao than the other heart compartments (Supplementary Fig. 10). It is required to form the elastic lamina, has a protective role against vascular injury, and its downregulation has been associated with aortic aneurysm⁴⁸. Interestingly, fibroblast growth factor 2 (FGF2) and ras-related protein (RAP1A), described as a key regulator of FGF2-induced angiogenesis⁴⁹, were three- to sixfold more abundant in the large arteries. As we had not identified FGF2 in a deep plasma proteome previously⁵⁰, it is unlikely to derive from blood remnants and may instead represent an ECM-bound form⁵¹. A total of 92 medium- to high-abundance proteins were exclusively quantified in the large arteries (Supplementary Data 4). This included the key focal

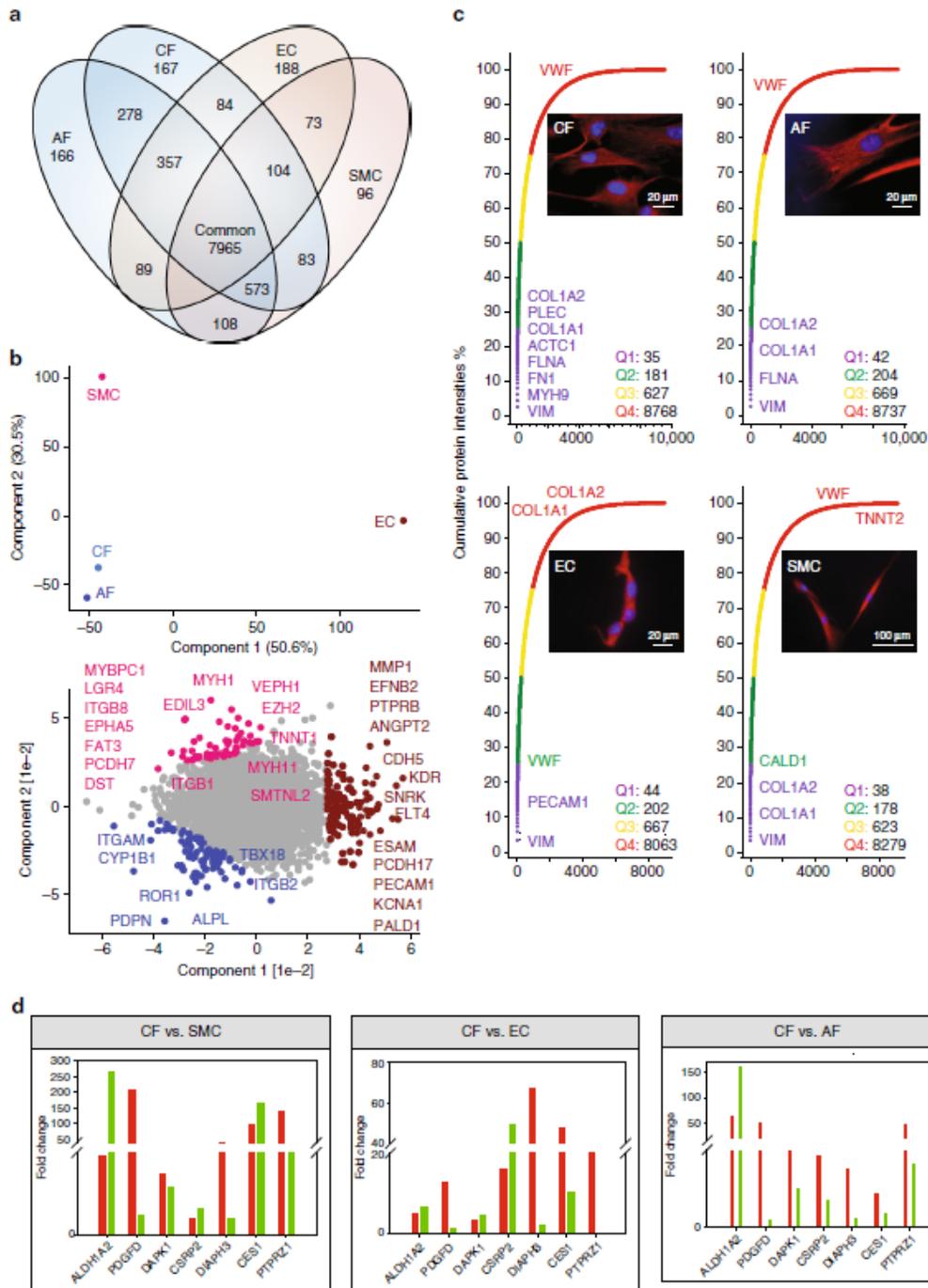


Fig. 5 Comparative analysis of the cell-type proteomes. **a** Commonly and exclusively quantified proteins in three cardiac cell types and adipose fibroblasts. **b** PCA comparing the four cell types based on component 1 and 2, which accounted for 50.6 and 30.5% of the variability, respectively. **c** Cumulative protein abundances for each cell type and total number of proteins constituting the quantiles (Q1-Q4). The corresponding cell types are illustrated with immunofluorescence pictures at $\times 100$ magnification. **d** RT-qPCR (green) and proteomic (red) fold-changes of the indicated genes in CF compared to all other cell types

adhesion protein SORBS1, COL26A1, and MYLK and MYH11, which are both involved in smooth muscle contraction, reflecting the higher proportion of SMCs in the large arteries' wall compared to coronary arteries.

The atrioventricular vs. semilunar valves proteome. The atrioventricular valves separating atria from ventricles (TV and MV) are morphologically quite different from the semilunar valves (AV and PV) preventing backflow of blood from aorta or pulmonary artery to ventricles. We found that only 497 (5.6%) proteins were significantly more abundant in the atrioventricular valves and 242 (2.7%) were more abundant in the semilunar valves (Fig. 4c). Valves are composed of highly organized ECM proteins and changes in its composition and their possible release are expected during valve deterioration, leading to dysfunction and failing heart valves^{52–54}. Although the overall changes were limited, we found that among the ECM proteins, plectin isoform 3 and GFAP were highly expressed in the semilunar valves, whereas peripheral plasma membrane CASK, collagen enzyme P4HA1, integrin ITGA8, and neurofibromin (NF1) were significantly higher in atrioventricular valves.

Cell-type-resolved proteome of the human heart. Our region-resolved proteome achieved great depths, but as we used homogenized tissue, we do not have cell-type-specific information about the origin of the proteins. To address this, we isolated CFs, ECs, and SMCs from tissue samples harvested during cardiac surgery (Supplementary Table 2, “Methods” section). AFs were included to help in defining the CF-specific proteome. Cardiomyocytes were not investigated because of the impossibility to culture these cells from surgical biopsies. We achieved highly purified cell populations with values for CFs, AFs, ECs, and SMCs of 96%, 97%, 96%, and 92%, respectively (Supplementary Fig. 11a–d). Of a total of 11,236 different proteins, 7965 were identified in all four cell types, indicating that the majority of the cardiac cell proteome is expressed in its major cell types (Fig. 5a). We found high correlation (0.92) in protein expression between the fibroblast cell types (CF and AF), whereas SMC and EC were somewhat less related (0.81) and this is also reflected in the PCA (Supplementary Fig. 12 and Fig. 5b). On average, the 40 most abundant proteins accounted for 25% of the total protein mass in all four cell types (Fig. 5c). Consistent with the mesodermal origin of these cell types, vimentin (VIM) was the most abundant protein, accounting for 3% of the total protein mass. In conjunction with LARP6, VIM stabilizes type I collagen mRNAs, leading to upregulation of the collagens COL1A1 and COL1A2⁵⁵. We found the collagens in the top quartile (Q1) of expression in CFs, AFs, and SMCs, whereas they were among the least abundant proteins (Q4) in ECs. Cell-type-enriched proteins—those with at least twofold higher expression in one of the cell types compared to all others—are listed in Supplementary Data 7.

CF-enriched cell surface markers. Over the past decade, CFs have been shown to hold great promise as a potential target population for cardiac regenerative therapies^{1,56,57}. Selection and targeting of CFs, however, remains challenging and currently relies on unspecific CF markers, including VIM, discoidin domain-containing receptor 2 (DDR2), periostin (POSTN), protein S100A4, ACTA2, platelet-derived growth factor receptors PDGFR α and β , T-box transcription factor TBX18, and the THY1 membrane glycoprotein⁵⁸. Among these, only PDGFR β , S100A4, and ACTA2 showed at least twofold enrichment in CFs compared to ECs and SMCs, whereas all other currently employed CF markers were not enriched in CFs. Remarkably, compared to another fibroblast cell type (AFs), none of these markers were

even twofold enriched. GATA4 and TBX20 have been reported as specific CF markers⁵⁹. GATA4 was indeed only identified and quantified in CFs; however, TBX20 was fourfold more abundant in SMCs compared to CFs.

Globally, 609 (5.8%) proteins were specifically enriched in CFs compared to AFs, ECs, and SMCs. These encompassed 25 cell membrane receptors (Supplementary Data 8). The presence of the tyrosine kinase ROR1 in CFs has not been reported before but we found it to be 200-fold more abundant than in the other cardiac cell types. Activin receptor ACVR1 is required for normal heart development⁶⁰ and it was also one of the most highest expressed proteins in CFs. Natriuretic peptide receptor NPR3 has a central role in vasodilatation, is known to be present in CFs⁶¹, and our data showed sixfold higher expression in CFs. The drug target hepatocyte growth factor receptor MET, was also increased in CFs (fourfold higher than ECs and SMCs). BDKRB2, the receptor for bradykinin plays a pivotal role in the cardiovascular system by regulating blood pressure. Interestingly, we exclusively identified it in fibroblasts, with 11-fold higher expression in CFs compared to AFs. Likewise, protein levels of the cell membrane phosphatase PTPRZ1 were more than 40-fold higher in CFs compared to all other investigated cell types, a finding supported by qPCR. Interestingly, the direction of expression changes was concordant between the mRNA and protein levels. Importantly, however, for this and six other genes, the fold-changes indicated by qPCR were not predictive of the actual protein-level changes (Fig. 5d). Our data provide a catalog of CF-enriched marker candidates that hold promise for better definition and targeting of human CFs.

EC and SMC proteome. Two further major cell types of the human heart include ECs, which form the inner lining of heart blood vessels and SMCs, the major constituents of the heart vasculature. ECs contain numerous storage granules filled with von Willebrand factor (VWF), which is involved in hemostasis. VWF was one of the most abundant proteins in ECs, whereas it was among the least abundant in CFs, AFs, and SMCs (Fig. 5c). The platelet endothelial cell adhesion molecule (PECAM1/CD31) was among the most abundant (Q1) proteins in ECs. Furthermore, EC-specific proteins ESAM (>22-fold) and ESM1 (exclusively) were overrepresented. Proteins involved in blood vessel morphogenesis such as the VEGF receptors FLT1, FLT4, KDR, as well as EPHB4 and its ligand Ephrin B2 showed 3- to 250-fold higher expressions compared to the other cell types. Comparing our data with single-cell transcriptomic data⁶² revealed that all genes identified as EC specific compared to CF (except CAV2, which only displayed a moderate increase) also showed at least several 10-fold upregulation in ECs at the proteomic level (Supplementary Data 8). SMCs contain the same muscle-contracting proteins as cardiac cells but do not have troponin. The low levels of cardiac TNNT2 detected (>100-fold less than in cavities), are likely due to the 8% impurity of isolated SMCs. In place of troponins, caldesmon (CALD1), which was among the highest abundant proteins in SMCs, blocks the myosin-head binding site on actin filaments⁶³. Proteins segregating the SMC group (Fig. 5b, highlighted in pink) from EC, AF, and CF included typical smooth muscle proteins such as SMTNL2 and MYH11 of which we quantified three splice variants and two were exclusively identified in SMCs. Our analysis also opens up for the investigation of new epigenetic mechanisms, for instance, based on the very significant enrichment for EZH2-regulated proteins ($p < 10^{-34}$).

Clinical application of the heart map to atrial fibrillation. Having generated a map of the healthy human heart, we next investigated if it could serve as a reference to pinpoint molecular

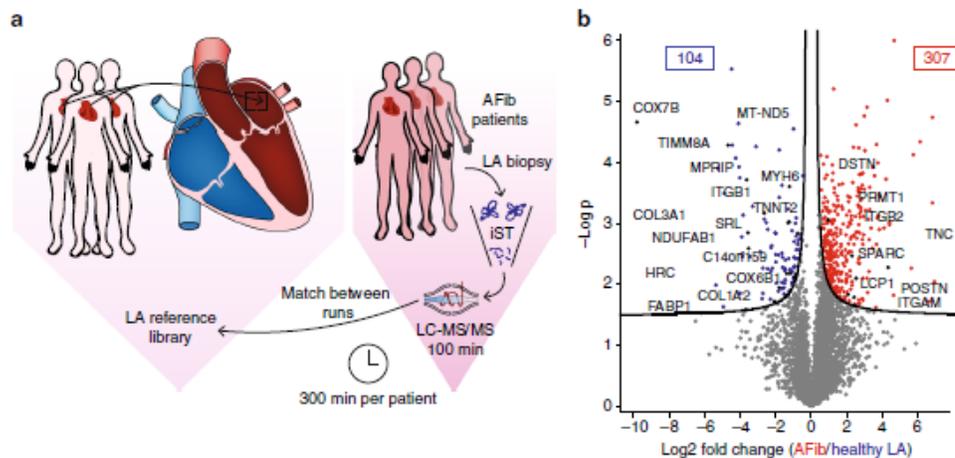


Fig. 6 Clinical application of the healthy human heart atlas to atrial fibrillation. **a** Experimental workflow: LA tissues from three atrial fibrillation patients (AFib) were single-runs of technical triplicates. Data were matched against the healthy human LA library. **b** Volcano plot of the p values vs. the \log_2 protein abundance differences in AFib compared to healthy LA. Significantly up- and downregulated proteins are highlighted in red and blue, respectively (FDR 0.05)

differences between healthy and diseased tissue. To this end, we collected LA samples from three patients suffering from AFib (Supplementary Table 3), the most common heart arrhythmia and a major cause of mortality⁶⁴. We applied a single-run method, in which an in-depth measurement of the proteomic system in question serves as a reference set of identified peptides for deep and high-throughput single-run measurements⁶⁵ to assess the proteomic changes in AFib patients from minimal material and in a timely manner. We found that combining the iST sample preparation, our established healthy reference heart “library,” and single-run, triplicate measurements, any cardiac sample can be profiled in less than two days, of which only 300 min are MS-measuring time (Fig. 6a). In this way, we quantified an average of 3681 proteins for the healthy LAs and 4147 proteins in the AFib group, with excellent average Pearson correlations for technical and “biological replicates” of 0.97 and 0.93, respectively (Supplementary Fig. 13). In the AFib group compared to healthy samples, 104 proteins were significantly downregulated and 307 were upregulated (Fig. 6b and Supplementary Data 9). Proteins with increased expression in AFib are involved in ribonucleoprotein complexes and transcription. Some ECM proteins, such as DSTN, ITGB2, ITGAM, and FLNB, were significantly upregulated in the AFib group, whereas others, including COL1A2, COL3A1, and ITGB1 were downregulated. These results point to a reorganization of the ECM in AFib, explaining previous observations at the level of expressed proteins⁶⁶. There is also evidence of significant contractile remodeling in AFib with several-fold lower expression of TNNT2, HRC, MYH6, SCN5A, and SRL, suggesting disruption of the cardiac tissue. Furthermore, MPRIP, a protein that has been previously associated with an increased number of stress fibers when downregulated⁶⁷ showed 12-fold lower expression. The most significantly downregulated proteins in the AFib group (Fig. 6b and lower yellow box in Supplementary Fig. 14) were enriched for “mitochondrion” ($p < 10^{-100}$). This included the most significantly downregulated protein (COX7B) and two other key mitochondrial proteins—IMMT and TIMM8A, all of which were >25-fold less abundant than in healthy tissue. Mitochondrial dysfunction has already been reported in AFib^{68–71}; however, the broader molecular nature of their defects and whether they are different between patients are not fully understood.

Interestingly, inspection of the hierarchical clustering plot within the AFib group revealed distinct and nonoverlapping clusters of up- and downregulated proteins for each of the three patients (orange and green clusters, Supplementary Fig. 14), whereas nondiseased biological controls showed similar expression to each other. Although the number of AFib patients is much too small to derive a general signature, our data clearly show that the mitochondrial defects reflected in the proteomes are very different between individuals. These proteomic patterns point to a potential molecular subclassification of AFib patients. These observations, however, will require a more thorough analysis including a larger patient cohort.

Discussion

Creating anatomical and cellular maps increases our understanding of human biology and diseases^{15,16,72}. Here, we used “loss-less” high-pH reversed-phase fractionation and high resolution, quantitative MS to generate a heart region, and cell-type-resolved human heart map. Starting from low microgram sample amounts, we quantified over 11,000 proteins, representing by far the deepest proteome of the healthy human heart, which is available in the online database MaxQB²⁵. The ability to work with minimal starting material enables in-depth proteomics analyses from heart biopsies that can be obtained during surgery. Furthermore, our streamlined proteomics workflow enables the profiling of any cardiac sample in less than two days, a realistic time frame for future clinical application. At the anatomic level, we found that the 16 heart regions clustered into the expected three main areas (cavities, vessels, and valves). Binary comparison of subgroups, such as the atria and ventricles, provides crucial information to understand the basis for atrioventricular differences in healthy as well as diseased human hearts, a precondition to identify more specific and reliable biomarkers. To complement the region-resolved heat map, we also established a comprehensive proteomic map of three noncardiomyocyte cell types. This should be particularly useful in future studies to better define and target the CF population. CFs are activated into myofibroblasts after acute myocardial infarction, leading to increased ECM production and wound healing by scar formation within the infarction area. The direct reprogramming of resident CFs after myocardial infarction into induced cardiomyocytes or cardiac

progenitor cells is currently a promising strategy for cardiac regeneration. Our quantitative proteomic data question the specificity of currently used CF markers while providing a promising panel of enriched cell surface markers for therapy. There is a lack of reliable biomarkers for aortic aneurysms (enlargement of the Ao) or more importantly aortic dissection (tear in the wall of the Ao). Currently, both diagnoses rely on clinical examinations and laborious imaging techniques. Our deep quantitative proteome of large arteries, in particular the Ao, could help in establishing a healthy baseline in future studies aiming to define protein expression indicative of these conditions. Likewise, our human cardiac valve proteome can be used as a background for future studies aiming to uncover biomarkers indicative of cardiac valve deterioration. Finally, we show that patients suffering from AFib present both common and distinct proteome profiles, potentially pointing to individual-specific disease manifestation. While our investigation is only a first step, it opens up as a yet-unexplored molecular classification at the level of expressed proteins. Further directions in human heart proteomics could include region- and cell-type-specific mapping at higher resolution, investigation of PTMs, and the combination of proteomics with detailed mechanistic investigations of disease etiology.

