Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto-Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
M.Sc. Katarzyna Buczak
born in: Kraków, Poland
Oral examination: 20.12.2018

# Spatial proteomics: from tissue organization to protein function

Referees: Dr. Peer Bork

Prof. Dr. Matthias Mayer

This study was carried out in the Structural and Computational Biology Unit (SCB) at the European Molecular Biology Laboratory (EMBL) Heidelberg from October 2014 to September 2018 supervised by Dr. Martin Beck

## Summary

Tumor heterogeneity is a term that refers to differences between tumors of the same type in distinct patients as well as to differences observed between cells within a tumor. The later is known as intratumoral heterogeneity (ITH) and is of high clinical relevance, since it directly affects the robustness of prognostic, diagnostics and prediction of biomarkers. Up to date ITH has been mainly investigated at the genomic level. Sequencing of multiple regions from the same cancer specimen have revealed that within a single tumor several clones of cells with distinct mutational landscapes exist, likely as a consequence of clonal evolution. However, ITH can also be driven by differences in the microenvironment that may rather be reflected in differential gene expression or protein turnover than in genomic changes. Nevertheless, to what extent the ITH is manifested on a proteome-wide scale remains largely unknown, mainly due to technical limitations. To overcome these limitations an efficient protocol that allows for proteomic analysis of limited amounts of formalin-fixed and paraffin-embedded (FFPE) material was developed and employed to characterize the proteomic changes in hepatocellular carcinoma (HCC). First, by comparing neoplastic to the adjacent, non-neoplastic tissues, I defined proteomic features that distinguish tumor from peritumoral tissues. The analysis revealed a decrease in abundances of various mitochondrial proteins including components of the NADH dehydrogenase complex I, possibly indicating the metabolic rearrangement in HCC. Subsequently, by analyzing different regions of HCC, I demonstrated the existence of a proteomic heterogeneity, beyond genetic variations, even in morphologically homogenous specimens, which affects various biological processes. Several clinically relevant proteins were identified as differentially expressed across the analyzed tumors or subject to ITH, thus underlying the importance of ITH studies for biomarker discovery and diagnostic applications.

In the second part of my thesis, I focused on the functional characterization of gp210 – a transmembrane component of the nuclear pore complex (NPC). In eukaryotic cells the nuclear envelope constitutes a barrier separating the nucleoplasm and cytoplasm. The transport of macromolecules between these compartments occurs through NPCs which form channels across the inner and outer membrane of the nuclear envelope. Apart from regulating the nucleocytoplasmic transport, NPCs are also involved in the other cellular processes such as chromatin organization, regulation of gene expression or differentiation. The NPC is comprised

of multiple copies of around 30 proteins called nucleoporins (~1000 protein in total). While the stoichiometry of scaffold components is constant across cell lines, differences in the composition of peripheral sites have been observed. One example of a nucleoporin with a cell-type specific expression is gp210. It is a transmembrane nucleoporin that associates with the NPC via its short C-terminal domain. The remaining larger part of the protein is localized within the perinuclear space and it is not required for the interaction with the NPC. The luminal function of gp210 so far has been linked to muscle cell differentiation but apart from this, its role remains largely unknown. In order to investigate the luminal function of gp210, I attempted to draft a map of potential interacting proteins. This was achieved by in-situ proximity labeling combined with mass spectrometry-based proteomics using the so-called BioID approach. Data obtained in BioID experiments indicate a functional link between gp210 and endoplasmic reticulum (ER) related biological functions. I have identified multiple factors involved in the regulation of ER stress and several proteins involved in glycophosphatidylinositol anchor attachment.

## Zusammenfassung

Tumorheterogenität ist ein Begriff, der sich auf Unterschiede zwischen Tumoren des gleichen Typs bei verschiedenen Patienten sowie auf Unterschiede zwischen Zellen innerhalb eines Tumors bezieht. Intratumorale Heterogenität (ITH) beeinflusst die Verlässlichkeit der Prognose, Diagnose und Vorhersage von Biomarkern und deshalb klinisch relevant. Bislang wurde die ITH hauptsächlich auf der genomischer Ebene untersucht. Die Sequenzierung mehrerer Regionen aus der gleichen Patientenprobe hat ergeben, dass innerhalb eines einzelnen Tumors mehrere Klone von Zellen mit unterschiedlichen Mutationslandschaften existieren, vermutlich als Folge einer klonalen Evolution. ITH kann jedoch auch durch Unterschiede in der Mikroumgebung des Tumors beeinflusst werden, die sich eher in der differenziellen Genexpression oder im Proteinumsatz als in genomischen Veränderungen widerspiegeln. Inwieweit sich die ITH Im Proteom manifestiert, bleibt jedoch weitgehend unbekannt, vor allem aufgrund technischer Einschränkungen. Um diese Einschränkungen zu überwinden, wurde ein effizientes Protokoll entwickelt, das die proteomische Analyse von kleinen Mengen an formalinfixierten und in Paraffin eingebettetem (FFPE) Material ermöglicht. Dieses Methode ermöglichte eine Charakterisierung der intratumoralen, proteomischen Veränderungen im hepatozellulären Karzinom (HCC). Durch den Vergleich neoplastischer mit benachbarter nicht-neoplastischer Gewebe habe ich proteomische Merkmale definiert, die den Tumor von peritumoralen Geweben unterscheiden. Eine Konzentrationsabnahme verschiedener mitochondrialer Proteine einschließlich Komponenten des NADH-Dehydrogenase-Komplexes I, weist möglicherweise auf eine metabolische Neuanordnung im HCC hin. In morphologisch homogenen Proben von HCC geht die proteomische Heterogenität über genetische Variationen hinaus und beeinflusst verschiedene biologische Prozesse. Mehrere klinisch relevante Marker waren differentiell exprimiert, im Vergleich mehrerer Tumore oder sogar innerhalb eines Tumors, was die Bedeutung von ITH-Studien unterstreicht.

Im zweiten Teil meiner Arbeit beschäftigte ich mich mit der funktionellen Charakterisierung von gp210 - einer Transmembrankomponente des Kernporenkomplexes (NPC). In eukaryotischen Zellen bildet die Kernhülle eine Barriere, die das Kernplasma und Zytoplasma trennt. Der Transport von Makromolekülen zwischen beiden Kompartimenten erfolgt durch Kernporen (NPCs), die Kanäle über die innere und äußere Membran der Kernhülle bilden. NPCs sind auch

an anderen zellulären Prozessen wie Chromatinorganisation, Regulation der Genexpression oder Differenzierung beteiligt. Der NPC besteht aus mehreren Kopien von etwa 30 Proteinen, die Nukleoporine genannt werden (insgesamt etwa 1000 Proteine). Während die Stöchiometrie der Gerüstkomponenten über Zelllinien hinweg konstant ist, wurden Unterschiede in der Zusammensetzung der peripheren Bereiche der Kernpore beobachtet. Ein Beispiel für ein zelltypspezifisches Nukleoporin ist gp210. Es ist ein transmembranes Nukleoporin, das über seine kurze C-terminale Domäne mit dem NPC assoziiert ist. Der verbleibende, größere Teil des Proteins befindet sich innerhalb des perinukleären Raums und ist für die Interaktion mit dem NPC nicht erforderlich. Die luminale Funktion von gp210 wurde bisher mit der Muskelzelldifferenzierung in Verbindung gebracht, aber abgesehen davon ist seine Rolle weitgehend unbekannt. Um die luminale Funktion von gp210 zu untersuchen, habe ich eine Karte mit möglichen interagierenden Proteinen entworfen. Dies wurde durch in situ Umgebungsmarkierung in Kombination mit auf Massenspektrometrie basierender Proteomik unter Verwendung der sogenannten BioID Methode erreicht. Daten, die in BioID Experimenten erhalten wurden, weisen auf eine funktionelle Verbindung zwischen gp210 und mit dem endoplasmatischen Retikulum (ER) verwandten biologischen Funktionen hin. Ich habe mehrere Faktoren identifiziert, die an der Regulierung von ER Stress beteiligt sind, und verschiedene Proteine, die an der Bindung von Glycophosphatidylinositol Ankern beteiligt sind.