Methods

Tissue preparation. In total, 16 healthy heart regions from three adult male individuals (Supplementary Table 1) were collected less than 72 h postmortem during an autopsy after a court order. The hearts of the subjects did not present any relevant injury or signs of cardiac malfunction and were therefore defined as healthy. The 16 heart regions included four main vessels (aorta (Ao), pulmonary artery (PA), vena cava inferior (IVC), and pulmonary vein (PVE)), four heart cavities (right atrium (RA), left atrium (LA), right ventricle (RV), and left ventricle (LV)), four heart valves (tricuspid valve (TV), pulmonary valve (PV), aortic valve (AV), and mitral valve (MV)), the ventricular septum (SepV), the atrial septum (SepA), the left coronary artery (LCA), and the right coronary artery (RCA) were explained by an official medicolegal expert. Samples were stored at -80°C after collection. The investigation was approved by the local ethical committee of the Medical Faculty of the Technical University of Munich (project no. 247/166). The ethical committee explicitly approved the use of human samples in the context of trauma.

Cell isolation. Atrial samples from patients undergoing cardiovascular surgery were cut into 1–2-mm² fragments and digested with 2 mg per ml of collagenase type II (Life Technologies, Cat. No. 17101-015, Carlsbad, CA) in PBS (1 h, 37 °C). After filtration (70- μm cell strainer (Greiner Bio-One, Cat. No. 542070, Frickenhausen, Germany) and red cell lysis (Red Blood Cell Lysis Solution, Miltenyi Biotec, Cat. No. 130-094-183, Bergisch Gladbach, Germany), the remaining cells were resuspended in 1 ml of auto-running MACS buffer (Miltenyi Biotec, Cat. No. 130-090-221). After prepreparation (30- μm filter, Miltenyi Biotec, Cat. No. 130-041-407), CD31-positive endothelial cells (ECs) were isolated using the CD31 MicroBead Kit (Miltenyi Biotec, Cat. No. 130-091-935) and the human FcR-Blocking Regent (Miltenyi Biotec, Cat. No. 130-059-901) in the MACS system (Miltenyi Biotec) according to the manufacturer's instruction. Isolated ECs were cultured in Endothelial Cell Growth Medium 2 (PromoCell, Cat. No. C-22011, Heidelberg, Germany) until confluence. Adipose fibroblasts (AF) were isolated from subcutaneous fat tissue and cardiac fibroblasts (CF) were isolated from atrial samples. Tissues were cut and digested using collagenase solution type II (2.5 h, 37 °C), and resuspended in DMEM high glucose (Biochrom-Millipore, Cat. No. FG 0435; Berlin, Germany) containing 10% fetal calf serum (FCS, Fisher Scientific, Schwerte, Germany), penicillin (100 U per ml), streptomycin (100 μg per ml), both from PanReacAppliChem (AppliChem, Darmstadt, Germany), and sodium-pyruvate (1 mM, Gibco, Karlsruhe, CA) and grown to confluence. Smooth muscle cells (SMCs) were isolated from the arteria mammaria interna (ITA). Vessels were cut longitudinally. With a scalpel, the tissue was cut into square pieces. The pieces were put on BD Primaria™ 6-well plates (Greiner Bio-One, Frickenhausen, Germany) and dried (2–3 h, 37 °C, 5% CO₂). Subsequently, SMC growth medium 2 (PromoCell GmbH, Heidelberg, Germany) was added. When the first SMCs migrated from the tissue, the pieces were removed and cells were grown to confluence. All biopsies from patients undergoing cardiovascular surgery were transferred within 10 min from the operation room in PBS. To passage the ECs, SMCs, and fibroblasts, cells were incubated with trypsin (Life Technologies) diluted 1 in 4 in PBS (5 min, 37 °C, 5% CO₂). Trypsin solution was neutralized using 2 times of media. ECs and SMCs were frozen in liquid nitrogen in 10% DMSO (Roth, Karlsruhe, Germany), 40% FCS, and 50% cell-type-specific medium. CFs and AFs were frozen in liquid nitrogen using 10% DMSO and 10% FCS in fibroblast-specific medium as described above. The generation of cell cultures from patient biopsies

was approved by the local ethical committee of the Medical Faculty of the Technical University of Munich (project no. 1588/06 (amendment) and 2919/10).

Flow cytometry-based quantification of the cell types. Frozen SMCs isolated from the ITA (passage p2) and cultured as described above were thawed and used for flow cytometry analysis. SMCs were fixed in 1% formaldehyde (20 min) and blocked in PBS containing 10% FCS (v/v) and 0.05% sodium azide (w/v). The first antibody anti- α smooth muscle actin (ab5694, Abcam, UK) was added at a ratio of 1:20 in wash/permeabilization buffer (PBS containing 5% FCS (v/v), saponin 0.5% (w/v), and sodium azide (0.05%, w/v)) for 30 min at 4 °C. The secondary antibody goat-anti-rabbit Alexa Fluor 488[®] (ab150077, Abcam, UK) was added at a ratio of 1:2000 in buffer containing saponin (30 min, ice, dark). ECs from atrial biopsies were isolated and cultured as described above. After preparation of a single-cell suspension, one sample of the total biopsy and the positive and negative fraction obtained after MACS were analyzed. All cell fractions were resuspended in ice-cold PBS/0.5% BSA/2 mM EDTA (FACS buffer) containing 5% anti-human CD31 PE-Cy7 (25-0319, eBioscience, Frankfurt, Germany) and incubated for 30 min in the dark. CFs and AFs were isolated as described before. Cells were cultured on gelatin-coated plates and analyzed at passage 1 or 2. Cells were detached using 0.25% trypsin (Life Technologies) diluted at a ratio of 1:4 in PBS. Monoclonal anti-human CD90 PE-Cy5 (eBioscience, 15-0909), anti-human CD105 APC (eBioscience, 17-1057), and anti-human CD45 FITC (eBioscience, 11-9459) antibodies were resuspended at a ratio of 1:20 in ice-cold FACS buffer and incubated for 30 min on ice in the dark. After staining, all cells undergoing flow cytometry analyses were resuspended in ice-cold PBS/0.5% BSA/4 mM EDTA and kept in the dark on ice until flow cytometric analysis was performed with a BD LSRFortessa (BD, San Jose, CA). Negative controls were unstained or stained with the secondary antibody alone. Cytometry data were analyzed with the FlowJo software version 7.6.5 (flowjo@treestar.com).

Immunocytochemistry. SMCs were fixed (4% PFA, 20 min) and permeabilized in PBS-T (0.1% Triton-X-100 in PBS, 10 min). Unspecific binding was blocked with 5% normal goat serum (Abcam, ab7841, Cambridge, UK) in PBST for 30 min. Polyclonal rabbit anti- α smooth muscle actin (Abcam, ab5694) was diluted at a ratio of 1:20 in PBS-T and SMCs were incubated with the first antibody overnight at 4 °C. The secondary antibody goat-anti-rabbit IgG (H&L) Alexa Fluor 555[®] (Abcam, ab150078) was diluted at a ratio of 1:200 in PBS-T. Cells were incubated in the dark for 60 min. ECs isolated from LA and the positive and the negative fraction after MACS sort were stained for CD31. Cells were isolated and processed using the MACS system as described above and plated on cover slides until confluence. Cells were fixed using 4% PFA (10 min) and blocked (5% goat serum in PBS). Polyclonal rabbit anti-CD31 (Abcam, ab28364) was diluted at a ratio of 1:25 and incubated overnight at 4 °C. Secondary antibody goat-anti-rabbit IgG (H&L) Alexa Fluor 555[®] (Abcam, ab150078) was diluted at a ratio of 1:500 and incubated (60 min, dark). Both antibodies were diluted in PBS. CFs and AFs were fixed using 4% PFA/sucrose in PBS (15 min). For VIM staining, cells were permeabilized (0.25% Triton-X-100 in PBS) and unspecific binding sites were blocked (5% goat serum in PBS-T, 1 h). Polyclonal rabbit anti-VIM (Abcam, ab45939) was used as cytoskeleton marker at 1 μg per ml of final concentration in PBS-T. For DDR2, staining cells were washed with PBS after fixation and blocked with 5% goat-serum in PBS. Polyclonal rabbit anti-DDR2 (LSBio, LS-C99151, Seattle, WA) was diluted at a ratio of 1:20 in PBS. Both antibodies were incubated for 1 h. Secondary antibody goat-anti-rabbit IgG (H&L) Alexa Fluor 555[®] (Abcam, ab150078) was diluted at a ratio of 1:500 in either PBS-T or PBS and incubated for 1 h in the dark. For immunocytochemistry, SMCs, ECs, CFs, and AFs were grown to approximately 80% confluence on 4-well chamber cover slides (Millipore EZ slides, Millipore, Darmstadt, Germany). All incubations were performed at room temperature, except overnight incubations. After the last wash, slides were air-dried, mounted in Abcam-mounting medium containing DAPI (Abcam, ab104139), sealed with coverslips, and evaluated under a fluorescent microscope (Axiovert 200 M, Zeiss, D-73447 Oberkochen).

RT-qPCR analysis. Expression of CF-specific proteins was confirmed on transcriptional level by RT-qPCR analysis. ECs, SMCs, CFs, and AFs (p0-4) were lysed with RNA lysis buffer (Qiagen, Erlangen, Germany). Total RNA was purified using the RNeasy total RNA kit (Qiagen) and reverse-transcribed into cDNA with M-MLV reverse transcriptase (Invitrogen, Darmstadt, Germany) according to the manufacturer's recommendation. Expression of ALDH1A2, PDGF, DAPK1, CSRP2, DIAPH3, CES1, and PTPRZ1 was evaluated on a QuantStudio3 (Applied Biosystems, Foster City, CA) using Power SYBR Green Master Mix (Applied Biosystems) and the following conditions: activation of Taq DNA polymerase (15 min at 95 °C) followed by 40 cycles with 15 s at 95 °C, 60 s at 60 °C. The sequences of the used primers are noted in Supplementary Table 4. Quantification was performed using the relative expression software tool REST[®]. Data were normalized to β -actin.

Sample preparation for MS analysis. All 16 heart regions dissected from three trauma victims were washed three times with cold PBS before being crushed in liquid nitrogen using a mortar and pestle. Powdered samples were then resuspended

in 500 μ l of SDC reduction and alkylation buffer and boiled for 10 min to denature proteins¹⁹. Samples were further mixed (six times for 30 s and cooled on ice in-between) using a FastPrep[®]-24 Instrument (MP Biomedicals). Protein concentration was measured using the TrypTohan assay and 300 μ g were further processed for overnight digestion by adding Lys-C and trypsin in a 1:50 ratio (μ g of enzyme to μ g of protein) at 37°C and 1700 rpm. On the following day, samples were sonicated using a bioruptor (15 cycles of 30 s) and further digested for 3 h with Lys-C and trypsin (1:100 ratio). Peptides were acidified to a final concentration of 0.1% trifluoroacetic acid (TFA) for SDB-RPS binding and 40 μ g of peptides were loaded on four 14-gauge Stage-Tip plugs. Peptides were washed first with isopropanol/1% TFA (200 μ l) and then 0.2% TFA (200 μ l) using an in-house-made Stage-Tip centrifuge at 2000 \times g. Peptides were eluted with 60 μ l of elution buffer (80% acetonitrile/1% ammonia) into auto sampler vials and dried at 60°C using a SpeedVac centrifuge (Eppendorf, Concentrator plus). Peptides were resuspended in 2% acetonitrile/0.1% TFA and sonicated (Branson Ultrasonics, Ultrasonics Cleaner Model 2510) before peptide concentration estimation using the Nanodrop. About 30 μ g of peptides of each sample were further fractionated into 54 fractions and concatenated into 8 fractions by high-pH reversed-phase fractionation using the recently described "loss-less" nano-fractionator²⁰. CFs, AFs, ECs, and SMCs were processed similarly to the heart tissue samples without liquid nitrogen crushing and FastPrep[®]-24 Instrument.

Liquid chromatography-MS analysis. Nanoflow LC-MS/MS analysis of tryptic peptides was conducted on a quadrupole Orbitrap mass spectrometer²¹ (Q Exactive HF, Thermo Fisher Scientific, Rockford, IL, USA) coupled to an EASYnLC 1200 ultra-high-pressure system (Thermo Fisher Scientific) via a nano-electrospray ion source. About 1 μ g of peptides were loaded on a 40-cm HPLC-column (75- μ m inner diameter; in-house packed using ReproSil-Pur C18-AQ 1.9- μ m silica beads; Dr Maisch GmbH, Germany). Peptides were separated using a linear gradient from 2 to 20% B in 55 min and stepped up to 40% in 40 min followed by a 5 min wash at 98% B at 350 nl per min where solvent A was 0.1% formic acid and 5% DMSO in water and solvent B was 80% acetonitrile, 5% DMSO, and 0.1% formic acid in water. The total duration of the run was 100 min. Column temperature was kept at 60°C by a peltier element-containing, in-house-developed oven. The mass spectrometer was operated in "top-15" data-dependent mode, collecting MS spectra in the Orbitrap mass analyzer (60,000 resolution, 300–1650 m/z range) with an automatic gain control (AGC) target of 3E6 and a maximum ion injection time of 15 ms. The most intense ions from the full scan were isolated with an isolation width of 1.5 m/z . Following higher-energy collisional dissociation (HCD) with a normalized collision energy (NCE) of 27%, MS/MS spectra were collected in the Orbitrap (15,000 resolution) with an AGC target of 5E4 and a maximum ion injection time of 25 ms. Precursor dynamic exclusion was enabled with a duration of 30 s. For clinical AFib samples, a "top-5" data-dependent acquisition method as described above was modified to increase the dynamic range on the MS1 level by including three segmented MS scans (12 segments each; total AGC target 1E6), covering a m/z range of 400–1200. MS1 resolution was set to 120,000 at m/z 200 throughout.

MS data analysis. Tandem mass spectra were searched against the 2015 Uniprot human databases (UP000005640_9606 and UP000005640_9606_additional) using MaxQuant²¹ version 1.5.5.6 with a 1% FDR at the peptide and protein level, peptides with a minimum length of seven amino acids with carbamidomethylation as a fixed modification, and N-terminal acetylation and methionine oxidations as variable modifications. Enzyme specificity was set as C-terminal to arginine and lysine using trypsin as protease and a maximum of two missed cleavages were allowed in the database search. The maximum mass tolerance for precursor and fragment ions was 4.5 ppm and 20 ppm, respectively. If applicable, peptide identifications by MS/MS were transferred between runs to minimize missing values for quantification with a 0.7-min window after retention time alignment. Label-free quantification was performed with the MaxLFQ algorithm using a minimum ratio count of 1. For clinical AFib samples, the identification transfer was restricted to the healthy LA library only, and we set a minimum ratio count of 2 for label-free quantification.

Statistical analysis. Statistical and bioinformatics analysis was performed with the Perseus software²² (version 1.5.5.0), Microsoft Excel, and R statistical software. Proteins that were identified in the decoy reverse database or only by site modification were not considered for data analysis. We also excluded potential contaminants. Data were further filtered to make sure that identified proteins showed expression in all biological triplicates of at least one heart region and missing values were imputed on the basis of normal distribution (down shift = 1.8, width = 0.15). PCA analysis of the heart region and cell types relied on singular value decomposition and the original feature (protein) space was orthogonally transformed into a set of linearly uncorrelated variables (principal components). These account for distinct types of variability in the data. For hierarchical clustering, LFQ intensities were first z-scored and clustered using Euclidean as a distance measure for column and row clustering. Gene set enrichment analysis (GSEA) was performed using gene set collections from the MSigDB²⁶. Mean log₂ ratios of biological triplicates and the corresponding *p* values were visualized with volcano plots. We used *t*-test

for binary comparisons and SAM with $s_0 = 0.1$ and FDR < 0.05 for the assessment of *t*-test results in volcano plots³².

Copy number calculation and subcellular heart proteome model. Conversion of LFQ intensities to copy number estimations was achieved using the proteomic ruler²⁷. The proteomic ruler plug-in v.0.1.6 was downloaded from the Perseus plug-in store, for use with Perseus version 1.5.5.0. Protein intensities were filtered for three valid values in at least one heart region. Proteins belonging to the GO term "blood microparticle" were removed from the analysis (see Supplementary Data 3 for a full list of removed proteins). Protein groups (proteins that can be distinguished based on the available peptide information) were annotated with amino acid sequence and tryptic peptide information for the leading protein ID, using the .FASTA file used for processing in MaxQuant. Copy numbers per diploid nucleus were estimated using the following settings: averaging mode—"All columns separately," molecular masses—"average molecular mass," detectability correction—"Number of theoretical peptides," scaling mode—"Histone proteomic ruler," ploidy "2," and total cellular protein concentration—"200 g per L." To build a subcellular model of the heart atlas proteome, subcellular localization predictions from spatial proteomics data^{28,29} were matched to the protein groups using the leading canonical protein ID. The median copy number of the three replicates was multiplied by the protein molecular weight to calculate protein mass. The mass of each protein was attributed to the nucleus, cytosol, or a specific organelle according to its distribution in HeLa cells. Since many highly abundant heart-specific proteins were not present in this spatial proteomics database, the top 100 proteins in each heart region were completed for subcellular localization using annotation from UniProt. These manually annotated protein masses were assigned entirely to the respective organelle. This led to a median of 94% of total protein mass being assigned to a specific location.

Data availability. All MS proteomics data have been deposited on ProteomeXchange via the PRIDE database with the data set identifier PXD006675 and can also be accessed in a user-friendly format at maxqb.biochem.mpg.de. All other data supporting the findings of this study are available within this article and in the supplementary material or from the corresponding authors on reasonable request.

Received: 19 June 2017 Accepted: 13 October 2017

Published online: 13 November 2017

References

- Xin, M., Olson, E. N. & Bassel-Duby, R. Mending broken hearts: cardiac development as a basis for adult heart regeneration and repair. *Nat. Rev. Mol. Cell Biol.* **14**, 529–541 (2013).
- Bergmann, O. et al. Dynamics of cell generation and turnover in the human heart. *Cell* **161**, 1566–1575 (2015).
- Pinto, A. R. et al. Revisiting cardiac cellular composition novelty and significance. *Circ. Res.* **118**, 400–409 (2016).
- Celermaier, D. S., Chow, C. K., Marjion, E., Anstey, N. M. & Woo, K. S. Cardiovascular disease in the developing world: prevalences, patterns, and the potential of early disease detection. *J. Am. Coll. Cardiol.* **60**, 1207–1216 (2012).
- Aye, T. T. et al. Proteome-wide protein concentrations in the human heart. *Mol. BioSyst.* **6**, 1917–1927 (2010).
- Lau, E. et al. A large dataset of protein dynamics in the mammalian heart proteome. *Sci. Data* **3**, 160015 (2016).
- Zhang, J. et al. Systematic Characterization of the murine mitochondrial proteome using functionally validated cardiac mitochondria. *Proteomics* **8**, 1564–1575 (2008).
- Peng, Y. et al. Top-down proteomics reveals concerted reductions in myofibrillar and Z-disc protein phosphorylation after acute myocardial infarction. *Mol. Cell Proteomics* **13**, 2752–2764 (2014).
- Lundby, A. & Olsen, J. V. Phosphoproteomics taken to heart. *Cell Cycle* **12**, 2707–2708 (2013).
- Ounzain, S. et al. Genome-wide profiling of the cardiac transcriptome after myocardial infarction identifies novel heart-specific long non-coding RNAs. *Eur. Heart J.* **36**, 353–368 (2015).
- Lindskog, C. et al. The human cardiac and skeletal muscle proteomes defined by transcriptomics and antibody-based profiling. *BMC Genomics* **16**, 475 (2015).
- Kääb, S. et al. Global gene expression in human myocardium—oligonucleotide microarray analysis of regional diversity and transcriptional regulation in heart failure. *J. Mol. Med.* **82**, 308–316 (2004).
- Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).
- Bekker-Jensen, D. B. et al. An optimized shotgun strategy for the rapid generation of comprehensive human proteomes. *Cell Syst.* <https://doi.org/10.1016/j.cels.2017.05.009> (2017).