# Table of Contents

## Contributions

The results described in the chapter “Spatial tissue proteomics as a tool to investigate the inter- and intra-tumor heterogeneity of hepatocellular carcinoma” and the corresponding method sections are part of the manuscript:

**Buczak, K**., Ori, A., Kirkpatrick, J.M., Holzer, K., Dauch, D., Roessler, S., Endris, V., Lasitschka, F.,Parca, L., Schmidt, A., Zender, L., Schirmacher, P., Krijgsveld, J., Singer, S., Beck, M., 2018. Spatial Tissue Proteomics Quantifies Inter- and Intratumor Heterogeneity in Hepatocellular Carcinoma (HCC). Mol. Cell. Proteomics 17, 810–825. doi:10.1074/mcp.RA117.000189

Data together with corresponding figures and methods that were not generated by myself are appropriately marked in the text.

# 1. General introduction to mass spectrometry based proteomics

## 1.1. Proteomics

Proteomics is a general term that describes areas of research focusing on large-scale characterization of proteins. This includes identification and quantification of proteins, their conformation, stability and turnover, the analysis of protein-protein interactions, characterization of post-translational modifications (PTMs) and others. The most common technique used in the field of proteomics is mass spectrometry (MS) – an analytical technique that measures the mass-to-charge ratio (m/z) of ionized molecules. In MS based proteomics, the two major approaches can be distinguished, namely the "top-down" and the "bottom-up" proteomics. The "top-down" approach is used to characterize intact proteins. In bottom up proteomic experiments, proteins from entire tissues, cells or purified organelles are solubilized and digested into shorter peptides and these are analyzed by the MS. Usually the trypsin is a protease of choice, mainly due to its high specificity. It cleaves a peptide chain at the carboxyl side of a lysine and arginine (unless followed by a proline). While the top-down strategy is powerful to analyze individual proteins in ensemble, i.e. to probe their conformation, sequence or even interactions, it is difficult to apply to complex protein mixtures. In bottom up proteomics, contextual information on the protein level is lost, however, compositionally complex samples can be analyzed.

## 1.2. Mass spectrometry based proteomics

The principal function of a mass spectrometer is the determination of the m/z values of ions. This in achieved in three steps: (i) ionization of the analyzed molecules, (ii) ion sorting according to their m/z, and (iii) recording of the signal intensity. A plot of intensity versus m/z ratio is called a mass spectrum. The initial role of top-down mass spectrometry in proteomics was limited to the identification of previously purified proteins and analysis of samples of moderate complexity. The number of proteins being identified in a single experiment was limited to few hundreds, primarily due to the available instrumentation. Several innovations led to the more comprehensive analysis of proteomes. The introduction of the electrospray ionization (see

1.2.1) that allows for generating gas-phase ions from a solution was an important steps towards enabling the MS analysis of proteins and peptides with high throughput (Fenn et al., 1989). This was an important prerequisite for the on-line coupling of mass spectrometers with liquid chromatography (LC) in order to enable the analysis of more complex samples by separation of peptides prior to MS analysis (Link et al., 1999). At last, the introduction of the orbitrap mass analyzer in the year 2000 (Makarov, 2000) is considered as a breakthrough innovation in mass spectrometry based proteomics. Thanks to outstanding sensitivity, mass accuracy and nowadays also scanning speed, the orbitrap mass analyzer allows for the identification of thousands of proteins in a single run. Since it was introduced, the proteomes of various organisms have been investigated with in-depth proteomic coverage.

### 1.2.1. Ionization

In MS-based proteomics, peptides are usually ionized using electrospray ionization (ESI). In ESI the solution with an analyte is passed through a thin capillary with a constant flow rate. In order to disperse the solution into fine aerosol, a high voltage is applied to the liquid via the capillary. In the presence of an electric field, the meniscus at the end of the capillary forms a Taylor cone that later transforms into the liquid jet. Due to electrostatic repulsion of charged ions the jet is unstable and collapses into small charged liquid droplets (Figure 1) (Fenn et al., 1989).

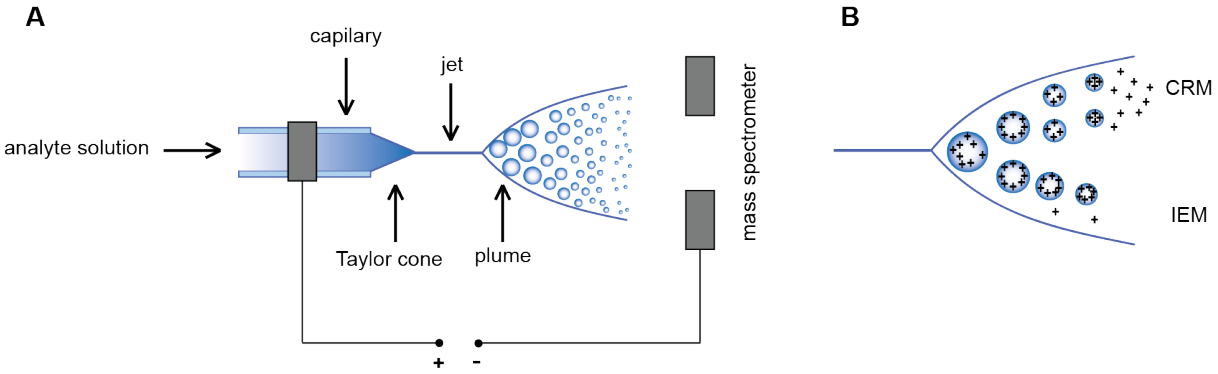



**Figure 1. Electrospray Ionization**
(A) Schematic representation of ESI. (B) Comparison of CRM and IEM models.

There are two models, namely the charge residue model (CRM) and the ion evaporation model (IEM), which propose how ions are transferred from solution to the gas phase. According to the CRM, solvent evaporates from droplets during ESI until the charge density at the surface reaches the Rayleigh limit, which leads to Coulomb fission of the droplet into multiple smaller droplets. This process is repeated until droplets consist of only a single analyte ion and the solvent fully

evaporates (Dole et al., 1968). Alternatively, the IEM has been proposed (Iribarne and Thomson, 1976). In this model droplets undergo fissions until they reach a size of about 20 nm in diameter and then as evaporation proceeds, ions are constantly expelled from the droplet. Currently, it is believed that larger molecules (above 1kDa) can be generated in agreement with the CRM model, while smaller ions can be ejected from nanodroplets according to IEM (Konermann et al., 2013).