15. Azimifar, S. B., Nagaraj, N., Cox, J. & Mann, M. Cell-type-resolved quantitative proteomics of murine liver. *Cell Metab.* **20**, 1076–1087 (2014).
16. Sharma, K. et al. Cell type- and brain region-resolved mouse brain proteome. *Nat. Neurosci.* **18**, 1819–1831 (2015).
17. Deshmukh, A. S. et al. Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors. *Mol. Cell. Proteomics* **14**, 841–853 (2015).
18. Murgia, M. et al. Single muscle fiber proteomics reveals fiber-type-specific features of human muscle aging. *Cell Rep.* **19**, 2396–2409 (2017).
19. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 (2014).
20. Kulak, N. A., Geyer, P. E. & Mann, M. Loss-less nano-fractionator for high sensitivity, high coverage proteomics. *Mol. Cell. Proteomics* **16**, 065136 (2017).
21. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
22. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
23. Kim, M.-S. et al. A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
24. Coscia, F. et al. Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nat. Commun.* **7**, 12645 (2016).
25. Schaab, C., Geiger, T., Stoehr, G., Cox, J. & Mann, M. Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol. Cell. Proteomics* **11** (M111), 014068 (2012).
26. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
27. Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A 'proteomic ruler' for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* **13**, 3497–3506 (2014).
28. Itzhak, D. N., Tyanova, S., Cox, J. & Bomer, G. H. Global, quantitative and dynamic mapping of protein subcellular localization. *eLife* **5**, e16950 (2016).
29. Calvo, S. E., Clauser, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–D1257 (2016).
30. Rakus, D., Gizak, A. & Wiśniewski, J. R. Proteomics unveils fibroblast-cardiomyocyte lactate shuttle and hexokinase paradox in mouse muscles. *J. Proteome Res.* **15**, 2479–2490 (2016).
31. Wiśniewski, J. R. et al. Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* **8**, 611 (2012).
32. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
33. Voelkel, T. et al. Lysine methyltransferase Smyd2 regulates Hsp90-mediated protection of the sarcomeric titin springs and cardiac function. *Biochim. Biophys. Acta* **1833**, 812–822 (2013).
34. Ouafik, L., May, V., Keutmann, H. T. & Eipper, B. A. Developmental regulation of peptidylglycine alpha-amidating monooxygenase (PAM) in rat heart atrium and ventricle. Tissue-specific changes in distribution of PAM activity, mRNA levels, and protein forms. *J. Biol. Chem.* **264**, 5839–5845 (1989).
35. England, J. & Loughna, S. Heavy and light roles: myosin in the morphogenesis of the heart. *Cell. Mol. Life Sci.* **70**, 1221–1239 (2013).
36. Lu, Z. Q., Sinha, A., Sharma, P., Kislinger, T. & Gramolini, A. O. Proteomic analysis of human fetal atria and ventricle. *J. Proteome Res.* **13**, 5869–5878 (2014).
37. Kumar, C., Saidapet, C., Delaney, P., Mendola, C. & Siddiqui, M. A. Expression of ventricular-type myosin light chain messenger RNA in spontaneously hypertensive rat atria. *Circ. Res.* **62**, 1093–1097 (1988).
38. Lesage, F. & Lazdunski, M. Molecular and functional properties of two-pore-domain potassium channels. *Am. J. Physiol.* **279**, F793–F801 (2000).
39. Yin, Z., Ren, J. & Guo, W. Sarcomeric protein isoform transitions in cardiac muscle: a journey to heart failure. *Biochim. Biophys. Acta* **1852**, 47–52 (2015).
40. Hamm, C. W. et al. The prognostic value of serum troponin T in unstable angina. *N. Engl. J. Med.* **327**, 146–150 (1992).
41. Weiss, A., Schiaffino, S. & Leinwand, L. A. Comparative sequence analysis of the complete human sarcomeric myosin heavy chain family: implications for functional diversity. Edited by J. Kam. *J. Mol. Biol.* **290**, 61–75 (1999).
42. Eggers, K. M., Johnston, N., James, S., Lindahl, B. & Venge, P. Cardiac troponin I levels in patients with non-ST-elevation acute coronary syndrome—The importance of gender. *Am. Heart J.* **168**, 317–324e1 (2014).
43. Shave, R. et al. Exercise-induced cardiac troponin elevation: evidence, mechanisms, and implications. *J. Am. Coll. Cardiol.* **56**, 169–176 (2010).
44. Keshishian, H. et al. Multiplexed, quantitative workflow for sensitive biomarker discovery in plasma yields novel candidates for early myocardial injury. *Mol. Cell. Proteomics* **14**, 2375–2393 (2015).
45. Govindan, S. et al. Increase in cardiac myosin binding protein-C plasma levels is a sensitive and cardiac-specific biomarker of myocardial infarction. *Am. J. Cardiovasc. Dis.* **3**, 60–70 (2013).
46. Members, W. G. et al. Heart disease and stroke statistics—2009 update. *Circulation* **119**, 480–486 (2009).
47. Ozcelik, F. et al. Diagnostic value of plasma fibronectin level in predicting the presence and severity of coronary artery disease. *Ann. Hematol.* **88**, 249–253 (2009).
48. Orriols, M. et al. Down-regulation of Fibulin-5 is associated with aortic dilation: role of inflammation and epigenetics. *Cardiovasc. Res.* **110**, 431–442 (2016).
49. Yan, J., Li, F., Ingram, D. A. & Quilliam, L. A. Rap1a is a key regulator of fibroblast growth factor 2-induced angiogenesis and together with Rap1b controls human endothelial cell functions. *Mol. Cell. Biol.* **28**, 5803–5810 (2008).
50. Geyer, P. E. et al. Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Mol. Syst. Biol.* **12**, 901 (2016).
51. Schiller, H. B. et al. Time- and compartment-resolved proteome profiling of the extracellular niche in lung injury and repair. *Mol. Syst. Biol.* **11**, 819 (2015).
52. Levine, R. A. et al. Mitral valve disease—morphology and mechanisms. *Nat. Rev. Cardiol.* **12**, 689–710 (2015).
53. Martin-Rojas, T. et al. iTRAQ proteomic analysis of extracellular matrix remodeling in aortic valve disease. *Sci. Rep.* **5**, 17290 (2015).
54. Tan, H. T. et al. Unravelling the proteome of degenerative human mitral valves. *Proteomics* **15**, 2934–2944 (2015).
55. Challa, A. A. & Stefanovic, B. A novel role of vimentin filaments: binding and stabilization of collagen mRNAs. *Mol. Cell. Biol.* **31**, 3773–3789 (2011).
56. Nam, Y.-J. et al. Reprogramming of human fibroblasts toward a cardiac fate. *Proc. Natl Acad. Sci. USA* **110**, 5588–5593 (2013).
57. Ieda, M. et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**, 375–386 (2010).
58. Souders, C. A., Bowers, S. L. K. & Baudino, T. A. Cardiac fibroblast. *Circ. Res.* **105**, 1164–1176 (2009).
59. Furtado, M. B. et al. Cardiogenic genes expressed in cardiac fibroblasts contribute to heart development and repairability and significance. *Circ. Res.* **114**, 1422–1434 (2014).
60. Thomas, P. S., Rajderkar, S., Lane, J., Mishina, Y. & Kaartinen, V. AcvR1-mediated BMP signaling in second heart field is required for arterial pole development: Implications for myocardial differentiation and regional identity. *Dev. Biol.* **390**, 191–207 (2014).
61. Rose, R. A. & Giles, W. R. Natriuretic peptide C receptor signalling in the heart and vasculature. *J. Physiol.* **586**, 353–366 (2008).
62. DeLaughter, D. M. et al. Single-cell resolution of temporal gene expression during heart development. *Dev. Cell* **39**, 480–490 (2016).
63. Hemric, M. E. & Chalovich, J. M. Effect of caldesmon on the ATPase activity and the binding of smooth and skeletal myosin subfragments to actin. *J. Biol. Chem.* **263**, 1878–1885 (1988).
64. Kirchhof, P. et al. 2016 ESC guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur. Heart J.* **37**, 2893–2962 (2016).
65. Nagaraj, N. et al. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top orbitrap. *Mol. Cell. Proteomics* **11**, M111013722 (2012).
66. Burstein, B. & Nattel, S. Atrial fibrosis: mechanisms and clinical relevance in atrial fibrillation. *J. Am. Coll. Cardiol.* **51**, 802–809 (2008).
67. Surks, H. K., Riddick, N. & Ohtani, K.-I. M-RIP targets myosin phosphatase to stress fibers to regulate myosin light chain phosphorylation in vascular smooth muscle cells. *J. Biol. Chem.* **280**, 42543–42551 (2005).
68. Ad, N., Schneider, A., Khalilulin, I., Borman, J. B. & Schwalb, H. Impaired mitochondrial response to simulated ischemic injury as a predictor of the development of atrial fibrillation after cardiac surgery: in vitro study in human myocardium. *J. Thorac. Cardiovasc. Surg.* **129**, 41–45 (2005).
69. Xie, W. et al. Mitochondrial oxidative stress promotes atrial fibrillation. *Sci. Rep.* **5**, 11427 (2015).
70. Chugh, S. S., Blackshear, J. L., Shen, W.-K., Hammill, S. C. & Gersh, B. J. Epidemiology and natural history of atrial fibrillation: clinical implications. *J. Am. Coll. Cardiol.* **37**, 371–378 (2001).
71. Goudami, M. et al. Development of a novel proteomic approach for mitochondrial proteomics from cardiac tissue from patients with atrial fibrillation. *J. Proteome Res.* **10**, 3484–3492 (2011).
72. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaab321 (2017).
73. Kelstrup, C. D. et al. Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field orbitrap mass spectrometer. *J. Proteome Res.* **13**, 6187–6195 (2014).

Acknowledgements

We thank Marta Murgia and Nagarjuna Nagaraj for their helpful suggestions and Korbinian Mayr, Igor Paron, and Gabry Sowa for their assistance in mass-spectrometric analysis. We also thank Jürgen Cox and Pavel Sinitcyn for bioinformatics tools and Sébastien Virreira Winter for critical reading of the manuscript. We thank Christoph Schaab for making the data accessible on MaxQB. We also thank Nazan Puluca, Stefanie Voss, Lena Eschenbach, and Nicole Beck for their generous support collecting patient materials and clinical data acquisition. This work was supported by the Max Planck Society for the Advancement of Science, the Novo Nordisk Foundation (Grant NNFI500001), and by the Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK B 15-005, DZHK B 15-039SE) and Deutsche Forschungsgemeinschaft – Sachmittelantrag (KR3770/7-1, KR3770/9-1).

Author contributions

S.D. acquired and interpreted the proteomics data, developed the concept, and wrote the manuscript. M.D. performed sample collection, FACS sorting, cell culture experiments, gene expression analysis, immunocytochemistry, and edited the paper. P.G. helped with the sample preparation. D.L. analyzed the subcellular proteome. C.B. collected samples from healthy probands. S.D. performed sample collection, immunocytochemistry, and FACS analysis. F.M. helped with the MS measurements. M.A.D. helped with the study design and provided clinical data of included patients. H.L. did the biobanking of selected patient materials, cell culture experiments, and edited the paper. R.L. designed and supervised the study and edited the paper. M.K. developed the concept, supervised the study, and edited the paper. M.M. designed and supervised the study and edited the paper. All the authors have approved the final version.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-01747-2.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

Article 7: Rapid proteomic workflow for solid tumors reveals LSD1 as a drug target in an end stage cancer patient

In preparation for submission to Molecular Oncology

Sophia Doll^{1,2}, Maximilian C. Kriegmair³, Alberto Santos Delgado², Michael Wierer¹, Helen Michele Neil², Stefan Porubsky³, Fabian Coscia^{1,2}, Philipp Geyer^{1,2}, Andreas Mund², Philipp Nuhn³, Matthias Mann^{1,2}

¹ Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, 82152, Germany

² Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark.

³ University of Mannheim, Mannheim, Germany.

Recent groundbreaking advances in MS-based proteomics have set in motion the transformation of translational MS-based cancer proteomics from an idea to a practice. In this manuscript, we present a rapid proteomic workflow for the analysis of clinically relevant cancer tissues allowing quantification of thousands of proteins in several hours of measurement time. Applying our workflow to an extremely rare cancer type, the urachus carcinoma, we uncovered a potential therapeutic target: Lysine specific histone demethylase 1 (LSD1). We created the possibility for medical doctors and scientists to truly use proteomics for end stage cancer patients to identify additional actionable therapeutic options.

.

Rapid proteomic workflow for solid tumors reveals LSD1 as a drug target in an end stage cancer patient

In preparation for submission to Molecular Oncology

Sophia Doll^{1,2}, Maximilian C. Kriegmair³, Alberto Santos Delgado², Michael Wierer¹, Helen Michele Neil², Stefan Porubsky⁴, Fabian Coscia^{1,2}, Philipp E. Geyer^{1,2}, Andreas Mund², Philipp Nuhn³, Matthias Mann^{1,2}

¹ Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany

² Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark.

³ Department of Urology, University Medical Center Mannheim, University of Heidelberg Germany.

⁴ Department of Pathology, University Medical Center Mannheim, University of Heidelberg, Germany.

Abstract

Recent advances in mass spectrometry (MS)-based technologies are now set to transform translational cancer proteomics from an idea to a practice. Here, we present a robust proteomic workflow for the analysis of clinically relevant cancer tissues, which allows quantitation of thousands of tumor proteins in several hours of measuring time and a total turnaround of a few days. We applied it to an extremely rare and chemorefractory urachal carcinoma. Quantitative comparison of lung metastasis and surrounding tissue revealed several statistically significantly upregulated proteins, among them lysine specific histone demethylase 1 (LSD1/KDM1A). LSD1 is an epigenetic regulator and is the target of active development efforts in oncology. This demonstrates that clinical cancer proteomics can rapidly and efficiently identify actionable therapeutic options even in end stage cancer patients. While currently described for a single case study, we envision that it can be applied broadly to other patients in a similar condition.

Introduction

Genomic and transcriptomic investigations based on next generation sequencing has revolutionized the field of oncology in the last decade and allowed the molecular profiling of thousands of tumors in different cancer types (Cancer Genome Atlas Research Network et al., 2013; Stratton et al., 2009). While these technologies have led to a better understanding of cancer origin and heterogeneity, it has often been challenging to turn mutation patterns into actionable therapeutic suggestions. It has also become evident that the development and complexity of cancer cannot be understood at the genetic or transcriptomic level alone. Clearly, proteins, the driving biological entities in cells, also play crucial roles in cancer. So far, proteomics – the large scale study of all proteins in a given system - has lagged behind genomics for technological reasons. However, following groundbreaking advances in mass spectrometry (MS) based proteomics, comprehensive characterization of nearly complete proteomes has now become a reality (Aebersold and Mann, 2016; Bekker-Jensen et al., 2017). In parallel, several proteomic tumor analysis consortia (e.g. CPTAC) have been launched and aim to systematically identify and characterize cancer-relevant proteins. So far, these consortia have focused on knowledge generation, rather than focusing on specific clinical applications.

Here we set out to use state of the art proteomics technology directly in a clinical oncology context. Our group has already established proteomic workflows enabling processing of clinically relevant tissue samples to great depth, even for formalin-fixed paraffin-embedded (FFPE) material (Wiśniewski et al., 2011, 2013). Recently, we have combined nearly all sample processing steps in a single reaction tube, thereby reducing preparation time, contamination and loss, while increasing quantification accuracy (inStageTip method) (Kulak et al., 2014; Doll et al., 2017). We reasoned that these advances would now enable rapid analysis of individual tumor tissues to inform treatment decisions, especially in patients with rare and end stage cancer malignancies, where evidence for therapeutic strategies and clinical trials are often lacking.

Urachal carcinomas originate from a remnant of the fetal structure connecting the allantois and the bladder. This form of cancer is very rare, accounting for less than 1% of all bladder cancers, aggressive, and consequently little studied. Patients with metastatic urachal cancer have poor prognosis and limited treatment options (Szarvas et al., 2016). Only a few cases have been investigated at the genomic level (Collazo-Lorduy et al., 2016; Singh et al., 2016a) and there are no global protein expression profiles of urachal carcinoma that could aid the search for biomarkers, therapeutic targets, or disease signatures.

A 57-year-old female presented with an urachal carcinoma that has metastasized to the lungs. The tumor had become refractory to all available chemo or radio therapy regimes but the patient wished to continue treatment. Based on the inStageTip sample preparation method, we developed a fast and reproducible workflow capable of producing analysis results in only about two days. Profiling the proteomic landscape of the metastasized tumor in comparison to the normal appearing surrounding tissue, we aimed to uncover potential therapeutic targets and gain a deeper understanding of the molecular mechanisms underlying this disease and its progression. We also employed proteomics to characterize the archived primary tumor and compared our results to deep sequencing data that we obtained from the same metastases.

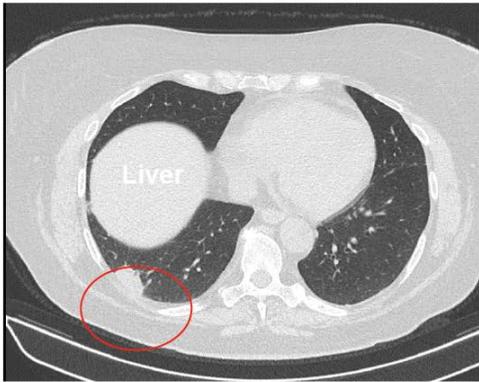
Results

Prior clinical course

Early symptoms of our patient included gross hematuria, which led us to perform a subsequent cystoscopy and bladder biopsy. Histopathology revealed a mucinous adenocarcinoma in the bladder, a finding consistent with a diagnosis of urachal carcinoma. As a first line of treatment, we performed a partial cystectomy and lymphadenectomy. Our final pathology showed a pT3b, pN1, L1, V1, R0 mucinous urachal carcinoma of the bladder (Supplementary figure 1A). Follow-up CT scans were performed on a three-month basis. Nine months after resection, the CT scan

revealed two suspicious hypodense lesions in the liver (Segment 5 and 4a) as well as a local recurrence found at the bladder dome (Supplementary Figure 1B and C). The local tumor board recommended chemotherapy, including one cycle of XELOX (oxaliplatin and capecitabine) and nine cycles of FOLFOX (folinic acid, fluorouracil, and oxaliplatin). Chemotherapy led to a partial hepatic response but was stopped due to severe peripheral neuropathy. To assess further treatment strategies the local recurrence was biopsied and confirmed transurethrally. After tumor board consultation, we performed a resection of the local recurrence combined with a partial hepatectomy and subsequent radiotherapy of the local recurrence side (59,4Gy). In later stages, two metastases were diagnosed at the introitus vaginae and the CT-scan of the thorax revealed bilateral noduli. Subsequent chemotherapy with four cycles of Gemcitabine/Cis-Platin led to a mixed response and further pulmonary progression of a predominant singular nodule was diagnosed (Figure 1A). At this point, all standard treatment options were exhausted and we set out to resect the lung metastasis and surrounding healthy tissue for subsequent proteomic analyses. Due to medical and psychological issues the resection was delayed for two months. In the thoracoscopy a disseminated pleural carcinosis was observed, that was most likely covered by pleural effusion in the preoperative CT-scan (Figure 1B). Pleural metastases and healthy pleura were biopsied, washed in PBS, flash frozen, and immediately transferred for proteomic analyses within one day.