### 1.2.2. Mass analyzers

Ionized peptides are subsequently separated in the mass analyzer. Currently, three types of MS analyzers are used in mass spectrometers dedicated to proteomic research. These are quadrupoles, ion traps and orbitraps. The quadrupole mass analyzer is built from four parallel cylindrical electrodes that are positioned around a common axis. The opposite electrodes are electrically paired by applying a radio frequency (RF) and direct current (DC) voltage and generate an electric field, which imposes a specific trajectory on the ions passing through the quadrupole that depends on their m/z. This can lead to one of the following situations: (i) the amplitude of the trajectory is too large and the ions are lost from the quadrupole, (ii) an ion's movement is not affected by the RF voltage leading to the DC voltage causing the ion to crash with the electrode, (iii) the ions have a stable trajectory throughout the quadrupole and thus reach the detector (Figure 2). By adjusting the parameters of the electric field, specific m/z ratios are selected.

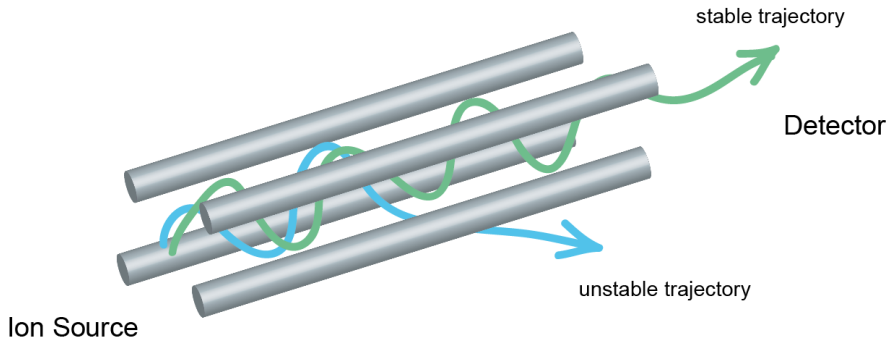



**Figure 2. Schematic representation of the quadrupole mass analyzer**

A quadrupole mass analyzer is built from four parallel rod shaped electrodes imposing an electric field. This can either cause unstable (in blue) or stable (in green) trajectories for analyzed ions.

An ion trap is built from a quadrupole and two hyperbolic cap electrodes positioned on opposite ends of the quadrupole. In an ion trap, the RF voltage applied to the quadrupole electrodes confines the ions radially. At the same time, a static voltage applied to the end cap electrodes

restricts the ions' movement axially. The combination of both RF and DC voltage is used to orient ions in the center of the trap. Changing the RF voltage can be used to alter the stability of ion positioning within the trap and, eventually, is used to expel ions from the trapping zone. To acquire an MS spectrum the RF voltage is adjusted in a controlled manner, thus allowing ions of a defined m/z to leave the trap and finally to be recorded by the detector. Ion traps can also act as mass filters by ejecting all ions from the trap except for the ones with selected m/z (Vachet and McElvany, 1999).

Orbitrap mass analyzers consist of an outer barrel-like electrode and an inner spindle-shaped electrode positioned along the same axis. These electrodes act together to trap ions in an orbital movement around the spindle. First, the electric field between the electrodes is reduced and ion packets are injected into the trap. Introduced ions then start oscillating around the spindle in a movement that can be described by the following parameters: i) frequency of radial oscillations, ii) frequency of rotation around the central electrode, and iii) frequency of oscillations along the z-axis. The last depends on an ion's m/z. As ions are injected, the voltage on the central electrode is ramped up to increase the electric field, thus squeezing ions towards the center of the analyzer, until the ions' radial oscillation is stabilized and they reach their final orbit within the orbitrap. At this point the electric field is stabilized and packets of ions with the same m/z form rings that oscillate along the central axis. The outer electrodes record these oscillations as an image current that is later transformed to a mass spectrum using Fourier transformation (Figure 3).

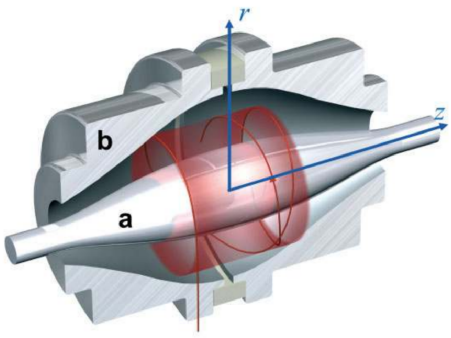



**Figure 3. Orbitrap mass analyzer**

A model of the orbitrap mass analyzer. Ions are oscillating between a central spindle-like electrode (a) and an outer barrel-shaped electrode (b). The frequency of oscillations along z-axis is recorded and further Fourier transformed to obtain a mass spectrum. Modified from (Scigelova and Makarov, 2006)

### 1.2.3. Peptide identification

Even a highly accurate m/z value of a given peptide ion is not sufficient to identify peptides in complex samples. Therefore in proteomics a so-called tandem MS approach (also known as tandem MS/MS) is used. In tandem MS, first the m/z of analyzed peptides is recorded as a full scan (MS1) and then the precursor ion of interest is isolated (either by a quadrupole or within an ion trap) and fragmented by breaking the peptide bonds. Several molecules of the same ion are fragmented, resulting in multiple different fragments. The mass spectra of all of them are then acquired and used for identification (MS2). In order to identify a peptide, measured MS2 spectra of fragment ions (MS2) are matched to a database containing all possible theoretical MS2 spectra of peptides that can be derived from the set of proteins in the database.

One of the most common fragmentation type used in MS based proteomics is collision-induced dissociation (CID) that takes place either in an ion trap or a quadrupole (Cooks, 1995). By raising the RF voltage on the quadrupole, ions are accelerated within the trap and their kinetic energy increases. A collision of accelerated ions with neutral gas molecules induces ion vibrations. As a consequence, a weak chemical bond breaks (usually the peptide bond), dividing the parental ion into so-called –b and –y ions (Figure 4). Mass spectra of the fragment ions are then acquired for each precursor peptide. A drawback of CID fragmentation is the fact that low-mass fragment ions are not efficiently trapped in ion traps (Louris et al., 1987). This is a major limiting factor for experiments where peptides are labeled with isobaric tags (see section 1.3.1), since the reporter ions are usually small and can be lost after fragmentation. To overcome this issue, higher-energy C-trap dissociation (HCD) was developed (Olsen et al., 2007). In HCD, the fragmentation is spatially separated from the ion trap and happens in a gas-tight shroud octopole directly connected to the C-trap (an ion storage prior the orbitrap analyzer). After the fragmentation ions are pushed to the C-trap and later injected into an orbitrap for mass analysis.

The other major type of fragmentation used in proteomics is electron transfer dissociation (ETD) (Louris et al., 1987). Mass spectrometers with ETD capability are equipped with an additional ETD source, in which radical anions are generated through chemical ionization of polycyclic aromatic hydrocarbon molecules. These are then transferred to an ion trap where they can react with multiply charged peptide ions. In the reaction an unpaired electron from the radical is transferred to the peptide ion forming an unstable radical cation with reduced charge that breaks along the N – C alpha bond forming so-called -c and -z fragment ions (Figure 4). ETD fragmentation is frequently used for the identification of post-translational modifications (Wiesner et al., 2008). It has been also shown that ETD is advantageous in cross-linking mass

spectrometry as it enhances identification of peptides covalently linked by MS-cleavable cross-linkers (Liu et al., 2015).