A) March 2017



B) June 2017

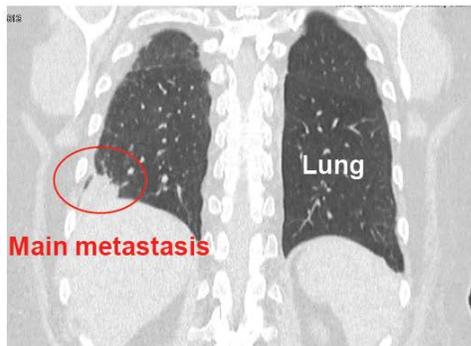
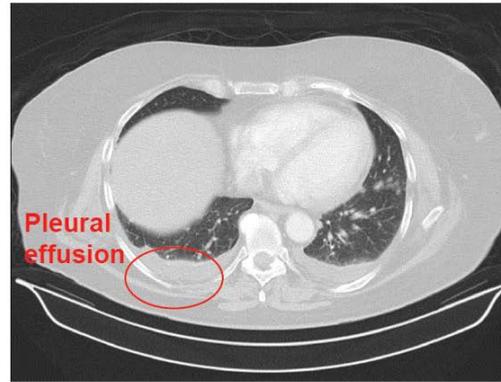


Figure 1 Preoperative CT-scans of the urachal carcinoma patient. A) CT scan in March 2017 showing a main metastases in the lungs. B) CT scan in June 2017 depicting a pleural effusion before the surgery, hiding a pleural carcinosis.

Streamlined proteomics workflow applied to chemorefractory carcinoma

To be useful in a clinical oncological setting, we reasoned that a proteomics workflow need to fulfill several criteria, including rapid overall analysis time (few days), extreme sensitivity (few thousand cells), depth of quantitative proteome coverage (several thousand proteins) along with robustness and reproducibility. The workflow that we adapted fulfills all these criteria (Methods): Briefly, we performed all sample preparation in a single reaction vial, based on the in-StageTip (iST) method sample preparation (Kulak et al., 2014). We chose a single-run LC-MS/MS workflow, rather than pre-fractionating the sample, to minimize measurement time and maximize quantitative accuracy. All bioinformatic analysis was done in the freely

available MaxQuant and Perseus software environments (Cox and Mann, 2008; Tyanova et al., 2016).

Upon shipment of the samples in the late afternoon, we started by lysing the samples and extracting the proteins. The surrounding fat of the tissues were removed by high speed centrifugation. Proteins were subsequently digested overnight using proteases. On the following day, we analyzed the peptide mixtures using a state-of-the-art label free workflow on a quadrupole – Orbitrap mass spectrometer (Figure 2B). Each sample, constituting a few μg of material, was measured in single shot triplicate measurements using 100 min high-performance liquid chromatography (HPLC) gradients. Analysis in MaxQuant specified a false discovery rate (FDR) of less than 1% at the peptide and protein levels. In total, we identified 50,870 sequence-unique peptides, corresponding to 5,562 protein groups (proteins that can be distinguished based on the available peptide information). The MaxLFQ algorithm (Cox et al., 2014) quantified 5,543 proteins in total and similar coverage in all samples. For further analysis, we only considered the subset of 4,857 proteins in our data with quantitative values with at least 70% valid values across the samples. Mean sequence coverage of all proteins by identified peptides was about 25%. Signal intensities for the quantified proteins spanned about five orders of magnitudes, with hemoglobin as one of the most abundant proteins, despite extensive washing of the samples with PBS before sample processing. Quantitative reproducibility was excellent, demonstrated by Pearson correlation coefficients between 0.97-0.99, and was on par, or even exceeded the values we previously achieved in cell lines systems (Coscia et al., 2016). We likewise observed high correlation values between control tissues taken from different locations (0.92) and between two different samplings of the metastases (0.97). Raw data and MaxQuant results are provided online and are available in our proteomic database MaxQB (Schaab et al., 2012). The complete workflow can be performed in less than 2.5 days and we conclude that it is well suited to application in the clinic.

A) Proteomics Timeline

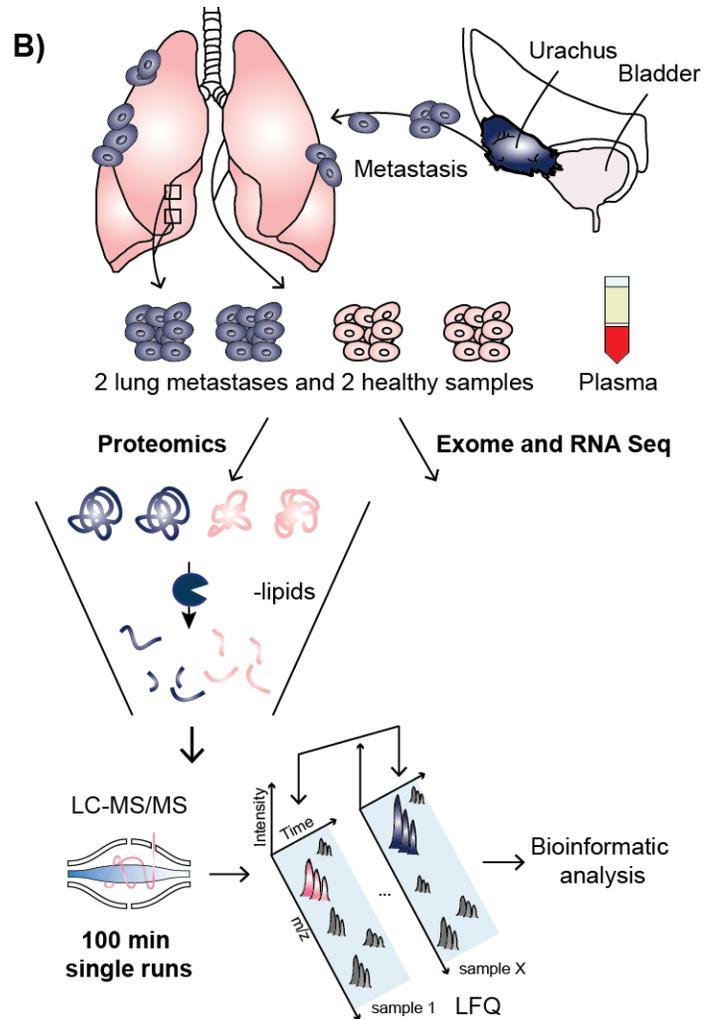
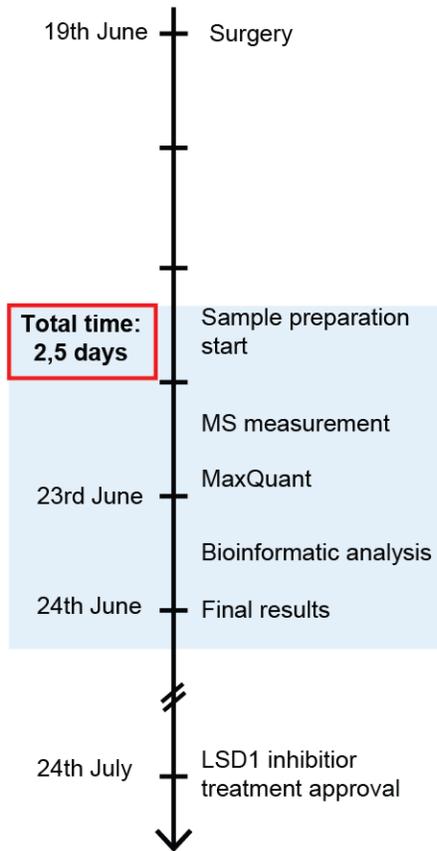


Figure 2 Case study proteomics workflow. A) Timeline of the project. B) Experimental design, including source of material, inStageTip sample preparation, and depiction of the analytical workflow.

Proteome analysis reveals LSD1 as a potential therapeutic target

For a functional view of the proteomic data, we used volcano plots – a type of scatter plot often used to present large-scale proteomic data sets - to compare expression differences between lung pleural metastases and healthy-appearing pleura. Based on a t-test for binary comparison and employing a 5% FDR, we found that 108 (2.2%) proteins showed significant alteration, of which 47 displayed significant up-regulation and 61 down-regulation in the metastases. Gene set enrichment analysis

(GSEA) using gene set collections from the MSigDB (Subramanian et al., 2005) revealed that proteins upregulated in the metastases were significantly enriched ($p < 5e-6$) for the terms epithelial mesenchymal transition, tumor invasiveness, and tumor metastasis. For example, periostin (POSTN) has previously been reported to promote cell motility in several cancer types, was 13-fold higher expressed in the metastases compared with non-diseased tissue (Gillan et al., 2002; Ishiba et al., 2014; Mikheev et al., 2015). The most up-regulated (>100-fold) protein in the metastases was thrombospondin-2 (THBS2) and is also involved in cell invasion as well as angiogenesis and correlates with poor survival (Bornstein, 2009; Iruela-Arispe et al., 2004; Lin et al., 2016; Qian et al., 2017; Wang et al., 2016). Another protein driving cell invasion, methylthioribose-1-phosphate isomerase (MRI1) was highly significantly upregulated but only 1.9-fold (Kabuyama et al., 2009). These observations demonstrate that the proteomics experiment performed as expected and suggest an important role of these proteins in the metastatic progression of urachal carcinoma.

In contrast, downregulated proteins were very significantly enriched in mitochondrial proteins ($p < 1e-17$), such as pyruvate carboxylase (PC), Acetyl-CoA carboxylase 2 (ACACB), and Acyl-coenzyme A thioesterase 2 (ACOT2). Interestingly, Ras suppressor protein 1 (RSU1) was about 4-fold down-regulated in the metastases. Apoptosis-inducing factor 2 (AIFM2) was 28-fold down-regulated in the metastases compared with non-diseased tissue. These observations suggest a regulatory role of RSU1 and AIFM2 in urachal carcinoma metastases.

In an effort to derive therapeutic options, we first reduced the total number of significantly upregulated proteins by applying a more stringent cutoff (1% FDR). This yielded four significantly upregulated proteins in the metastatic tissue: methylthioribose-1-phosphate isomerase (MRI1), solute carrier family 22 member 18 (SLC22A18), collagen alpha-1 (XI) chain (COL11A1), and lysine-specific histone demethylase 1A (KDM1A, also known as LSD1) (Figure 3A). Next, we asked which of these proteins were potentially druggable, which left us with LSD1 as the sole remaining candidate. We quantified LSD1 with 11 unique peptides, reaching an

approximate sequence coverage of 20%, and found that it was 16-fold more highly expressed in the metastases compared to the control.

LSD1 is an epigenetic regulator that demethylates both the activating histone mark H3K4me and the repressive mark H3K9me, thereby acting as a coactivator or corepressor, depending on cellular context. LSD1 has previously been reported as upregulated in multiple cancer types and its inhibition has antitumor activity in lung cancer (Mohammad and Kruger, 2016; Singh et al., 2015). These findings led to the development of multiple LSD1 inhibitors that are currently in clinical trials (Alsaqer et al., 2017; Mohammad and Kruger, 2016; Schmidt and McCafferty, 2007). Even though it was unclear whether the lung metastases would respond to a LSD1 inhibitor, there were no other rational or reasonable treatment options available at this point. Unfortunately, extensive efforts to obtain one of these drugs for use in our patient ultimately proved unsuccessful. Luckily, tranylcypromine a drug developed decades ago and FDA approved for the treatment of depression and anxiety (Burger and Yost, 1948) has recently been shown to irreversibly inhibit LSD1 as a side effect (Binda et al., 2010; Ulrich et al., 2017; Zheng et al., 2016). This analogue of amphetamine is a monoamine oxidase (MAO) inhibitor, an enzyme family that is mechanistically related to LSD1. Tranylcypromine and derivatives of this drug already showed clinical efficacy for several conditions in clinical trials, including the treatment of AML. The local tumor board approved treatment with this drug and our patient was prescribed a tyramine-free diet, to prevent accumulation of tyramine (which is normally metabolized by MAO) leading to high blood pressure, which may culminate in a hypertensive crisis (Gillman, 2011; Ulrich et al., 2017). However, a baseline CT at the initiation of therapy revealed dramatic metastatic progression to the liver, concurrent with hepatic failure (Supplementary Figure 1D). The patient was then transferred to palliative care ward and died soon after.

MS-based proteomics is a multifaceted technology and further allowed us to investigate the plasma proteome of our patient. Based on our previously developed 'plasma proteome profiling' pipeline (Geyer et al., 2016a) we quantified

approximately 460 proteins in triplicate LCMS measurements enabling quantification of inflammatory proteins, such as CRP and the majority of the complement system (Supplementary Figure 2). Furthermore, we identified the entire inflammatory panel which we have previously reported and found it to be clearly elevated compared to normal controls (Geyer et al., 2016b), as expected in a patient with end-stage malignancy and heavy metastatic load.

We also investigated whether the patient would be likely to respond to immunotherapy. MS-based measurements did not reveal any expression of PD1 or PDL1 proteins, an observation that was later confirmed by immunohistochemistry (Supplementary Figure 3D). We also did not observe any immune cells infiltration in the metastases, suggesting a poor response to immunotherapy-based treatments.

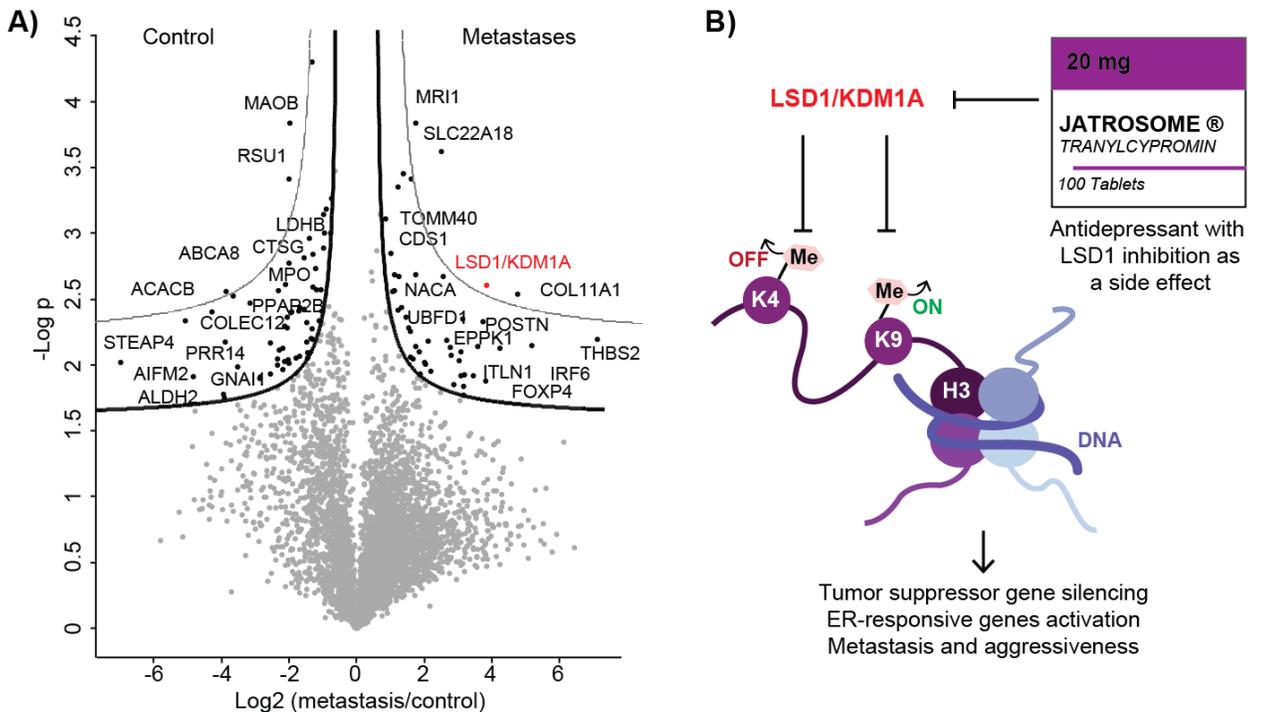


Figure 3 Proteins differentially expressed in the urachal carcinoma lung metastases.

A) Volcano plot of the p-values (y-axis) vs. the log2 protein abundance differences (x-axis) between metastases and control, with lines of significance colored in black or grey lines corresponding to a 5% or 1% FDR, respectively. B) Mechanisms of action of LSD1/KDM1A.

Proteomic analysis of the primary tumor

To further investigate the proteomic landscape of our quantitative and in depth proteomic case study, we next analyzed the proteome of the primary tumor, which had been preserved as FFPE material for several years. H&E staining revealed that the primary tumor was rich in extracellular mucin and stroma compared to healthy control tissue (Figure 4B and C). Our proteomic analysis revealed major differences between the primary and healthy surrounding tissue (Figure 4A). In total, we quantified approximately 4,300 proteins and found that mucinous (MUC1 and MUC2) and mesenchymal proteins (such as THBS2, COL11A1, and CTHRC1) were significantly upregulated in the primary tumor compared to healthy surrounding tissue. Generally, the epithelial mesenchymal transition, and thus mesenchymal gene upregulation is associated with poor prognosis in various malignancies including colorectal cancer and ovarian cancer (Chen et al., 2014; Rokavec et al., 2017; Sleeman and Thiery, 2011). The fact that mesenchymal proteins were highly enriched in the primary tumor, is concordant with the later development of multiple and aggressive metastases. Interestingly, we also found that LSD1 appeared to be upregulated in the metastases compared to the primary tumor, albeit not significantly.

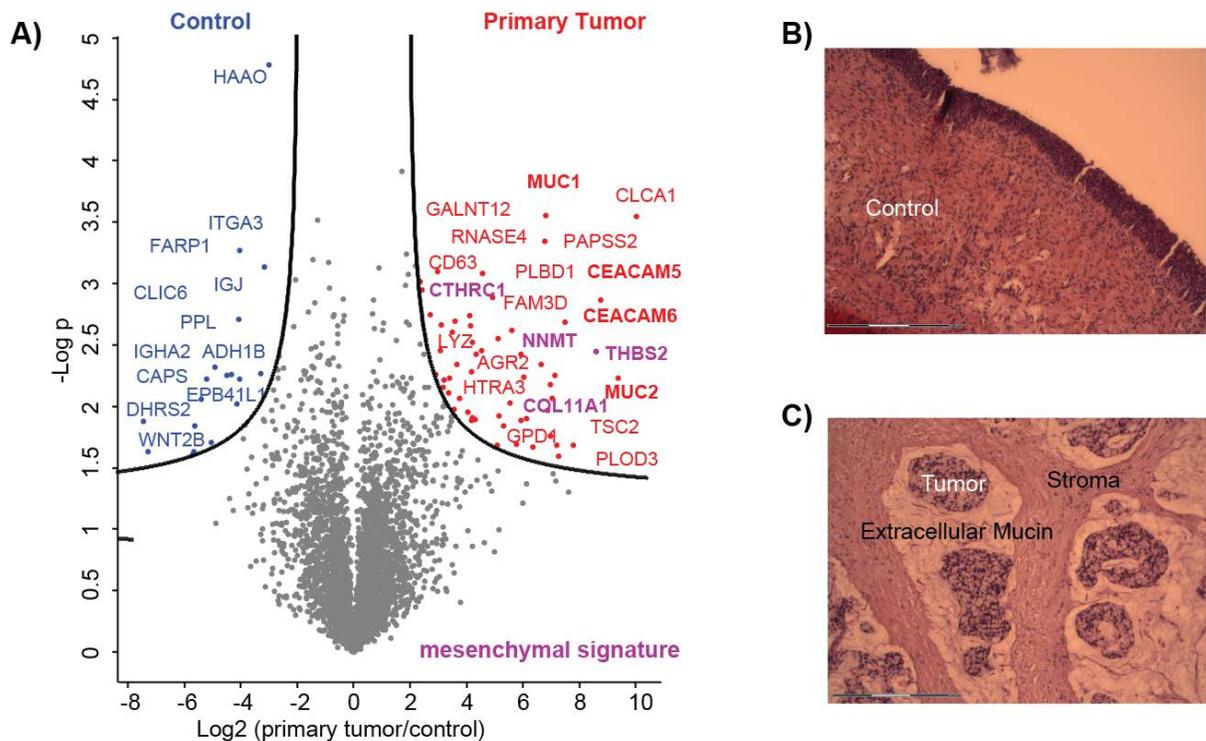


Figure 4 Differentially expressed proteins in the primary tumor. A) Volcano plot of the p-values vs. the \log_2 protein abundance differences between primary tumor and control, with significance lines (5% FDR) colored in black. B) H&E stainings of healthy control tissue surrounding the primary tumor (C), reveals prominent stroma formation.

Next generation sequencing analysis of the metastases

To gain additional insights into the overall molecular mechanisms underlying urachal carcinoma, tumor etiology and to compare transcriptomics to proteomics, we also extracted RNA and DNA extraction for subsequent next generation sequencing. The quality of the extracted RNA from the metastatic samples, however, was poor, prohibiting transcriptomic analysis. DNA is more stable, allowing us to perform exome sequencing on our sample. Overall, we observed hundreds of mutations in coding regions, indicating a hypermutated phenotype, consistent with a previous report (Kardos et al., 2017). Comparing the mutations to a database of druggable genes (Broad Institute), yielded a total of 160 mutations of potential therapeutic interest. Among those, we examined the mutation spectrum of the

growth factor receptor (EGFR) as EGFR-inhibitors (e.g. gefitinib), have been described for urachal carcinoma recently (Collazo-Lorduy et al., 2016; Singh et al., 2016b; Sirintrapun et al., 2014). The EGFR pathway member K-RAS is frequently mutated in urachal cancer and we identified a missense mutation at position 117 (K117N, exon 4) that has previously been associated with various cancer forms such as bladder and colorectal cancer. Unfortunately, multiple studies reported that patients with K-RAS mutations in exons 2, 3 and 4 did not respond to EGFR-targeted therapy (Bokemeyer et al., 2015; Douillard et al., 2013). We also found two intronic and two exonic somatic mutations of the LSD1 gene, which, however, scored neutral by mutation effect predictors (Supplementary Figure 4).

Discussion

Modern oncology is at a turning point, where systemic cancer treatment is moving from multi-cytotoxic chemotherapies towards individual targeted therapies. This is particularly promising for patients suffering from rare cancer forms, where standard chemotherapies often fail and large clinical studies are unlikely to be performed. In the near future, sequencing at the genomic, transcriptomic and proteomic levels might provide the basis for individual targeted treatment prescription and thereby change clinical practice. However, the large spectrum of mutations does not necessarily lead to clear therapeutic options, a problem that becomes even more acute when considering mutational heterogeneity of most tumors. These general challenges were reflected in our case study, where mutational analysis did not lead to a clear treatment recommendation. In contrast, our personalized MS-based proteomic analysis worked robustly and quickly on both the lung metastases and the archived primary tumor.

The current standard treatment for localized urachal cancer is surgery, whereas chemotherapy is used on the metastatic disease. Given the rarity of this cancer type, robust data from prospective trials on chemotherapy regimens is unlikely to be obtainable and evidence mainly consists of small retrospective cohorts. Due to the similarity of urachal cancer to colorectal adenocarcinoma and urothelial carcinomas,

treatment regimens are generally extrapolated from these diseases, justifying the FOLFOX therapy prescribed to our patient. Targeted epidermal growth factor receptor (EGFR)-inhibitors (e.g. gefitinib), have been prescribed for urachal carcinoma recently. To guide decisions concerning this alternative therapy option, we further looked into K-RAS mutations and uncovered a missense mutation. However, EGFR-targeted therapy was not prescribed because multiple studies reported that patients with similar K-RAS mutations as our patient, did not respond to therapy (Bokemeyer et al., 2015; Douillard et al., 2013). Furthermore, the elevated liver enzymes of our patients were contraindicated such a therapy. In the search for possible treatment options in this patient, we found that PDL1 and CD8 immunohistochemistry were also negative, suggesting a poor response to check-point inhibitors.

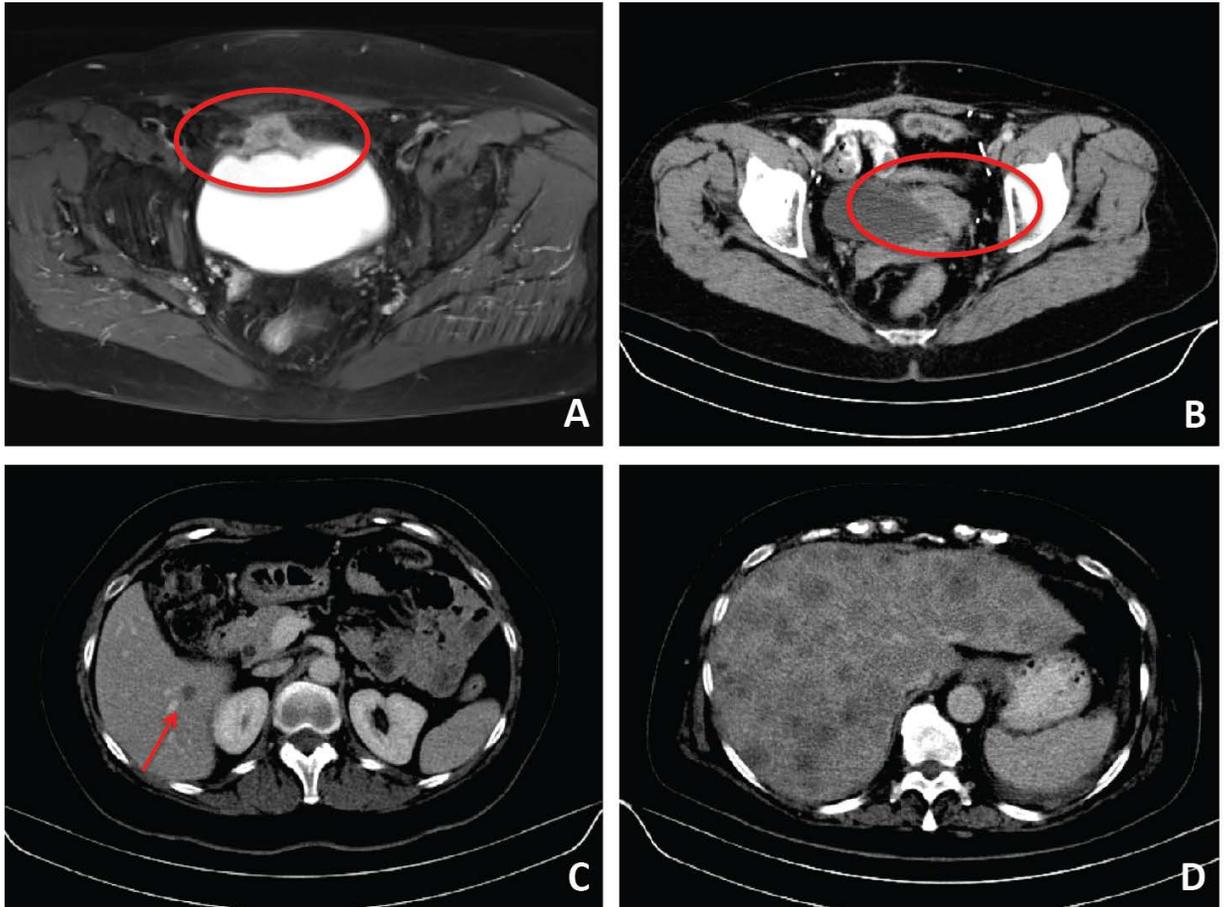
Lacking evidence-based treatment options for our end-stage patient, who was willing to exhaust all possibilities, we turned to our MS-based proteomic analysis, which identified LSD1 as a therapeutic target highly enriched in metastatic tissue, thus providing a promising treatment opportunity. Of note, our workflow allowed fast proteomic analysis of clinical tumor tissue providing timely results to the patient and the clinicians. The proteomic sample preparation and data analysis were accomplished in only about two days, faster than the genomic analysis. This highlights the promise of MS-based proteomics in clinical routine, where fast target identification for cancer patients beyond standard treatment could be highly beneficial.

In summary, we demonstrated a fast and reproducible proteomic workflow that created the possibility for clinicians to use proteomics for personalized diagnosis and treatment in the clinical setting. By combining genomic with proteomics data, we further informed the therapeutic decision. We aim to apply this workflow to cancer patients in a variety of chemorefractive tumors, in the hope of identifying additional treatment options for at least some of them.

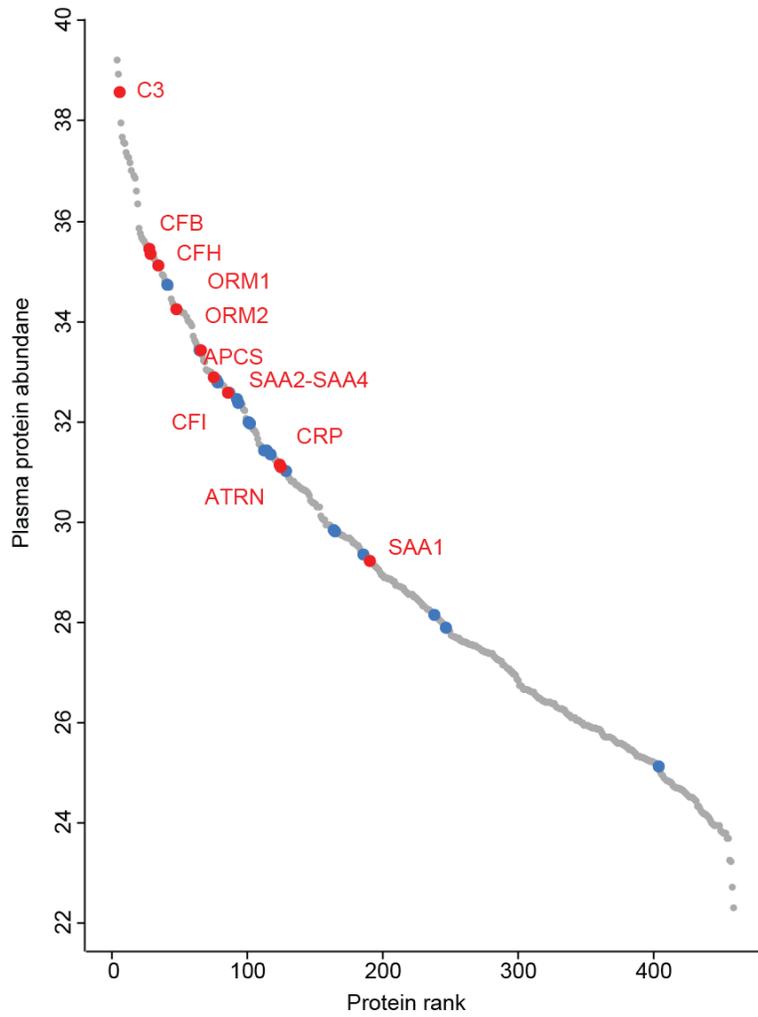
Acknowledgments

We thank Peter Treit for critical reading of the manuscript and Natalie Krahmer for sample preparation input. We also thank Igor Paron and Gaby Sowa for their assistance in mass-spectrometric measurements. This work was supported by the Max Planck Society for the Advancement of Science and by the Novo Nordisk Foundation (Grant NNF15CC0001).

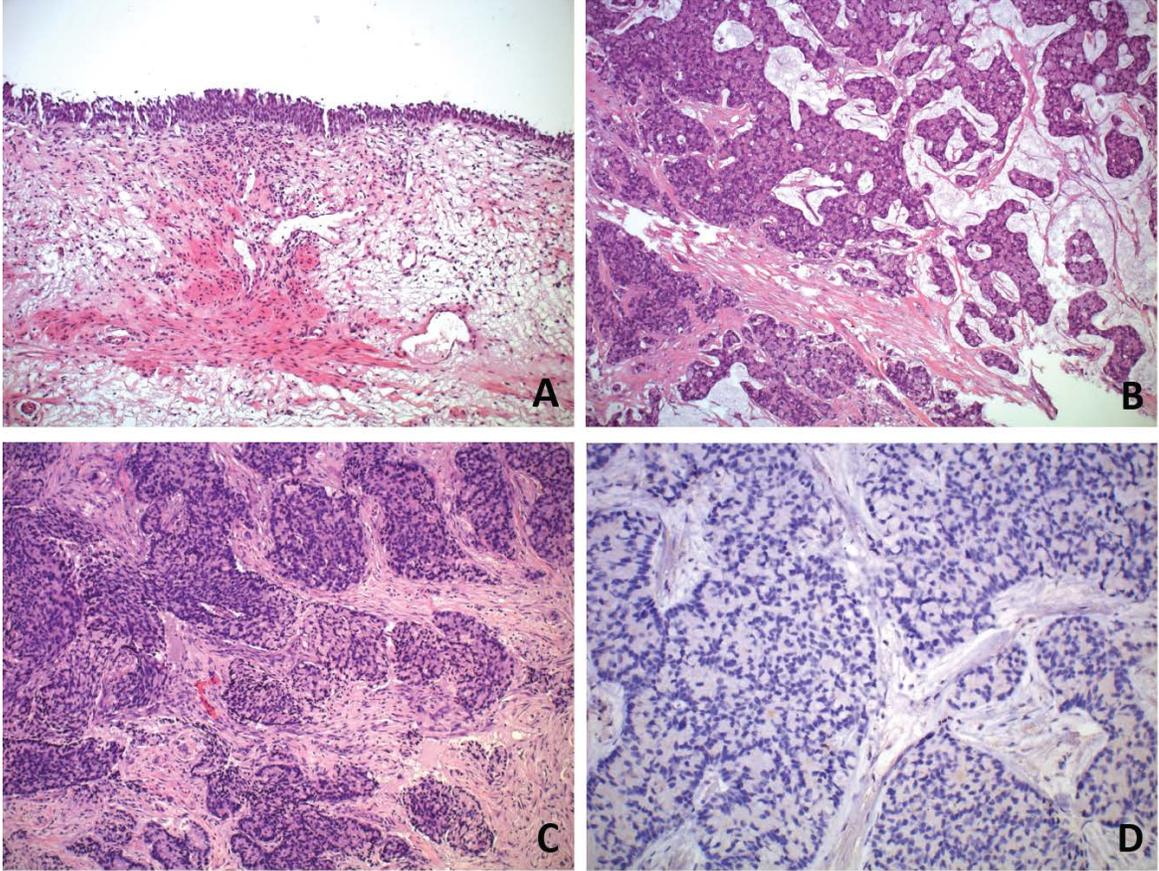
Supplementary Figures



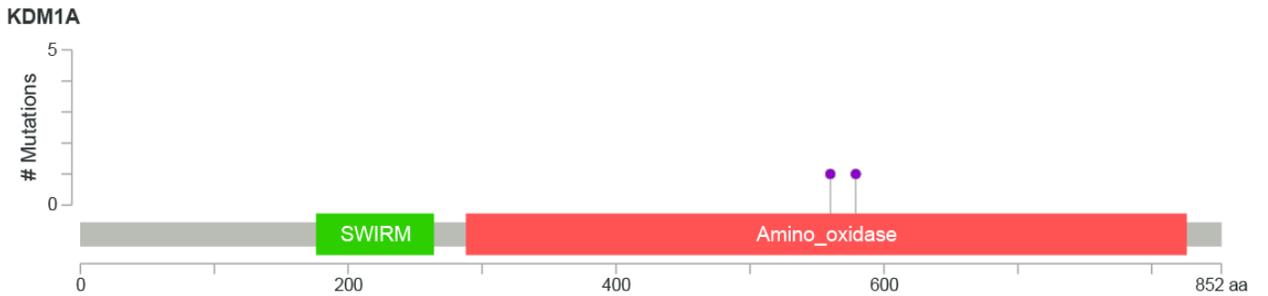
Supplementary Figure 1: A) MR-imagery of the primary tumor showed a suspicious mass at the anterior bladder wall (red circle). B-C) Follow-up CT-imagery revealed local recurrence nine months after partial nephrectomy (red circle) and a hepatic metastasis (red arrow). D) Massive progression of hepatic metastases in the CT scan prior to LSD1 therapy.



Supplementary Figure 2: Plasma proteome abundance rank. The previously reported inflammatory panel is highlighted in red, other complement proteins are represented in blue.



Supplementary Figure 3: H&E stainings. A) Healthy urothelium of the partial cystectomy specimen. B) Primary mucinous urachal adenocarcinoma. C-D) Hepatic metastasis of the urachal carcinoma with negative PDL1 immunohistochemistry.



Supplementary Figure 4: Mutation diagram of LSD1.

Material and methods

Sample preparation for mass spectrometry analysis

The lung metastases were collected during surgery and were washed three times with cold PBS before flash freezing the samples in liquid nitrogen and shipping on dry ice. The samples were cut in half to enable genomic and proteomic analysis.