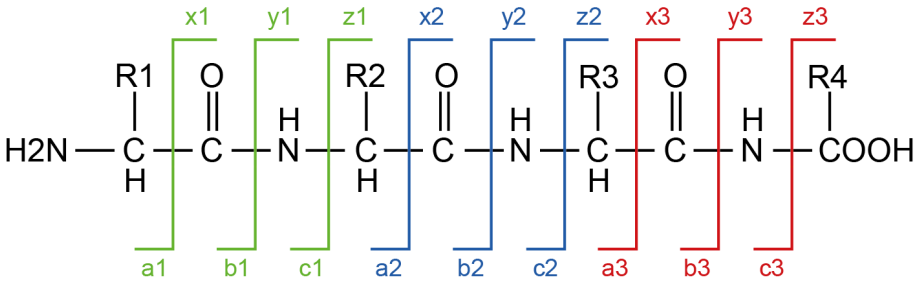



**Figure 4. Peptide fragmentation pattern**

Fragment ions called a, b and c correspond to the N-terminal part of the fragmented peptide, while ions x, y, z correspond to the C-terminal side.

## 1.3. Quantitative proteomics

The general aim of quantitative proteomics is to identify and accurately quantify proteins in a sample, usually in order to compare their abundances across different biological conditions. Various approaches, in which different acquisition and quantification strategies are employed, have been developed to enable such analysis. These can be divided into three major categories: (i) data dependent acquisition (DDA), (ii) targeted proteomics and (iii) data independent acquisition (DIA). The appropriate method from the proteomic toolbox is selected based on the particular question being asked, the number of samples to be analyzed, available instrumentation etc.

### 1.3.1. Data dependent acquisition (DDA) – shotgun proteomics

Data dependent acquisition (DDA), also known as shotgun or discovery proteomics, refers to all types of measurements where ions are selected for fragmentation based on their intensity in the MS1 scan. After each full MS1 scan, the most intense ions are isolated, fragmented and analyzed on MS2 level (Figure 5). This approach is called TopN mode, where N is a number of ions selected for fragmentation from each MS1 scan.

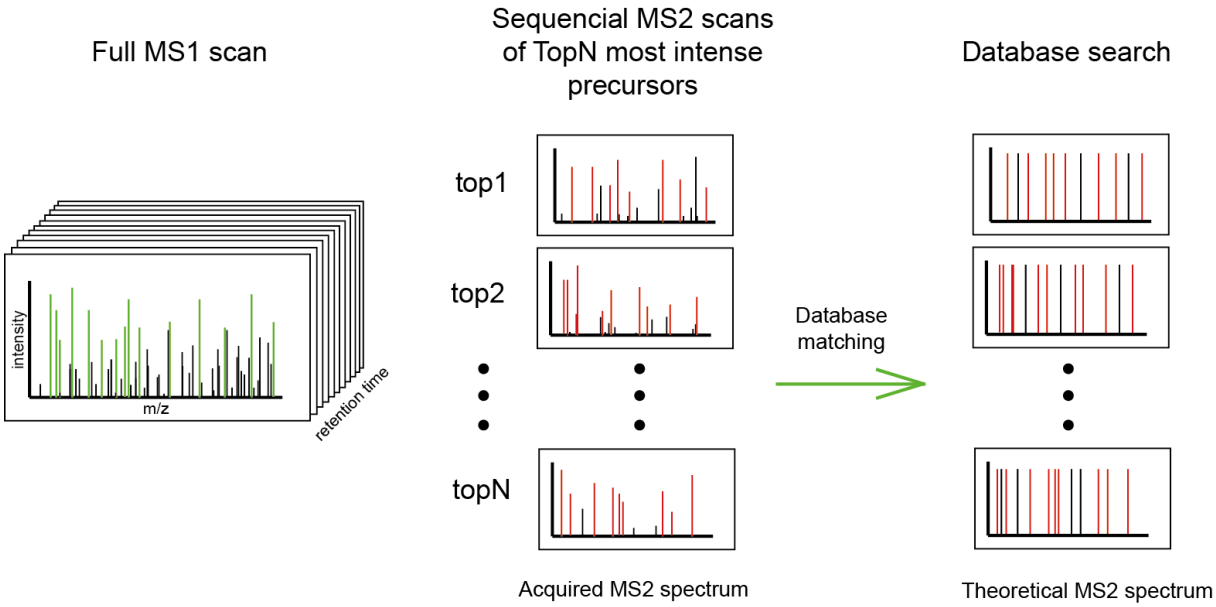



**Figure 5. Data dependent acquisition strategy**

Throughout the entire length of the LC gradient the full MS1 scans are acquired. The top most intense ions (in green) are selected for fragmentation. MS2 spectra of fragment ions are recorded and during the analysis step matched against the database. Red peaks correspond to the ions included in the database. Peaks in black represents ions that cannot be explained by the theoretical spectrum.

Label free quantification

Although each peptide has different properties that affect its overall MS intensity (efficiency of ionization, transfer, and detection etc.), the signal of ions derived from ESI is highly reproducible and the MS intensity of a peptide correlates with its concentration. These two properties are the basis for the label free quantification (LFQ) approach. In LFQ, MS2 spectra are used to identify peptides (using the database search described earlier), while MS1 intensities of corresponding peptides are extracted (based on their m/z and retention time) and used to calculate the relative quantities of peptides. The advantage of LFQ over label-based approaches is its cost and time efficiency and that the number of samples is not limited to the number of possible labels. However, the LFQ approach has certain limitations. Due to the low number of fragmentation events and the semi-stochastic nature of precursor selection, any DDA measurement (and LFQ in particular) suffers from the missing values problem. Because of this, it is not the method of choice for experiments with large sample sizes. Additionally, it is not a suitable method for experiments where deep proteome coverage is desired, since an additional off-line fractionation step is usually required to significantly increase the depth of coverage. Since in LFQ each sample is measured separately, such an approach would drastically elongate the total measurement time.

Label-based quantification

In label-based approaches, peptides derived from different samples are pooled and simultaneously analyzed during a single MS run. This way, errors derived from sample handling and instrument instabilities are minimized and the total measurement time is reduced. One of the first successful labeling approaches was stable isotopic labeling by aminoacids in cell culture (SILAC)(Ong et al., 2002). In SILAC, cells are grown in a culture medium where lysine and arginine are substituted with the same amino acids containing the heavy $^{13}C$ or $^{13}C$ and $^{15}N$ isotopes in all positions of the molecules. Since trypsin cleaves after lysines and arginines, each peptide in a tryptic digest carries at least one heavy labeled aminoacid and can be therefore distinguished from the non-labeled sample based on the defined mass shift in the mass spectrum. Since not all types of samples are amenable for metabolic labeling other methods including either enzymatic (Yao et al., 2001) or chemical (Gygi et al., 1999; Jue-Liang Hsu et al., 2003) labeling have also been introduced. Experiments using isotopic labels allow for simultaneous analysis of up to three conditions at the same time. However, pooling differentially labeled peptides increases sample complexity. Enzymatic and chemical isotopic labeling have been recently almost entirely replaced by isobaric labeling, which does not affect sample complexity (Dayon et al., 2008; Ross et al., 2004; Thompson et al., 2003). An example of such labeling is known as tandem mass tagging (TMT) that currently allows for multiplexing up to 10 samples during a single MS run (Figure 6).