Proteomic sample preparation

Control and lung metastases samples were thawed on ice and prepared according to the in stage tip sample preparation method (Kulak et al. Nat Methods, 2014). Briefly, 100 µl of the reducing alkylating sodium deoxycholate buffer (PreOmics) was added to the samples before protein denaturation at 100°C for 20 min. Proteins were then digested by LysC and trypsin overnight at 37°C and 1700 rpm. Peptides were acidified to a final concentration of 0.1% trifluoroacetic acid (TFA) for SDB-RPS binding and desalted before LC-MS/MS analysis.

Liquid chromatography-MS analysis

Samples were measured on an on a quadrupole Orbitrap mass spectrometer (Scheltema et al., 2014; Kelstrup et al., 2014) (Q Exactive HF, Thermo Fisher Scientific, Rockford, IL, USA) coupled to an EASYnLC 1200 ultra-high-pressure system (Thermo Fisher Scientific) via a nano-electrospray ion source. About 1 µg of peptides were loaded on a 40 cm HPLC-column (75 µm inner diameter; in-house packed using ReproSil-Pur C18-AQ 1.9 µm silica beads; Dr Maisch GmbH, Germany). Peptides were separated using a linear gradient from 3% to 23% B in 82 min and stepped up to 40% in 8 min at 350 nl per min where solvent A was 0.1% formic acid in water and solvent B was 80% acetonitrile and 0.1% formic acid in water. The total duration of the gradient was 100 min. Column temperature was kept at 60 °C by a Peltier element-containing, in-house developed oven. The mass spectrometer was operated in 'top-15' data-dependent mode, collecting MS spectra in the Orbitrap mass analyzer (60,000 resolution, 300-1,650 m/z range) with an

automatic gain control (AGC) target of 3E6 and a maximum ion injection time of 25 ms. The most intense ions from the full scan were isolated with a width of 1.4 m/z. Following higher-energy collisional dissociation (HCD) with a normalized collision energy (NCE) of 27%, MS/MS spectra were collected in the Orbitrap (15,000 resolution) with an AGC target of 1E5 and a maximum ion injection time of 25 ms. Precursor dynamic exclusion was enabled with a duration of 20 s.

MS data analysis

Tandem mass spectra were searched against the 2015 Uniprot human databases (UP000005640_9606 and UP000005640_9606_additional) using MaxQuant version 1.5.3.34 with a 1% FDR at the peptide and protein level, peptides with a minimum length of seven amino acids with carbamidomethylation as a fixed modification and N-terminal acetylation and methionine oxidations as variable modifications. Enzyme specificity was set as C-terminal to arginine and lysine using trypsin as protease and a maximum of two missed cleavages were allowed in the database search. The maximum initial mass tolerance for precursor and fragment ions were 4.5 ppm and 20 ppm, respectively. If applicable, peptide identifications by MS/MS were transferred between runs to minimize missing values for quantification with a 0.7 min window after retention time alignment. Label-free quantification was performed with the MaxLFQ algorithm using a minimum ratio count of 1.

Statistical analysis

Statistical and bioinformatics analysis was performed with the Perseus software (Tyanova et al., 2016) (version 1.5.5.0), Microsoft Excel, and R statistical software. Proteins that were identified in the decoy reverse database or only by site modification were not considered for data analysis. Mean log₂ ratios of biological triplicates and the corresponding p-values were visualized with volcano plots. We used t-test for binary comparisons and SAM with $s_0=0.1$ and $FDR<0.05$ for the assessment of t-test results in volcano plots.

References

- Aebersold, R., and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 347–355.
- Alsaqer, S.F., Tashkandi, M.M., Kartha, V.K., Yang, Y.-T., Alkheriji, Y., Salama, A., Varelas, X., Kukuruzinska, M., Monti, S., Bais, M.V., et al. (2017). Inhibition of LSD1 epigenetically attenuates oral cancer growth and metastasis. *Oncotarget* 8, 73372–73386.
- Bekker-Jensen, D.B., Kelstrup, C.D., Batth, T.S., Larsen, S.C., Haldrup, C., Bramsen, J.B., Sørensen, K.D., Høyer, S., Ørntoft, T.F., Andersen, C.L., et al. (2017). An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* 4, 587–599.e4.
- Binda, C., Valente, S., Romanenghi, M., Pilotto, S., Cirilli, R., Karytinis, A., Ciossani, G., Botrugno, O.A., Forneris, F., Tardugno, M., et al. (2010). Biochemical, Structural, and Biological Evaluation of Tranylcypromine Derivatives as Inhibitors of Histone Demethylases LSD1 and LSD2. *J. Am. Chem. Soc.* 132, 6827–6833.
- Bokemeyer, C., Köhne, C.-H., Ciardiello, F., Lenz, H.-J., Heinemann, V., Klinkhardt, U., Beier, F., Duecker, K., van Krieken, J.H., and Tejpar, S. (2015). FOLFOX4 plus cetuximab treatment and RAS mutations in colorectal cancer. *Eur. J. Cancer* 51, 1243–1252.
- Bornstein, P. (2009). Thrombospondins function as regulators of angiogenesis. *J. Cell Commun. Signal.* 3, 189–200.
- Burger, A., and Yost, W. (1948). Arylcycloalkylamines. I. 2-phenylcyclopropylamine (*J. Am. Chem. Soc.*).
- Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120.
- Chen, D., Gassenmaier, M., Maruschke, M., Riesenberger, R., Pohla, H., Stief, C.G., Zimmermann, W., and Buchner, A. (2014). Expression and Prognostic Significance of a Comprehensive Epithelial-Mesenchymal Transition Gene Set in Renal Cell Carcinoma. *J. Urol.* 191, 479–486.
- Collazo-Lorduy, A., Castillo-Martin, M., Wang, L., Patel, V., Iyer, G., Jordan, E., Al-Ahmadie, H., Leonard, I., Oh, W.K., Zhu, J., et al. (2016). Urachal Carcinoma Shares Genomic Alterations with Colorectal Carcinoma and May Respond to Epidermal Growth Factor Inhibition. *Eur. Urol.* 70, 771–775.
- Coscia, F., Watters, K.M., Curtis, M., Eckert, M.A., Chiang, C.Y., Tyanova, S., Montag, A., Lastra, R.R., Lengyel, E., and Mann, M. (2016). Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nat. Commun.* 7, 12645.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.

- Cox, J., Hein, M.Y., Lubner, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 13, 2513–2526.
- Doll, S., Dreßen, M., Geyer, P.E., Itzhak, D.N., Braun, C., Doppler, S.A., Meier, F., Deutsch, M.-A., Lahm, H., Lange, R., et al. (2017). Region and cell-type resolved quantitative proteomic map of the human heart. *Nat. Commun.* 8, 1469.
- Douillard, J.-Y., Oliner, K.S., Siena, S., Tabernero, J., Burkes, R., Barugel, M., Humblet, Y., Bodoky, G., Cunningham, D., Jassem, J., et al. (2013). Panitumumab–FOLFOX4 Treatment and RAS Mutations in Colorectal Cancer. *N. Engl. J. Med.* 369, 1023–1034.
- Geyer, P.E., Kulak, N.A., Pichler, G., Holdt, L.M., Teupser, D., and Mann, M. (2016a). Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* 2, 185–195.
- Geyer, P.E., Wewer Albrechtsen, N.J., Tyanova, S., Grassl, N., Iepsen, E.W., Lundgren, J., Madsbad, S., Holst, J.J., Torekov, S.S., and Mann, M. (2016b). Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Mol. Syst. Biol.* 12, 901.
- Gillan, L., Matei, D., Fishman, D.A., Gerbin, C.S., Karlan, B.Y., and Chang, D.D. (2002). Periostin Secreted by Epithelial Ovarian Carcinoma Is a Ligand for $\alpha V\beta 3$ and $\alpha V\beta 5$ Integrins and Promotes Cell Motility. *Cancer Res.* 62, 5358–5364.
- Gillman, P.K. (2011). Advances Pertaining to the Pharmacology and Interactions of Irreversible Nonselective Monoamine Oxidase Inhibitors. *J. Clin. Psychopharmacol.* 31, 66–74.
- Iruela-Arispe, M.L., Luque, A., and Lee, N. (2004). Thrombospondin modules and angiogenesis. *Int. J. Biochem. Cell Biol.* 36, 1070–1078.
- Ishiba, T., Nagahara, M., Nakagawa, T., Sato, T., Ishikawa, T., Uetake, H., Sugihara, K., Miki, Y., and Nakanishi, A. (2014). Periostin suppression induces decorin secretion leading to reduced breast cancer cell motility and invasion. *Sci. Rep.* 4, 7069.
- Kabuyama, Y., Litman, E.S., Templeton, P.D., Metzner, S.I., Witze, E.S., Argast, G.M., Langer, S.J., Polvinen, K., Shellman, Y., Chan, D., et al. (2009). A Mediator of Rho-dependent Invasion Moonlights as a Methionine Salvage Enzyme. *Mol. Cell. Proteomics* 8, 2308–2320.
- Kardos, J., Wobker, S.E., Woods, M.E., Nielsen, M.E., Smith, A.B., Wallen, E.M., Pruthi, R.S., Hayward, M.C., McGinty, K.A., Grilley-Olson, J.E., et al. (2017). Comprehensive Molecular Characterization of Urachal Adenocarcinoma Reveals Commonalities With Colorectal Cancer, Including a Hypermutable Phenotype. *JCO Precis. Oncol.* 1–12.
- Kelstrup, C.D., Jersie-Christensen, R.R., Batth, T.S., Arrey, T.N., Kuehn, A., Kellmann, M., and Olsen, J.V. (2014). Rapid and Deep Proteomes by Faster Sequencing on a Benchtop Quadrupole Ultra-High-Field Orbitrap Mass Spectrometer. *J. Proteome Res.* 13, 6187–6195.
- Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N., and Mann, M. (2014). Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* 11, 319–324.

- Lin, X., Hu, D., Chen, G., Shi, Y., Zhang, H., Wang, X., Guo, X., Lu, L., Black, D., Zheng, X., et al. (2016). Associations of THBS2 and THBS4 polymorphisms to gastric cancer in a Southeast Chinese population. *Cancer Genet.* 209, 215–222.
- Mikheev, A.M., Mikheeva, S.A., Trister, A.D., Tokita, M.J., Emerson, S.N., Parada, C.A., Born, D.E., Carnemolla, B., Frankel, S., Kim, D.-H., et al. (2015). Periostin is a novel therapeutic target that predicts and regulates glioma malignancy. *Neuro-Oncol.* 17, 372–382.
- Mohammad, H.P., and Kruger, R.G. (2016). Antitumor activity of LSD1 inhibitors in lung cancer. *Mol. Cell. Oncol.* 3, e1117700.
- Qian, Z., Zhang, G., Song, G., Shi, J., Gong, L., Mou, Y., Han, Y., Qian, Z., Zhang, G., Song, G., et al. (2017). Integrated analysis of genes associated with poor prognosis of patients with colorectal cancer liver metastasis. *Oncotarget* 8, 25500–25512.
- Rokavec, M., Kaller, M., Horst, D., and Hermeking, H. (2017). Pan-cancer EMT-signature identifies RBM47 down-regulation during colorectal cancer progression. *Sci. Rep.* 7, 4687.
- Schaab, C., Geiger, T., Stoehr, G., Cox, J., and Mann, M. (2012). Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol. Cell. Proteomics MCP* 11, M111.014068.
- Scheltema, R.A., Hauschild, J.-P., Lange, O., Hornburg, D., Denisov, E., Damoc, E., Kuehn, A., Makarov, A., and Mann, M. (2014). The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer. *Mol. Cell. Proteomics* 13, 3698–3708.
- Schmidt, D.M.Z., and McCafferty, D.G. (2007). trans-2-Phenylcyclopropylamine Is a Mechanism-Based Inactivator of the Histone Demethylase LSD1. *Biochemistry (Mosc.)* 46, 4408–4416.
- Schork, N.J. (2015). Personalized medicine: Time for one-person trials. *Nat. News* 520, 609.
- Singh, H., Liu, Y., Xiao, X., Lin, L., Kim, J., Hummelen, P.V., Wu, C.-L., Bass, A.J., Saylor, P.J., Singh, H., et al. (2016a). Whole exome sequencing of urachal adenocarcinoma reveals recurrent NF1 mutations. *Oncotarget* 7, 29211–29215.
- Singh, H., Liu, Y., Xiao, X., Lin, L., Kim, J., Hummelen, P.V., Wu, C.-L., Bass, A.J., Saylor, P.J., Singh, H., et al. (2016b). Whole exome sequencing of urachal adenocarcinoma reveals recurrent NF1 mutations. *Oncotarget* 7, 29211–29215.
- Singh, M.M., Johnson, B., Venkatarayan, A., Flores, E.R., Zhang, J., Su, X., Barton, M., Lang, F., and Chandra, J. (2015). Preclinical activity of combined HDAC and KDM1A inhibition in glioblastoma. *Neuro-Oncol.* 17, 1463–1473.
- Sirintrapun, S.J., Ward, M., Woo, J., and Cimic, A. (2014). High-stage urachal adenocarcinoma can be associated with microsatellite instability and KRAS mutations. *Hum. Pathol.* 45, 327–330.
- Sleeman, J.P., and Thiery, J.P. (2011). SnapShot: The Epithelial-Mesenchymal Transition. *Cell* 145, 162–162.e1.
- Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. *Nature* 458, 719.

- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* *102*, 15545–15550.
- Szarvas, T., Módos, O., Niedworok, C., Reis, H., Szendrői, A., Szász, M.A., and Nyirády, P. (2016). Clinical, prognostic, and therapeutic aspects of urachal carcinoma—A comprehensive review with meta-analysis of 1,010 cases. *Urol. Oncol. Semin. Orig. Investig.* *34*, 388–398.
- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* *13*, 731–740.
- Ulrich, S., Ricken, R., and Adli, M. (2017). Tranylcypromine in mind (Part I): Review of pharmacology. *Eur. Neuropsychopharmacol.* *27*, 697–713.
- Wang, X., Zhang, L., Li, H., Sun, W., Zhang, H., and Lai, M. (2016). THBS2 is a Potential Prognostic Biomarker in Colorectal Cancer. *Sci. Rep.* *6*, 33366.
- Wiśniewski, J.R., Ostasiewicz, P., and Mann, M. (2011). High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J. Proteome Res.* *10*, 3040–3049.
- Wiśniewski, J.R., Duś, K., and Mann, M. (2013). Proteomic workflow for analysis of archival formalin-fixed and paraffin-embedded clinical samples to a depth of 10 000 proteins. *Proteomics Clin. Appl.* *7*, 225–233.
- Zheng, Y.C., Yu, B., Jiang, G.Z., Feng, X.J., He, P.X., Chu, X.Y., Zhao, W., and Liu, H.M. (2016). Irreversible LSD1 Inhibitors: Application of Tranylcypromine and Its Derivatives in Cancer Treatment. *Curr. Top. Med. Chem.* *16*, 2179–2188.

IV) Discussion

Approximately nine million deaths worldwide are attributed to cancer, and its poor clinical outcome is related to a diverse array of factors, including late diagnosis, lack of specific treatment, therapy resistance, and the limited connection of pre-clinical with clinical research. Many cancer subtypes remain poorly understood and few targeted therapeutic treatments are available. Revolutionary advances in genomics technologies have so far mainly contributed to our understanding of cancer origin and heterogeneity but unfortunately not had much impact on the clinic. In parallel, technology improvements in MS-based proteomics now finally enable the analysis of the molecular drivers of cellular function – proteins. These advances enable the identification and quantification of nearly complete proteomes, as more than 10,000 proteins and PTMs can be identified in cells and tissues, including challenging muscle tissues (Article 6). It is therefore possible to use MS-based proteomics to study cellular changes at a near comprehensive, systems-wide level by examining changes in protein and PTM expression, protein subcellular localization and protein interaction partners.

To address the fundamental question of how the genotype is mechanistically translated into phenotype it is important to interrogate the role of specific genes, for instance by deleting or inserting a gene or mutation of interest, on a global proteomic and phosphoproteomic scale (Articles 4 and 5). Likewise, with small molecule inhibitors we can investigate the perturbation of phosphorylation signaling cascades following kinase inhibition, on a systems-wide scale (Article 3). For example, a recent study showed that cancer driven alterations in human tumors can be mapped in human cancer cell lines and correlated with drug treatment sensitivity²¹³. Cell line models could therefore be used in the future to guide and link the development and application of therapies in the clinic.

Further proteomic developments now permit in-depth tissue analysis in about two days from obtaining the sample to final analysis result, this speed being a prerequisite for clinical application (Article 7). Based on such a rapid proteomic

analysis, we uncovered a potential therapeutic target in a chemorefractory cancer patient, which enabled clinicians devise a third line treatment. We envision that this workflow can be applied to cancer patients at much earlier stages of disease progression, with the ultimate goal to distinguish patients most likely to respond and benefit from a given treatment from those who will only endure its negative side effects.

Together, the technical advances detailed in this thesis lead us into a new paradigm, in which MS-based proteomics is transformed from being a tool for specialist laboratories to a powerful technology for translational cancer proteomics used in cutting edge clinics to deliver personalized treatment options.

Building on the work presented here, we aim to expand our robust, rapid and sensitive proteomic workflow to the analysis of human cancer tissues at both greater depth and higher throughput and accuracy. We plan to apply this workflow to micro- and macro-dissected cancer tissues in an automated fashion to explore the proteome of large FFPE cancer tissue cohorts found in multiple biobanks as well as biopsies collected during surgery. By facilitating the analysis of thousands of samples in a reasonable timeframe, one could study the underlying molecular mechanisms of cancer at the proteome level, testing the validity of known biomarkers, while potentially uncovering new ones. In two preliminary studies, we have already analyzed the proteome of 60 melanoma FFPE samples and adenoma samples, which yielded relevant results.

To achieve more reproducible measurements (which is a prerequisite in clinical practice) we envision using data independent acquisition (DIA) measurement strategies. DIA has become particularly attractive with the latest technological developments, predominantly the sequencing speed and sensitivity of the latest Orbitrap instruments²¹⁴. With the latest instrumentation, DIA largely eliminates the missing value problem. It also provides the means to couple MS to fast LC separation techniques. The recently developed Evosep LC, which is based on a rapid elution concept²¹⁵, significantly reduces the overhead time between sample pick up and MS measurement start point. This new design of LC makes use of a

pre-formed gradient that already contains the sample. Furthermore, sample to sample carry over, and thus contamination, is considerably reduced as peptides are loaded on disposable StageTips, which are used for direct elution of the samples. Based on a 90 min method, one can reach a throughput of up to 16 tissue samples per day and instrument. Chemical multiplexing, such as in 11-plex TMT would enable an even higher throughput, enabling direct comparison of samples in the same spectrum. This might obviate the need for technical triplicate measurements that are still often used in label free approaches to ensure high accuracy. This combination of robustness and performance technologies would make it possible to tackle clinical studies at relatively high throughput.