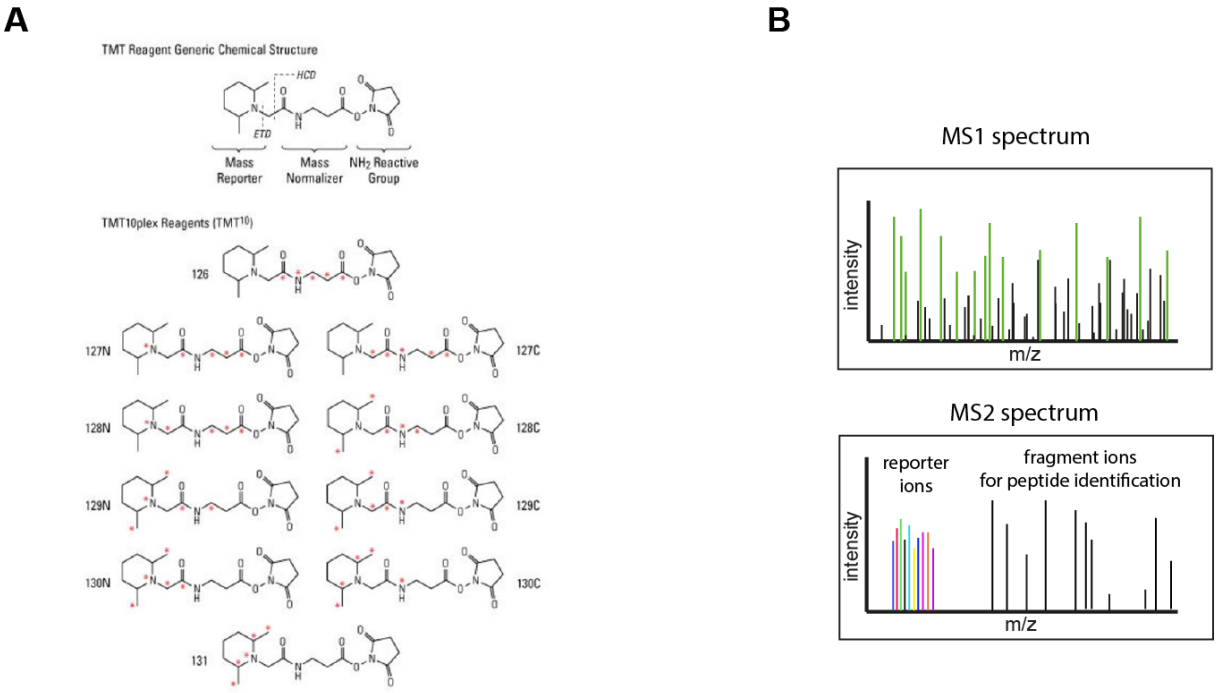



**Figure 6. Tandem mass tagging**

(A) Chemistry of the TMT reagent set. (B) Peptides labeled with different TMT channels are indistinguishable at the MS1 level. Upon HCD, TMT fragments and generated reporter ions can be quantified

In TMT, all 10 tags are different isotopomers of the same reagent therefore they all have exactly the same mass. The tagging reagent is built from three regions: (i) an amine reactive group, (ii) a mass normalizer group and (iii) a reporter group. During labeling, the amine reactive group covalently binds to primary amines (N-termini of peptides or the amine group of lysine residues). Thusly labeled peptides are not distinguishable at the MS1 level, however upon the HCD fragmentation the bond between the reporter group and mass normalizer breaks generating reporter fragment ions of different m/z. The intensities of these reporter ions are used for quantification.

### 1.3.2. Targeted proteomics

In shotgun proteomics a large number of proteins can be identified, however the dynamic range of such measurements is limited. Due to preferential selection of highly intense ions (in DDA), proteins of low abundance are often not detected in such analyses. In contrast to DDA, in targeted proteomics precursor ions are selected for fragmentation based on a predefined inclusion list of m/z values regardless of their intensities. Thus even low abundant proteins can be quantified (Lange et al., 2008). On the other hand, such strategy applies only for hypothesis driven research, since it provides information only about pre-selected proteins. In a classical targeted proteomic experiment (also called selected reaction monitoring, SRM), after LC separation peptides are injected into a mass spectrometer that contains three connected quadrupoles (Q1-Q2-Q3). In the first quadrupole Q1, selected precursor ions are filtered and passed to the second quadrupole Q2 that acts as a collision cell. Here peptides are CID fragmented and transferred to Q3, which selects for predefined fragment ions (Figure 7A). These ions are then passed to a detector, resulting in traces of signal intensity versus retention time for each selected precursor-fragment ion pair (also called SRM transitions). The intensity of a MS2 signal is used for quantification. To increase the total number of peptides that can be measured during a single run, the acquisition of specific transitions is scheduled only around a known peptide elution time (Escher et al., 2012). A similar strategy has been developed for instruments with a quadrupole followed by a high mass accuracy analyzer, like the Q-Exactive (quadrupole – orbitrap). Similarly to SRM, the precursor ion is selected by the quadrupole and passed further for fragmentation. The difference is that all fragment ions are analyzed at once in the orbitrap

mass analyzer, therefore such approach is called parallel reaction monitoring (PRM, Figure 7B) (Gallien et al., 2012). In addition, synthetic heavy labeled peptides of the exactly same sequence are often spiked into samples to act as internal standards for the quantification. Using accurately quantified heavy labeled (AQUA) peptides allows also for the determination of absolute abundance of selected peptides (Kettenbach et al., 2011).

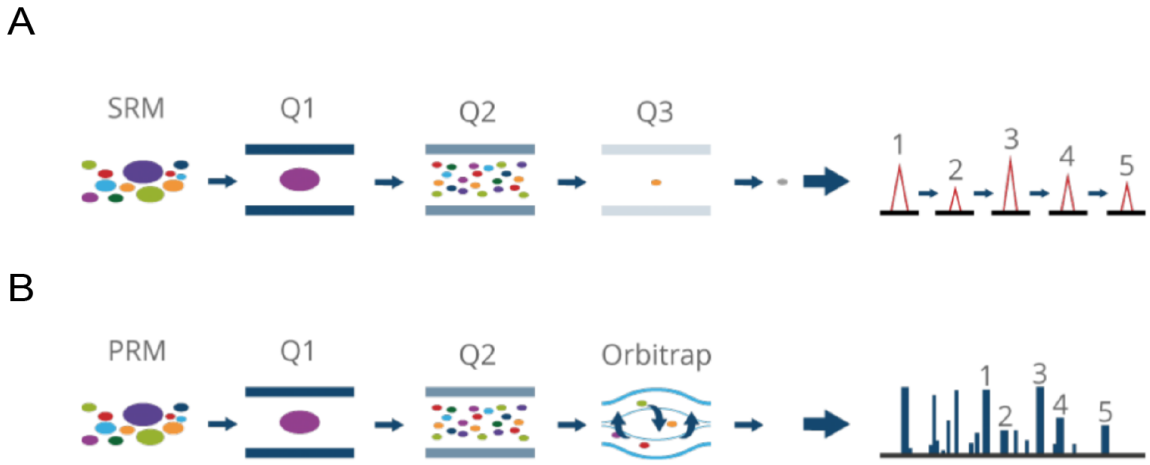



**Figure 7. Targeted proteomic approaches**

In SRM (A) each fragment ion is analyzed separately by the third quadrupole while in PRM (B) are fragments are simultaneously analyzed by the orbitrap mass analyzer. Modified from (Titz et al., 2014).