The integration with various omics approaches, such as in proteogenomics, will play even more important roles in cancer research and metabolomics also holds great promise for precision surgery in particular. This will generate exponential data flows that will necessitate concurrent advances in bioinformatics and computational proteomics. Machine-learning technologies have the power to uncover cancer drivers and facilitate the ability to generate biological insights from large datasets. The integration of large scale omics data sets have led to a new era of data-driven medicine, termed high definition medicine²¹⁶. Likewise, in a truly personalized medicine approach, one would begin by determining the personal healthy baseline of an individual to later ascertain the exact pathology in a patient-resolved fashion. In cases where this is not possible, one can at least establish typical reference ranges, as we have begun to do with the proteome of a healthy human heart (in this case, compared to diseased atrial fibrillation hearts) (Article 6). On a global scale, large data sets would also enable better prevention and treatment options if these big data are managed more effectively. Apart from improving the health of millions, this could result in billions of dollar savings to health care systems, as better management of current clinical trials could allow more efficient data sharing and advance clinical trials more rapidly²¹⁷.

In my journey from analyzing a single protein through a comprehensive understanding of deep proteome networks, I also explored the scientific

interconnections between industry, academia, and the clinic during my PhD. It is my hope that collaborations between these different institutions will play increasingly crucial roles in the future to develop more effective and affordable drugs, manage big data and fight cancer more efficiently.

V) References

1. Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* **45**, 1113–1120 (2013).
2. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347 (2016).
3. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413 (2008).
4. Walsh, C. T. *Posttranslational Modification of Proteins: Expanding Nature's Inventory*. (Roberts and Company Publishers, 2006).
5. Smith, L. M. *et al.* Proteoform: a single term describing protein complexity. *Nat. Methods* **10**, 186 (2013).
6. <https://www.phosphosite.org>.
7. Bekker-Jensen, D. B. *et al.* An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* **4**, 587–599.e4 (2017).
8. Olsen, J. V. *et al.* Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **127**, 635–648 (2006).
9. Francavilla, C. *et al.* Functional Proteomics Defines the Molecular Switch Underlying FGF Receptor Trafficking and Cellular Outputs. *Mol. Cell* **51**, 707–722 (2013).
10. Klaeger, S. *et al.* The target landscape of clinical kinase drugs. *Science* **358**, eaan4368 (2017).
11. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).
12. Wilm, M. Principles of electrospray ionization. *Mol. Cell. Proteomics* mcp.R111.009407 (2011). doi:10.1074/mcp.R111.009407
13. Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68**, 1–8 (1996).
14. Fornelli, L. *et al.* Top-down proteomics: Where we are, where we are going? *J. Proteomics* (2017). doi:10.1016/j.jprot.2017.02.002

15. Hein, M. Y., Sharma, Kirti, K., Cox, J. & Mann, M. *Hein, M.Y., Sharma, K., Cox, J., and Mann, M. (2013). Handbook of Systems Biology: Concepts and Insights: Proteomic Analysis of Cellular Systems. Academic Press, 3-25. (Academic Press, 2013).*
16. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
17. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2006).
18. Mann, M. & Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**, 4390–4399 (1994).
19. Mann, M. The Rise of Mass Spectrometry and the Fall of Edman Degradation. *Clin. Chem.* **62**, 293–294 (2016).
20. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
21. Yates, J. R. Mass spectrometry and the age of the proteome. *J. Mass Spectrom. JMS* **33**, 1–19 (1998).
22. Yates, J. R. Mass spectrometry: from genomics to proteomics. *Trends Genet.* **16**, 5–8 (2000).
23. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359 (2009).
24. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319 (2014).
25. Burkhardt, J. M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A. & Zahedi, R. P. Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J. Proteomics* **75**, 1454–1462 (2012).
26. Olsen, J. V., Ong, S.-E. & Mann, M. Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Mol. Cell. Proteomics* **3**, 608–614 (2004).
27. Giansanti, P., Tsiatsiani, L., Low, T. Y. & Heck, A. J. R. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat. Protoc.* **11**, 993 (2016).
28. Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat. Protoc.* **2**, 933 (2007).

29. Maile, T. M. *et al.* Mass Spectrometric Quantification of Histone Post-translational Modifications by a Hybrid Chemical Labeling Method. *Mol. Cell. Proteomics* **14**, 1148–1158 (2015).
30. Humphrey, S. J., Azimifar, S. B. & Mann, M. High-throughput phosphoproteomics reveals *in vivo* insulin signaling dynamics. *Nat. Biotechnol.* **33**, 990 (2015).
31. Thingholm, T. E., Jensen, O. N., Robinson, P. J. & Larsen, M. R. SIMAC (Sequential Elution from IMAC), a Phosphoproteomics Strategy for the Rapid Separation of Monophosphorylated from Multiply Phosphorylated Peptides. *Mol. Cell. Proteomics* **7**, 661–671 (2008).
32. Larsen, M. R., Graham, M. E., Robinson, P. J. & Roepstorff, P. Improved Detection of Hydrophilic Phosphopeptides Using Graphite Powder Microcolumns and Mass Spectrometry Evidence for *in Vivo* Doubly Phosphorylated Dynamin I and Dynamin III. *Mol. Cell. Proteomics* **3**, 456–465 (2004).
33. Zhou, H. *et al.* Enhancing the Identification of Phosphopeptides from Putative Basophilic Kinase Substrates Using Ti (IV) Based IMAC Enrichment. *Mol. Cell. Proteomics* **10**, M110.006452 (2011).
34. Abe, Y. *et al.* Deep Phospho- and Phosphotyrosine Proteomics Identified Active Kinases and Phosphorylation Networks in Colorectal Cancer Cell Lines Resistant to Cetuximab. *Sci. Rep.* **7**, 10463 (2017).
35. van der Mijn, J. C. *et al.* Evaluation of different phospho-tyrosine antibodies for label-free phosphoproteomics. *J. Proteomics* **127**, 259–263 (2015).
36. Wagner, S. A. *et al.* A Proteome-wide, Quantitative Survey of *In Vivo* Ubiquitylation Sites Reveals Widespread Regulatory Roles. *Mol. Cell. Proteomics* **10**, M111.013284 (2011).
37. Kim, W. *et al.* Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. *Mol. Cell* **44**, 325–340 (2011).
38. Carlson, S. M. & Gozani, O. Emerging Technologies to Map the Protein Methylome. *J. Mol. Biol.* **426**, 3350–3362 (2014).
39. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat. Rev. Mol. Cell Biol.* **15**, 536 (2014).
40. Wiita, A. P., Seaman, J. E. & Wells, J. A. Global Analysis of Cellular Proteolysis by Selective Enzymatic Labeling of Protein N-Termini. *Methods Enzymol.* **544**, 327–358 (2014).
41. Melchior, K. *et al.* Protein- versus peptide fractionation in the first dimension of two-dimensional high-performance liquid chromatography-matrix-assisted laser

- desorption/ionization tandem mass spectrometry for qualitative proteome analysis of tissue samples. *J. Chromatogr. A* **1217**, 6159–6168 (2010).
42. Wang, Y. *et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *PROTEOMICS* **11**, 2019–2026 (2011).
 43. Delmotte, N., Lasaosa, M., Tholey, A., Heinzle, E. & Huber, C. G. Two-Dimensional Reversed-Phase × Ion-Pair Reversed-Phase HPLC: An Alternative Approach to High-Resolution Peptide Separation for Shotgun Proteome Analysis. *J. Proteome Res.* **6**, 4363–4373 (2007).
 44. Song, C. *et al.* Reversed-Phase-Reversed-Phase Liquid Chromatography Approach with High Orthogonality for Multidimensional Separation of Phosphopeptides. *Anal. Chem.* **82**, 53–56 (2010).
 45. Gilar, M., Olivova, P., Daly, A. E. & Gebler, J. C. Orthogonality of Separation in Two-Dimensional Liquid Chromatography. *Anal. Chem.* **77**, 6426–6434 (2005).
 46. Kulak, N. A., Geyer, P. E. & Mann, M. Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics. *Mol. Cell. Proteomics* **16**, 694–705 (2017).
 47. Hahne, H. *et al.* DMSO enhances electrospray response, boosting sensitivity of proteomic experiments. *Nat. Methods* **10**, 989 (2013).
 48. Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal. Chem.* **8**, 61–80 (2015).
 49. Grinfeld, D., Aizikov, K., Kreutzmann, A., Damoc, E. & Makarov, A. Phase-Constrained Spectrum Deconvolution for Fourier Transform Mass Spectrometry. *Anal. Chem.* **89**, 1202–1211 (2017).
 50. Paizs, B. & Suhai, S. Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **24**, 508–548 (2005).
 51. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing | Nature Genetics. Available at: <https://www.nature.com/articles/ng.259>. (Accessed: 7th December 2017)
 52. Frese, C. K. *et al.* Improved Peptide Identification by Targeted Fragmentation Using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* **10**, 2377–2388 (2011).
 53. Biemann, K. Appendix 5. Nomenclature for peptide fragment ions (positive ions). in *Methods in Enzymology* **193**, 886–887 (Academic Press, 1990).

54. Nagaraj, N., D'Souza, R. C. J., Cox, J., Olsen, J. V. & Mann, M. Feasibility of Large-Scale Phosphoproteomics with Higher Energy Collisional Dissociation Fragmentation. *J. Proteome Res.* **9**, 6786–6794 (2010).
55. Eliuk, S. M., Maltby, D., Panning, B. & Burlingame, A. L. High Resolution Electron Transfer Dissociation Studies of Unfractionated Intact Histones from Murine Embryonic Stem Cells Using On-line Capillary LC Separation. *Mol. Cell. Proteomics MCP* **9**, 824–837 (2010).
56. Myers, S. A., Daou, S., Affar, E. B. & Burlingame, A. Electron transfer dissociation (ETD): The mass spectrometric breakthrough essential for O-GlcNAc protein site assignments – A study of the O-GlcNAcylated protein Host Cell Factor C1. *Proteomics* **13**, 982–991 (2013).
57. <http://planetorbitrap.com/q-exactive-hf#tab:schematic>.
58. Scheltema, R. A. *et al.* The Q Exactive HF, a Benchtop Mass Spectrometer with a Pre-filter, High-performance Quadrupole and an Ultra-high-field Orbitrap Analyzer. *Mol. Cell. Proteomics MCP* **13**, 3698–3708 (2014).
59. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367 (2008).
60. Nagaraj, N. *et al.* System-wide Perturbation Analysis with Nearly Complete Coverage of the Yeast Proteome by Single-shot Ultra HPLC Runs on a Bench Top Orbitrap. *Mol. Cell. Proteomics MCP* **11**, (2012).
61. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins. *Mol. Cell. Proteomics MCP* **11**, (2012).
62. Geyer, P. E. *et al.* Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2**, 185–195 (2016).
63. Meier, F., Geyer, P. E., Cox, J. & Mann, M. BoxCar method enables single shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat Methods* (2018).
64. Orbitrap Mass Spectrometry - Analytical Chemistry (ACS Publications). Available at: <http://pubs.acs.org/doi/10.1021/ac4001223>. (Accessed: 7th December 2017)
65. Chapman, J. D., Goodlett, D. R. & Masselon, C. D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom. Rev.* **33**, 452–470 (2014).

66. Gillet, L. C. *et al.* Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics MCP* **11**, (2012).
67. Guo, T. *et al.* Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.* **21**, 407 (2015).
68. Bruderer, R. *et al.* Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol. Cell. Proteomics* mcp.RA117.000314 (2017). doi:10.1074/mcp.RA117.000314
69. Picotti, P. & Aebersold, R. Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9**, 555 (2012).
70. Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S. & Coon, J. J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics MCP* **11**, 1475–1488 (2012).
71. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 (1999).
72. Cox, J. *et al.* Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
73. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
74. Protein Mass Spectrometry/Proteomics Lecture by Alma Burlingame.
75. Ong, S.-E. *et al.* Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
76. Krüger, M. *et al.* SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell* **134**, 353–364 (2008).
77. Godoy, L. M. F. de. SILAC Yeast: From Labeling to Comprehensive Proteome Quantification. in *Shotgun Proteomics* 81–109 (Humana Press, New York, NY, 2014). doi:10.1007/978-1-4939-0685-7_6
78. Sury, M. D., Chen, J.-X. & Selbach, M. The SILAC Fly Allows for Accurate Protein Quantification in Vivo. *Mol. Cell. Proteomics MCP* **9**, 2173–2183 (2010).

79. Matthes, A., Köhl, K. & Schulze, W. X. SILAC and Alternatives in Studying Cellular Proteomes of Plants. in *Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC)* 65–83 (Humana Press, New York, NY, 2014). doi:10.1007/978-1-4939-1142-4_6
80. Rose, C. M. *et al.* Neutron Encoded Labeling for Peptide Identification. *Anal. Chem.* **85**, (2013).
81. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Methods* **7**, 383 (2010).
82. Zeiler, M., Straube, W. L., Lundberg, E., Uhlen, M. & Mann, M. A Protein Epitope Signature Tag (PrEST) Library Allows SILAC-based Absolute Quantification and Multiplexed Determination of Protein Copy Numbers in Cell Lines. *Mol. Cell. Proteomics MCP* **11**, (2012).
83. Kaiser, S. E. *et al.* Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. *Nat. Methods* **8**, 691–696 (2011).
84. Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A “Proteomic Ruler” for Protein Copy Number and Concentration Estimation without Spike-in Standards. *Mol. Cell. Proteomics MCP* **13**, 3497–3506 (2014).
85. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* **4**, 484 (2009).
86. Ross, P. L. *et al.* Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **3**, 1154–1169 (2004).
87. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003).
88. Ow, S. Y., Salim, M., Noirel, J., Evans, C. & Wright, P. C. Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation. *PROTEOMICS* **11**, 2341–2346 (2011).
89. Rauniyar, N. & Yates, J. R. Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *J. Proteome Res.* **13**, 5293–5309 (2014).
90. Savitski, M. M. *et al.* Measuring and Managing Ratio Compression for Accurate iTRAQ/TMT Quantification. *J. Proteome Res.* **12**, 3586–3598 (2013).
91. Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* **8**, 937 (2011).
92. Wühr, M. *et al.* Accurate Multiplexed Proteomics at the MS2 Level Using the Complement Reporter Ion Cluster. *Anal. Chem.* **84**, 9214–9221 (2012).

93. Ishihama, Y. *et al.* Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol. Cell. Proteomics* **4**, 1265–1272 (2005).
94. Mann, M., Kulak, N. A., Nagaraj, N. & Cox, J. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. *Mol. Cell* **49**, 583–590 (2013).
95. Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55 (2016).
96. Robles, M. S., Humphrey, S. J. & Mann, M. Phosphorylation Is a Central Mechanism for Circadian Control of Metabolism and Physiology. *Cell Metab.* **25**, 118–127 (2017).
97. Andersen, J. S. *et al.* Nucleolar proteome dynamics. *Nature* **433**, 77 (2005).
98. Dunkley, T. P. J., Watson, R., Griffin, J. L., Dupree, P. & Lilley, K. S. Localization of Organelle Proteins by Isotope Tagging (LOPIT). *Mol. Cell. Proteomics* **3**, 1128–1134 (2004).
99. Itzhak, D. N., Tyanova, S., Cox, J. & Borner, G. H. Global, quantitative and dynamic mapping of protein subcellular localization. *eLife* **5**, e16950 (2016).
100. Hein, M. Y. *et al.* A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **163**, 712–723 (2015).
101. Huttlin, E. L. *et al.* Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505 (2017).
102. Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582 (2014).
103. Kim, M.-S. *et al.* A draft map of the human proteome. *Nature* **509**, 575 (2014).
104. Ezkurdia, I., Vázquez, J., Valencia, A. & Tress, M. Analyzing the First Drafts of the Human Proteome. *J. Proteome Res.* **13**, 3854–3855 (2014).
105. Mukherjee, S. *The Emperor of All Maladies*. (Scribner, 2010).
106. www.who.int/mediacentre/factsheets/fs297/en/.
107. Hanahan, D. & Weinberg, R. A. The Hallmarks of Cancer. *Cell* **100**, 57–70 (2000).
108. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
109. Lantz, P. M., Remington, P. L. & Newcomb, P. A. Mammography screening and increased incidence of breast cancer in Wisconsin. *J. Natl. Cancer Inst.* **83**, 1540–1546 (1991).
110. Williams, S. C. P. Link between obesity and cancer. *Proc. Natl. Acad. Sci.* **110**, 8753–8754 (2013).

111. Very rare cancers--a problem neglected. *Lancet Oncol.* **2**, 189 (2001).
112. Loeb, S. *et al.* What is the true number needed to screen and treat to save a life with prostate-specific antigen testing? *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **29**, 464–467 (2011).
113. Cohen, M. H. *et al.* Approval Summary for Imatinib Mesylate Capsules in the Treatment of Chronic Myelogenous Leukemia. *Clin. Cancer Res.* **8**, 935–942 (2002).
114. Druker, B. J. *et al.* Five-Year Follow-up of Patients Receiving Imatinib for Chronic Myeloid Leukemia. *N. Engl. J. Med.* **355**, 2408–2417 (2006).
115. Lemieux, J. *et al.* The Role of Neoadjuvant HER2-Targeted Therapies in HER2 Overexpressing Breast Cancers. *Curr. Oncol.* **16**, 48–57 (2009).
116. Siena, S., Sartore-Bianchi, A., Di Nicolantonio, F., Balfour, J. & Bardelli, A. Biomarkers Predicting Clinical Outcome of Epidermal Growth Factor Receptor-Targeted Therapy in Metastatic Colorectal Cancer. *JNCI J. Natl. Cancer Inst.* **101**, 1308–1324 (2009).
117. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **28**, 1057 (2010).
118. Fraga, M. F. *et al.* Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat. Genet.* **37**, 391 (2005).
119. Choi, J. D. & Lee, J.-S. Interplay between Epigenetics and Genetics in Cancer, Interplay between Epigenetics and Genetics in Cancer. *Genomics Inform. Genomics Inform.* **11**, 164–173 (2013).
120. You, J. S. & Jones, P. A. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* **22**, 9–20 (2012).
121. Vinogradova, M. *et al.* An inhibitor of KDM5 demethylases reduces survival of drug-tolerant cancer cells. *Nat. Chem. Biol.* **12**, 531 (2016).
122. Sidoli, S. & Garcia, B. A. Middle-down proteomics: a still unexploited resource for chromatin biology. *Expert Rev. Proteomics* **14**, 617–626 (2017).
123. Shvartsburg, A. A., Zheng, Y., Smith, R. D. & Kelleher, N. L. Ion Mobility Separation of Variant Histone Tails Extending to the “Middle-Down” Range. *Anal. Chem.* **84**, 4271–4276 (2012).
124. Shvartsburg, A. A., Zheng, Y., Smith, R. D. & Kelleher, N. L. Separation of Variant Methylated Histone Tails by Differential Ion Mobility. *Anal. Chem.* **84**, 6317–6320 (2012).
125. Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb. Perspect. Biol.* **7**, a025064 (2015).