### 1.3.3. Data Independent Acquisition

Data independent acquisition (DIA), also known as Sequential window acquisition of all theoretical fragment ions (SWATH), has been introduced as an approach that is not affected by missing value issues (like DDA), but at the same time is not limited to only a small number of identified proteins (like SRM) (Doerr, 2015; Gillet et al., 2012). In DIA, each full MS1 scan is divided into consecutive small m/z window scans. In each of these scans, all ions within the given m/z range are isolated by a quadrupole. They are then subjected to HCD fragmentation and then simultaneously analyzed on a high-resolution mass analyzer. Short cycles of isolations are repeated until the entire m/z range is covered (Figure 8). The challenging part of DIA experiments is the data analysis step. Due to the fact that multiple peptides are fragmented together at the same time, MS2 spectra are highly convoluted and as such they cannot be used directly to search against a database. Instead, before a DIA data analysis, a DDA-based spectral library needs to be generated. For this, a representative sample is measured multiple times in DDA mode to obtain sufficient proteomic coverage. All acquired MS files are then subjected to

the standard library based search in order to assign peptides to the experimental fragmentation patterns. Identified MS2 spectra are compiled into the spectral library. During the analysis step, MS2 spectra of DIA files are searched against the previously generated library and fragment ion spectra corresponding to each peptide are extracted from the complex MS2 scans. In order to reduce the computational requirement and ensure confident assignment, fragment ions of particular peptides are only considered in spectra acquired around the expected retention time of the peptide (Bruderer et al., 2015).

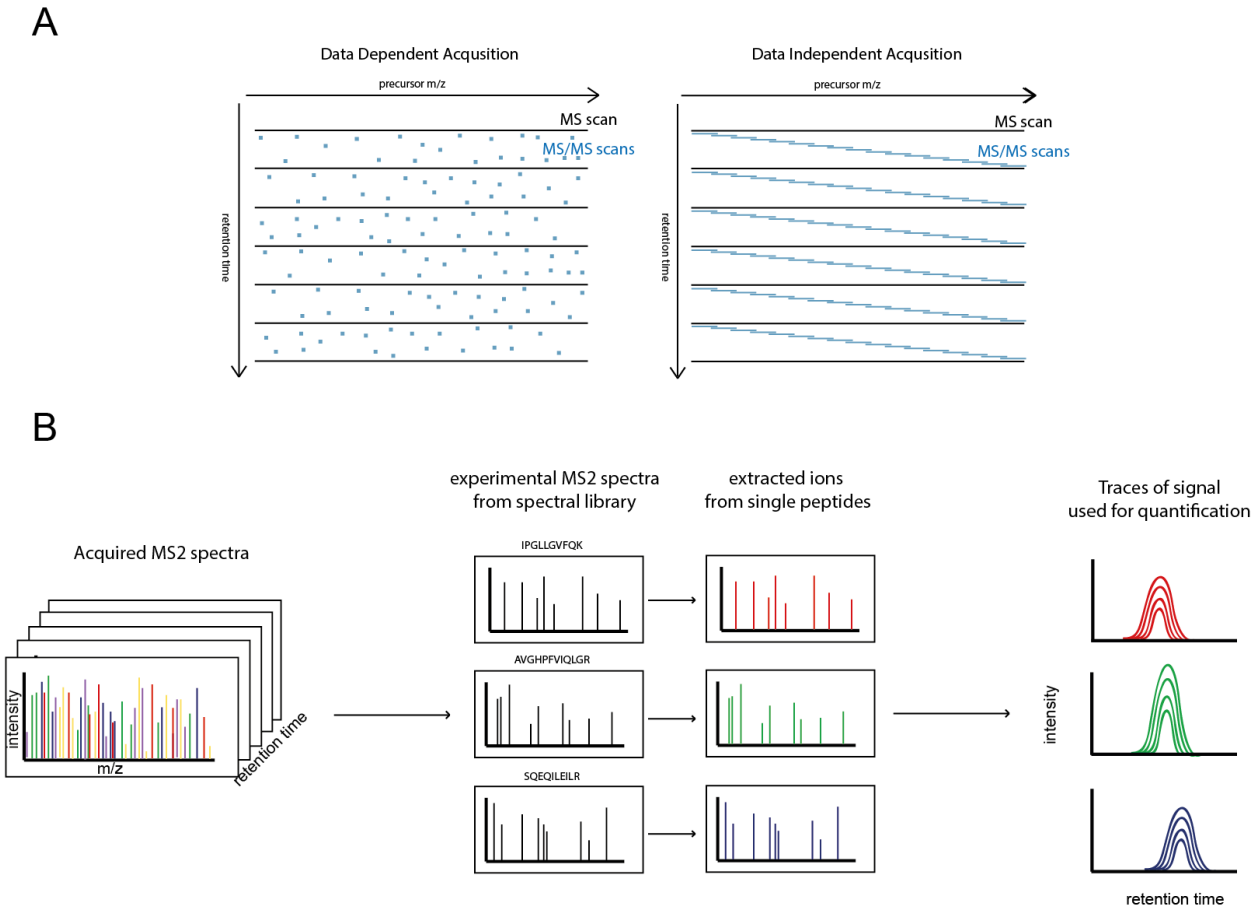



**Figure 8. Data Independent Acquisition Strategy**

(A) Comparison of DDA (on the left) and DIA (on the right) acquisition approaches. In DDA, narrow m/z window is used to isolate only the most intense ion. In DIA, all ions within a wide m/z window are subjected to fragmentation. Cycles of isolation are subsequently repeated until the whole m/z range is covered. (B) Data analysis pipeline. Complex MS2 spectra are matched against a DDA-based peptide library. Fragment ion peptides are extracted based on the experimental fragmentation pattern throughout the predicted retention time resulting in traces of signal intensity versus retention time.

# 2. Spatial tissue proteomics as a tool to investigate the inter- and intra-tumor heterogeneity of hepatocellular carcinoma

## 2.1. Introduction

### 2.1.1. Intratumoral heterogeneity

Although a single mutation is sufficient to initiate tumor formation (Fialkow, 1979), it does not mean that a single tumor entity is comprised of identical cells that are clones of a precursor cancer cell. Most common cancer mutations affect genes involved in processes that regulate cell fate, cell survival and chromosome maintenance (Vogelstein et al., 2013). Because of this, cancer cells divide in a fast, uncontrolled manner and often fail during DNA repair mechanisms. As a consequence, they acquire and propagate mutations more frequently as compared to normal cells. Genetic instability is therefore an intrinsic feature of cancer cells, and distinct cell populations (each carrying its own particular set of mutations) can often be identified within a single tumor mass. This phenomenon is known as intra-tumor heterogeneity (ITH) and has been described for a range of solid tumors as well as for hematopoietic malignancies (Gay et al., 2016). Currently, there are two models to explain the development of ITH, namely the cancer stem cell model and the clonal evolution model. Although slightly different, they do not contradict each other and both are believed to contribute to ITH (Kreso and Dick, 2014; Marusyk and Polyak, 2010). The cancer stem cell (CSC) model defines a tumor as a mixture of cells of different properties. Among these, only certain subsets of cells have features that resemble the characteristics of stem cells: a high capacity of cell-renewal and ability to differentiate into various other cell types. All other cells present within a tumor are considered to be an outcome of the aberrant differentiation of CSCs. In comparison to CSCs, the remaining cells have limited mitotic capacity and are thought not to be involved in tumor progression (Dick, 2008). The experimental evidence underlying the CSC model originally came from a study in which acute myeloid leukemia (AML) cells were sorted based on their surface antigen and transplanted into severe combined immune-deficient (SCID) mice (Lapidot et al., 1994). Only in mice transplanted with certain types of AML cell types ($CD34^+$, $CD38^-$), the formation of re-growing colonies was observed, whereas the remaining ones did not induce tumor growth. Even though the CSC model was originally proposed for hematopoietic malignancies, it has also been shown

to be applicable for multiple solid tumors (Alvero et al., 2009; Li et al., 2007; Schatton et al., 2008; Zhang et al., 2008). Nevertheless, certain aspects of it still remain under debate (Marusyk and Polyak, 2010). It has been shown that variations in experimental conditions can drastically increase the number of cells identified as CSCs, thus suggesting a significant role for the microenvironment in tumor progression (Quintana et al., 2008).

In the clonal evolution model, a tumor is compared to an evolving population that undergoes Darwinian selection (Nowell, 1976). In this model, the whole population of cancer cells frequently acquires random mutations or epigenetic marks, out of which some may have an advantage for tumor progression. As a consequence, cells carrying such advantageous mutations expand faster and overgrow the remaining cells. At the beginning of tumor development usually only few such dominant clones can be distinguished, but as tumorigenesis progresses more branching events can usually be observed (Gerlinger et al., 2012). In the case of clonal evolution, the selective pressure is driven by the complex tumor microenvironment. This includes: (i) systemic regulators of a host (hormones, growth factors, components of immune response etc.), (ii) local conditions such as presence of metabolites, access to oxygen and nutrients, as well as physical constraints which include mechanic pressure from surrounding tissues, composition of extracellular matrix etc. (Greaves and Maley, 2012). As these factors may vary both spatially and temporally, there is no guarantee that a certain mutation will have an equal selective advantage during all stages of tumor progression or within all of the tumor's volume, therefore tumors constantly co-evolve together with their microenvironment (Polyak et al., 2009).

### 2.1.2. Biological implication of the intratumoral heterogeneity

Intratumoral heterogeneity and diagnostics:

In standard pathology workflow, a diagnosis is frequently primarily based on phenotypic traits. The workflow usually combines overall histological analysis of tissue architecture with immunohistological assays that monitor the expression of cancer-specific biomarkers. For example, breast cancer biopsies are routinely tested for the expression of three biomarkers: (i) estrogen hormone receptor (ER), (ii) progesterone hormone receptor (PR) and (iii) receptor tyrosine-protein kinase erbB-2 (HER2) (Hammond et al., 2010). Based on the expression of these receptors, breast cancers can be classified into three major groups: (i) luminal (positive for ER, PR or both), (ii) HER2 positive (positive for HER2 but negative for ER and PR) and (iii) triple negative, in which the expression of none of the markers is observed (Sinn and Kreipe,

2013). This classification has prognostic value (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005) and influences further patient treatment (Harris et al., 2016). However it has been shown that the expression of standard breast cancer markers can be highly variable within a single tumor entity (Nassar et al., 2010; Turashvili and Brogi, 2017) resulting in difficulties during the diagnostic process. To minimize the impact of ITH on tumor diagnostic, several biopsies of the same tumor are being independently inspected, if possible, and the overall diagnosis is based on the region with the most aggressive phenotype. Nevertheless, there is still a risk of undersampling, and as such analyzed regions may not reflect the actual state of the tumor. This is of particular importance especially for larger tumors, where only a small percentage of the total tumor mass is inspected by a pathologist.

Intratumoral heterogeneity and cancer therapies

Even though a tremendous effort is dedicated into cancer research, most malignancies still remain incurable. The only way for successful treatment and prevention of disease recurrence is the entire removal of cancer cells. However in many cases surgical resections are not possible, for example, due to the localization of the tumor close to important organs or blood vessels. Another issue is that tumors are often detected at the metastatic stage, meaning that cancer cells have already migrated into other tissues. In the case of non-operative malignancies, patients are usually treated either with unspecific (chemo- or radiotherapy) or targeted therapies. Although the initial response to the treatment is usually very promising, in many cases the patients relapse and resulting recurrent tumors are no longer sensitive to the same type of treatment. For example, the mechanism of resistance to lung cancer therapy has been described for a patient with a *ROS1* translocation (Awad et al., 2013). In this case study, the patient received treatment with Crizotinib (Sahu et al., 2013), a multi-target tyrosine kinase inhibitor. The treatment resulted in a strong initial response, but a relapse was observed. Sequencing analysis of the recurrent primary tumor as well as metastatic sites revealed a mutation in the ROS1 kinase domain. The substitution (G2032R) caused steric interference affecting inhibitor binding, and therefore made the mutated ROS1 resistant to inhibition by the drug. Similar mechanisms have been observed for the resistance of lung adenocarcinomas to the treatment with epidermal growth factor receptor (EGFR) inhibitors. Here, the treatment frequently results in the T970M mutation in the kinase domain of EGFR leading to insensitivity to the initial drugs (Pao et al., 2005; Sequist et al., 2011). It is frequently proposed, that during therapy, cancer cells acquire mutations providing resistance to the drugs and these are then propagated according to the clonal expansion model(Greaves and Maley, 2012; Iwasa et al., 2005). Other research suggests, that

such mutations occur independently to the drug admission and cells carrying the resistance mutations are already present in pretreated tumors, albeit at very low abundance (Shaw et al., 2016; Yu et al., 2014). This would indicate that tumors are dynamic entities capable of fine-tuning their clonal composition to adjust to the current microenvironmental requirements. This model is supported by the observation that for some patients the post-treatment domination of EGFR inhibitor-resistant clones is only temporal, and few months after the drug withdrawal the T970M mutation is not detectable. Possibly, in the absence of the inhibitor, the mutation is no longer advantageous for tumor progression. Nevertheless, the drug readmission immediately triggers outgrowth of T970M positive cells (Sequist et al., 2011).

### 2.1.3. Methods to study intratumoral heterogeneity

Genetic heterogeneity

First reports regarding the genetic heterogeneity of tumors were based on karyotype analysis and revealed that within a single tumor, different cells may have different types of chromosomal abnormalities (Nowell, 1976). Currently, most ITH studies are based on next generation sequencing (NGS) approaches that allow detecting sequence variants and calculating their frequency within the ensemble of DNA molecules derived from the bulk of a tumor. A whole cancer genome sequencing was reported for the first time for a patient with FAB M1 subtype of AML (Ley et al., 2008). This particular subtype was selected for the analysis due to the lack of signs of heterogeneity, meaning that no chromosomal abnormalities and no somatic copy number alterations had been observed before. Authors of this study identified 10 somatic mutations, out of which one had a significantly lower allelic frequency indicating a presence of more than one clone of AML cells. Later, the same approach was used to characterize a large cohort of both primary and recurrent cases of AML revealing that heterogeneity frequently occurs in AML (Ley et al., 2013). NGS analysis of ITH has also been employed for a number of solid tumors (Jacoby et al., 2015). For example, multiregional whole exome sequencing of renal carcinoma indicated a large degree of clonal heterogeneity but also revealed that the majority of mutations are not uniformly distributed throughout an individual's tumors, thus revealing spatial heterogeneity (Gerlinger et al., 2014, 2012).