126. Kapoor, A. *et al.* The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. *Nature* **468**, 1105 (2010).
127. Evertts, A. G. *et al.* Quantitative Dynamics of the Link between Cellular Metabolism and Histone Acetylation. *J. Biol. Chem.* **288**, 12142–12151 (2013).
128. Hendriks, I. A. *et al.* Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat. Struct. Mol. Biol.* **21**, 927 (2014).
129. Ruthenburg, A. J., Li, H., Patel, D. J. & Allis, C. D. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **8**, 983 (2007).
130. Yoo, K. H. & Hennighausen, L. EZH2 Methyltransferase and H3K27 Methylation in Breast Cancer. *Int. J. Biol. Sci.* **8**, 59–65 (2011).
131. Velichutina, I. *et al.* EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. *Blood* **116**, 5247–5255 (2010).
132. Kim, E. *et al.* Phosphorylation of EZH2 Activates STAT3 Signaling via STAT3 Methylation and Promotes Tumorigenicity of Glioblastoma Stem-like Cells. *Cancer Cell* **23**, 839–852 (2013).
133. Gonzalez, M. E. *et al.* Histone Methyltransferase EZH2 Induces Akt-Dependent Genomic Instability and BRCA1 Inhibition in Breast Cancer. *Cancer Res.* **71**, 2360–2370 (2011).
134. Chase, A. & Cross, N. C. P. Aberrations of EZH2 in Cancer. *Clin. Cancer Res.* **17**, 2613–2618 (2011).
135. Kondo, Y. *et al.* Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. *Nat. Genet.* **40**, 741 (2008).
136. Knutson, S. K. *et al.* A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nat. Chem. Biol.* **8**, 890 (2012).
137. Cao, J. & Yan, Q. Histone Demethylases Set the Stage for Cancer Metastasis. *Sci Signal* **6**, pe15-pe15 (2013).
138. Turcan, S. *et al.* IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* **483**, 479 (2012).
139. Parsons, D. W. *et al.* An Integrated Genomic Analysis of Human Glioblastoma Multiforme. *Science* **321**, 1807–1812 (2008).
140. Xu, W. *et al.* Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α -Ketoglutarate-Dependent Dioxygenases. *Cancer Cell* **19**, 17–30 (2011).
141. Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* **12**, 463–469 (2011).

142. Shi, Y. *et al.* Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **119**, 941–953 (2004).
143. Rudolph, T., Beuch, S. & Reuter, G. Lysine-specific histone demethylase LSD1 and the dynamic control of chromatin. *Biol. Chem.* **394**, 1019–1028 (2013).
144. Lynch, J. T., Harris, W. J. & Somerville, T. C. P. LSD1 inhibition: a therapeutic strategy in cancer? *Expert Opin. Ther. Targets* **16**, 1239–1249 (2012).
145. Mohammad, H. P. & Kruger, R. G. Antitumor activity of LSD1 inhibitors in lung cancer. *Mol. Cell. Oncol.* **3**, e1117700 (2016).
146. Singh, M. M. *et al.* Preclinical activity of combined HDAC and KDM1A inhibition in glioblastoma. *Neuro-Oncol.* **17**, 1463–1473 (2015).
147. Alsaqer, S. F. *et al.* Inhibition of LSD1 epigenetically attenuates oral cancer growth and metastasis. *Oncotarget* **8**, 73372–73386 (2017).
148. A, B. & W, Y. Arylcycloalkylamines. I. 2-phenylcyclopropylamine. *J Am Chem Soc* **70**, 2198–2201 (1948).
149. Zheng, Y. C. *et al.* Irreversible LSD1 Inhibitors: Application of Tranylcpromine and Its Derivatives in Cancer Treatment. *Curr. Top. Med. Chem.* **16**, 2179–2188 (2016).
150. Binda, C. *et al.* Biochemical, Structural, and Biological Evaluation of Tranylcpromine Derivatives as Inhibitors of Histone Demethylases LSD1 and LSD2. *J. Am. Chem. Soc.* **132**, 6827–6833 (2010).
151. <https://clinicaltrials.gov/ct2/show/NCT02177812>.
152. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **298**, 1912–1934 (2002).
153. Gomperts, B. D., Kramer, I. M. & Tatham, P. E. R. Signal Transduction (Academic Press). (2004).
154. Avruch, J. MAP kinase pathways: The first twenty years. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1773**, 1150–1160 (2007).
155. Pearson, G. *et al.* Mitogen-Activated Protein (MAP) Kinase Pathways: Regulation and Physiological Functions. *Endocr. Rev.* **22**, 153–183 (2001).
156. Whitman, M., Downes, C. P., Keeler, M., Keller, T. & Cantley, L. Type I phosphatidylinositol kinase makes a novel inositol phospholipid, phosphatidylinositol-3-phosphate. *Nature* **332**, 644 (1988).
157. Vanhaesebroeck, B., Stephens, L. & Hawkins, P. PI3K signalling: the path to discovery and understanding. *Nat. Rev. Mol. Cell Biol.* **13**, 195 (2012).

158. Lien, E. C., Lyssiotis, C. A. & Cantley, L. C. Metabolic Reprogramming by the PI3K-Akt-mTOR Pathway in Cancer. *Recent Results Cancer Res. Fortschritte Krebsforsch. Progres Dans Rech. Sur Cancer* **207**, 39–72 (2016).
159. Harvey, J. J. AN UNIDENTIFIED VIRUS WHICH CAUSES THE RAPID PRODUCTION OF TUMOURS IN MICE. *Nature* **204**, 1104–1105 (1964).
160. Kirsten, W. H., Schauf, V. & McCoy, J. Properties of a murine sarcoma virus. *Bibl. Haematol.* 246–249 (1970).
161. Shields, J. M., Pruitt, K., McFall, A., Shaub, A. & Der, C. J. Understanding Ras: 'it ain't over 'til it's over'. *Trends Cell Biol.* **10**, 147–154 (2000).
162. Hoeflich, K. P. *et al.* In vivo Antitumor Activity of MEK and Phosphatidylinositol 3-Kinase Inhibitors in Basal-Like Breast Cancer Models. *Clin. Cancer Res.* **15**, 4649–4664 (2009).
163. Network, T. C. G. A. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330 (2012).
164. Agrawal, N. *et al.* Integrated Genomic Characterization of Papillary Thyroid Carcinoma. *Cell* **159**, 676–690 (2014).
165. <http://cbioportal.org>.
166. Wu, P., Nielsen, T. E. & Clausen, M. H. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* **36**, 422–439 (2015).
167. Where is the Future of Drug Discovery for Cancer? *Cell* **168**, 564–565 (2017).
168. Larkin, J. *et al.* Combined Vemurafenib and Cobimetinib in BRAF-Mutated Melanoma. *N. Engl. J. Med.* **371**, 1867–1876 (2014).
169. Cobimetinib Approved for Advanced Melanoma. *National Cancer Institute* Available at: <https://www.cancer.gov/news-events/cancer-currents-blog/2015/cobimetinib-melanoma>. (Accessed: 10th December 2017)
170. Folkes, A. J. *et al.* The Identification of 2-(1H-Indazol-4-yl)-6-(4-methanesulfonyl-piperazin-1-ylmethyl)-4-morpholin-4-yl-thieno[3,2-d]pyrimidine (GDC-0941) as a Potent, Selective, Orally Bioavailable Inhibitor of Class I PI3 Kinase for the Treatment of Cancer. *J. Med. Chem.* **51**, 5522–5532 (2008).
171. Shi, F. *et al.* The PI3K inhibitor GDC-0941 enhances radiosensitization and reduces chemoresistance to temozolomide in GBM cell lines. *Neuroscience* **346**, 298–308 (2017).

172. Sarker, D. *et al.* First-in-human Phase I study of Pictilisib (GDC-0941), a potent pan-class I phosphatidylinositol-3-kinase (PI3K) inhibitor, in patients with advanced solid tumors. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **21**, 77–86 (2015).
173. Yamamoto, N. *et al.* Phase Ia/Ib study of the pan-class I PI3K inhibitor pictilisib (GDC-0941) administered as a single agent in Japanese patients with solid tumors and in combination in Japanese patients with non-squamous non-small cell lung cancer. *Invest. New Drugs* **35**, 37–46 (2017).
174. Ndubaku, C. O. *et al.* Discovery of 2-{3-[2-(1-Isopropyl-3-methyl-1H-1,2,4-triazol-5-yl)-5,6-dihydrobenzo[f]imidazo[1,2-d][1,4]oxazepin-9-yl]-1H-pyrazol-1-yl}-2-methylpropanamide (GDC-0032): A β -Sparing Phosphoinositide 3-Kinase Inhibitor with High Unbound Exposure and Robust in Vivo Antitumor Activity. *J. Med. Chem.* **56**, 4597–4610 (2013).
175. Lopez, S. *et al.* Taselisib, a selective inhibitor of PIK3CA, is highly effective on PIK3CA-mutated and HER2/neu amplified uterine serous carcinoma in vitro and in vivo. *Gynecol. Oncol.* **135**, 312–317 (2014).
176. Juric, D. *et al.* Phase I Dose-Escalation Study of Taselisib, an Oral PI3K Inhibitor, in Patients with Advanced Solid Tumors. *Cancer Discov.* **7**, 704–715 (2017).
177. Neklesa, T. K., Winkler, J. D. & Crews, C. M. Targeted protein degradation by PROTACs. *Pharmacol. Ther.* **174**, 138–144 (2017).
178. Deshaies, R. J. Protein degradation: Prime time for PROTACs. *Nat. Chem. Biol.* **11**, 634 (2015).
179. Kirkpatrick, D. S. *et al.* Phosphoproteomic characterization of DNA damage response in melanoma cells following MEK/PI3K dual inhibition. *Proc. Natl. Acad. Sci.* **110**, 19426–19431 (2013).
180. Wu, X. *et al.* Activation of diverse signalling pathways by oncogenic *PIK3CA* mutations. *Nat. Commun.* **5**, 4961 (2014).
181. Abelin, J. G. *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* **15**, 1622–1641 (2016).
182. Zhang, T., Schneider, J. D., Zhu, N. & Chen, S. Identification of MAPK Substrates Using Quantitative Phosphoproteomics. in *Plant Pattern Recognition Receptors* 133–142 (Humana Press, New York, NY, 2017). doi:10.1007/978-1-4939-6859-6_10
183. Britten, C. D. PI3K and MEK inhibitor combinations: examining the evidence in selected tumor types. *Cancer Chemother. Pharmacol.* **71**, 1395–1409 (2013).

184. Geyer, P. E., Holdt, L. M., Teupser, D. & Mann, M. Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* **13**, (2017).
185. Fung, E. T. A Recipe for Proteomics Diagnostic Test Development: The OVA1 Test, from Biomarker Discovery to FDA Clearance. *Clin. Chem.* **56**, 327–329 (2010).
186. Ueland, F. R. *et al.* Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. *Obstet. Gynecol.* **117**, 1289–1297 (2011).
187. Rai, A. J. *et al.* Proteomic approaches to tumor marker discovery. *Arch. Pathol. Lab. Med.* **126**, 1518–1526 (2002).
188. Zhang, Z. *et al.* Three Biomarkers Identified from Serum Proteomic Analysis for the Detection of Early Stage Ovarian Cancer. *Cancer Res.* **64**, 5882–5890 (2004).
189. Moore, R. G. *et al.* Evaluation of the Diagnostic Accuracy of the Risk of Ovarian Malignancy Algorithm in Women With a Pelvic Mass. *Obstet. Gynecol.* **118**, 280–288 (2011).
190. Bristow, R. E. *et al.* Ovarian malignancy risk stratification of the adnexal mass using a multivariate index assay. *Gynecol. Oncol.* **128**, 252–259 (2013).
191. Wu, W. & Choudhry, H. *Next Generation Sequencing in Cancer Research, Volume 2*. **2**, (Springer, 2015).
192. Hewitt, S. M. *et al.* Tissue Handling and Specimen Preparation in Surgical Pathology: Issues Concerning the Recovery of Nucleic Acids From Formalin-Fixed, Paraffin-Embedded Tissue. *Arch. Pathol. Lab. Med.* **132**, 1929–1935 (2008).
193. Balgley, B. M. *et al.* Evaluation of Archival Time on Shotgun Proteomics of Formalin-Fixed and Paraffin-Embedded Tissues. *J. Proteome Res.* **8**, 917–925 (2009).
194. Ostasiewicz, P., Zielinska, D. F., Mann, M. & Wiśniewski, J. R. Proteome, Phosphoproteome, and N-Glycoproteome Are Quantitatively Preserved in Formalin-Fixed Paraffin-Embedded Tissue and Analyzable by High-Resolution Mass Spectrometry. *J. Proteome Res.* **9**, 3688–3700 (2010).
195. Wiśniewski, J. R. *et al.* Absolute Proteome Analysis of Colorectal Mucosa, Adenoma, and Cancer Reveals Drastic Changes in Fatty Acid Metabolism and Plasma Membrane Transporters. *J. Proteome Res.* **14**, 4005–4018 (2015).
196. Shi, S.-R., Liu, C., Balgley, B. M., Lee, C. & Taylor, C. R. Protein Extraction from Formalin-fixed, Paraffin-embedded Tissue Sections: Quality Evaluation by Mass Spectrometry. *J. Histochem. Cytochem.* **54**, 739–743 (2006).

197. Wiśniewski, J. R. *et al.* Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* **8**, 611 (2012).
198. Deeb, S. J. *et al.* Machine Learning-based Classification of Diffuse Large B-cell Lymphoma Patients by Their Protein Expression Profiles. *Mol. Cell. Proteomics* **14**, 2947–2960 (2015).
199. Farmer, P. *et al.* A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat. Med.* **15**, 68 (2009).
200. Wang, W. *et al.* Effector T Cells Abrogate Stroma-Mediated Chemoresistance in Ovarian Cancer. *Cell* **165**, 1092–1105 (2016).
201. Calon, A. *et al.* Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat. Genet.* **47**, 320 (2015).
202. Sawyers, C. L. The cancer biomarker problem. *Nature* **452**, 548–552 (2008).
203. Zhang, Z. *et al.* EGFR-mutated lung cancer: a paradigm of molecular oncology. *Oncotarget* **1**, 497–514 (2010).
204. Etzioni, R. *et al.* Early detection: The case for early detection. *Nat. Rev. Cancer* **3**, 243 (2003).
205. Zhang, B. *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382 (2014).
206. Zhang, H. *et al.* Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **166**, 755–765 (2016).
207. Doll, S. & Burlingame, A. L. Mass Spectrometry-Based Detection and Assignment of Protein Posttranslational Modifications. *ACS Chem. Biol.* **10**, 63–71 (2015).
208. Gnad, F., Doll, S., Manning, G., Arnott, D. & Zhang, Z. Bioinformatics analysis of thousands of TCGA tumors to determine the involvement of epigenetic regulators in human cancer. *BMC Genomics* **16**, S5 (2015).
209. Doll, S., Urisman, A., Oses-Prieto, J. A., Arnott, D. & Burlingame, A. L. Quantitative Proteomics Reveals Fundamental Regulatory Differences in Oncogenic HRAS and Isocitrate Dehydrogenase (IDH1) Driven Astrocytoma. *Mol. Cell. Proteomics* **16**, 39–56 (2017).
210. Gnad, F. *et al.* Phosphoproteome analysis of the MAPK pathway reveals previously undetected feedback mechanisms. *PROTEOMICS* **16**, 1998–2004 (2016).
211. Gnad, F. *et al.* Quantitative phosphoproteomic analysis of the PI3K-regulated signaling network. *PROTEOMICS* **16**, 1992–1997 (2016).
212. Doll, S. *et al.* Region and cell-type resolved quantitative proteomic map of the human heart. *Nat. Commun.* **8**, 1469 (2017).

213. Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **166**, 740–754 (2016).
214. Kelstrup, C. D. *et al.* Performance evaluation of the Q Exactive HF-X for shotgun proteomics. *J. Proteome Res.* (2017). doi:10.1021/acs.jproteome.7b00602
215. Falkenby, L. G. *et al.* Integrated Solid-Phase Extraction–Capillary Liquid Chromatography (speLC) Interfaced to ESI–MS/MS for Fast Characterization and Quantification of Protein and Proteomes. *J. Proteome Res.* **13**, 6169–6175 (2014).
216. Torkamani, A., Andersen, K. G., Steinhubl, S. R. & Topol, E. J. High-Definition Medicine. *Cell* **170**, 828–843 (2017).
217. Manyika, J. *et al.* Big data: The next frontier for innovation, competition, and productivity | McKinsey & Company. Available at: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>. (Accessed: 19th December 2017)

VI) Acknowledgments

I would like to acknowledge many people that contributed to this work and made my PhD time an unforgettable, international and fun experience.

First of all, I would like to thank my PhD supervisor, Prof. Matthias Mann that made all of this possible. Thank you for the amazing mentorship over the last years. I am very grateful that you supported my initiative to realize my PhD at three different institutions, including Genentech Inc., the University of California San Francisco (UCSF), and the Max Planck Institute (MPI) of Biochemistry. Thank you for involving me in fascinating projects, your advices, and instructive editing sessions!

I would like to thank my mentor at Genentech, David Arnott, for introducing me to the mass spec world and answering my millions of questions. Thank you Jennie Lill for making my stay at Genentech possible and all the support. Al Burlingame, thank you for having welcomed me at UCSF, all the discussions about the proteomics community, how Hawaii looked like in the 70's, and of course the great wine. I would also like to thank my doctoral examination board members for their time and review.

Many thanks to Philipp Geyer for the best welcome in the pub office, introducing me to the plasma proteome world, all the great laughs, and being a real Bavarian that always supports another French Bavarian.

Florian Meier for putting all my cardiac proteins in big boxes, merci mon ami!

Sebastian with two last names Virreira Winner for making our life EASI and training for the next relay marathon.

Peter Treit for just being Peter, all the support in the lab, feedback, laughs, and updating my music library.

Dan Itzhak for everything that you transform into maps, hours spent in the cold room to create even more maps, helpful input, and your friendship.

Fabian Coscia for the great cancer project collaborations and many more to come!

My former colleague and seatmate, Scarlet Beck, for all the talks, come back to Munich soon!

Thanks to my new colleagues Andreas Brunner and Catherine Vasilopoulou for the great work environment. Ozge Karayel, thanks for the pink Chance. Marta Murgia for helpful discussions about muscles.

Thank you to our amazingly skilled mass spec experts Igor Paron and Korbinian Mayr. Gaby, many Thanks for all the pink columns you made for me. Alison Dalfovo, thanks for all the organizational help.

Thank you Lesca Holdt and Daniel Teupser for the great collaboration and the lovely food.

Nils Brünner and Jose Moreira for the exciting adenoma project collaboration.

I would like to acknowledge my funding from the Studienstiftung des deutschen Volkes that also allowed me to meet wonderful people, in particular my dear friend and colleague Romy Conzade.

I am grateful to have wonderful friends around me: My dear friend Dr. Rebecca Rogers for all the hours we spent together studying, your support without limitation, and being the coolest sustainable resource manager, Anna Buchberger for being the most talented pianist in the world and always being there for me, and Margaux Barthélémy for always making sure I don't forget my French roots.

Last but definitely not least, my biggest thank you goes to my wonderful family. Mum, you are simply the best, I am so grateful for everything you have done for us. Thanks Dad for having made a scientist of me. My brother and sisters Florian, Anaïs, and Elisa - you mean the world to me! Oma und Opa ihr seid die Besten, vielen Dank für alles! Flo, for an unforgettable time in San Francisco and having always supported me. I learned a lot from you and I can't thank you enough for everything you have done for me. I would also like to thank my Banana Queen Aïshi for all the love you gave us, we miss you.