Robust whole genome sequencing of tissue biopsies has significantly improved our understanding about tumor evolution and its heterogeneity. Nevertheless, some limitations especially for clinical applications remain. First of all, the detection of low frequency mutations

with whole genome/exome approaches (with allelic frequency > 1%) is still technically challenging (Alizadeh et al., 2015). This may be of high importance for designing targeted therapy strategies in cases where a drug resistant clone is present at very low abundance and cannot be detected with standard approaches. Detection of such mutations can be significantly improved by sequencing only selected genes. For example, a panel of 360 common cancer-related mutations has been assembled and used to investigate ITH in a cohort of breast cancer biopsies (Yates et al., 2015). Another important clinical limitation directly derives from the large degree of intratumoral spatial heterogeneity. As shown for renal carcinoma, not all mutations are evenly distributed and the number of identified genetic variants increases with the number of sequenced regions (Gerlinger et al., 2012). Furthermore, biopsy collection is frequently an invasive procedure and as such there is no possibility for multiple region probing during most diagnostic procedures. As a consequence, there is a high probability of missing important mutations due to undersampling of the tumor. It has been proposed that sequencing of free circulating tumor DNA (ctDNA) obtained through non-invasive liquid biopsies (from blood) could provide a solution to this problem (Dagogo-Jack and Shaw, 2017). Circulating DNAs are small fragments of nucleic acids found in the bloodstream. Although, the exact mechanism of ctDNA release still remains unclear, it has been proposed that it occurs during necrotic and apoptotic events (Siravegna and Bardelli, 2016). Sequencing of ctDNA from breast, ovarian and lung cancer resulted in identification of known mutations that were previously detected from standard biopsies of the same tumors (Murtaza et al., 2013). The same study showed that sequencing ctDNA from liquid biopsies is a suitable tool for rapid detection of drug related mutations and as such could be used to monitor drug response during therapies. In a cohort of 140 patients with metastatic colorectal cancer it has been shown that mutational analysis of ctDNA can identify more mutations than sequencing of tissue biopsies from the same patients (Thierry et al., 2017).

Non-genetic heterogeneity

Genetic heterogeneity provides comprehensive information regarding tumor evolution and its current state of development. The information about clonal composition has been shown to have an impact on both diagnostics and targeted treatment design. In addition, the level of genetic ITH can be easily assessed even from the whole tumor, which minimizes experimental effort. It is therefore understandable that the majority of ITH research focuses solely on genetics. Nevertheless, ITH can be manifested also in different ways. Long-term exposure of a tumor to unfavorable conditions (for example drug treatment) will impose selective pressures and trigger

clonal evolution. However, local or short-term changes in the tumor microenvironment are more likely to be reflected at the transcriptome and proteome level rather than the genome. Due to various technical limitations, however, these types of ITH remain understudied. Thus, understanding non-genetic ITH requires further method development.

Although RNA sequencing (RNA-seq) methods are well established, their applications to study transcriptomic heterogeneity are still limited. Traditional RNA-seq of bulk tumors provides only an averaged expression profile of all cells present in the analyzed sample. Since it is nearly impossible to estimate the contribution of particular cells to the entire pool, information about heterogeneity remains lacking. In addition, unrelated cell types such as infiltrating immune cells or components of the vascular system are frequently present within solid tumors, and thus also contribute to the overall RNA pool. These are likely to be unevenly distributed across the whole tumor and as such random sampling of multiple tumor regions does not necessarily solve the issue of convoluted signals. Single cell RNA sequencing has been proposed as an approach that allows to overcome aforementioned problems (Patel et al., 2014). However, since transcriptomes are highly dynamic it is also possible that cells respond to the stress derived from the cell separation procedure, which may be a source of additional bias.

With the contemporary pathology toolset, however, analysis of heterogeneous protein expression is more feasible and it can be performed also with fixed tissues. Using immunohistochemistry (IHC), the presence of a protein of interest can be visualized with high spatial resolution, which additionally allows direct comparison of protein expression with morphological features. Indeed, heterogeneous protein expression assessed by IHC has been reported for some tumors (Nassar et al., 2010; Turashvili and Brogi, 2017). Importantly, it has been shown that ITH also affects clinically relevant biomarkers, thus underlining the importance of proteomic ITH studies. A major drawback of antibody-based approaches is their low throughput. Even with automated systems, immuno-staining can only be performed for a limited number of targets, for which suitable antibodies are available. MS-based approaches, which could circumvent many of these limitations, and which have been successfully used for general cancer proteomics, have not previously been applied to the study of ITH due to multiple technical limitations (described in section 2.1.6.).

### 2.1.4. Hepatocellular Carcinoma

Hepatocellular carcinoma (HCC) is the fifth most common malignancy worldwide and the second most frequent cause of cancer related death (El-Serag and Kanwal, 2014). Its incident

rate is rapidly increasing with an annual increase of 2%, as estimated in the USA (Stewart and Wild, 2014). Around half of HCC cases are associated with Hepatitis B Virus (HBV) infection and around 25% are the consequence of Hepatitis C Virus (HCV). The most important among the remaining risk factors are alcoholic fatty liver disease and nonalcoholic steatohepatitis (Sanyal et al., 2010). Current therapeutic options for HCC patients include partial hepatectomy or liver transplantation, however less than 20% of HCC patients are amenable for such treatment (Raza and Sood, 2014). The prognosis of HCC patients is very poor, with a five-year-survival rate of less than 5%. Successful therapy accompanied by the increase in survival rate largely depends on the time and accuracy of diagnosis, indicating a strong and urgent need for better understanding of the disease and identification of effective biomarkers. Importantly, at the histomorphological level, the heterogeneity of HCC is less apparent when compared to other solid tumors. Therefore, changes at the protein expression level, if present, are not direct consequences of differences in cellular composition, which make HCC a suitable model system for proteomic ITH studies. In addition, the molecular basis of HCC development and diversity have so far been mainly studied using genomic and transcriptomics approaches (Boyault et al., 2007; Guichard et al., 2012; Roessler et al., 2010; Schulze et al., 2015). However, disease relevant alterations at the proteomic level, particularly in the spatial context, remain poorly defined.

### 2.1.5. Tissue preservation

To prevent degradation of biological material derived from patients, surgically removed tissues need to be preserved as quickly as possible. Preservation happens either through immediate tissue freezing or via fixation with a 4% formaldehyde solution (formalin). Thanks to preservation of non-modified proteins and nucleic acids, freezing is advantageous for molecular biology and biochemical analyses. On the other hand, freezing often affects tissue architecture and morphology, which makes frozen samples not ideal for histological analysis. Additionally, storage of freshly frozen tissues is inefficient in terms of space and cost. Due to these reasons, in most cases freezing is not the preservation method of choice and the availability of these types of samples is limited only to specialized research groups that are in close collaboration with tissue banks and hospitals. In contrast, formalin-fixation and paraffin embedding (FFPE) provides excellent preservation of tissue architecture, which is of high importance especially for ITH studies. FFPE fixation also allows long-term storage at room temperature without any impact on the sample quality (Karlsson and Karlsson, 2011). Because of this, FFPE tissues are the standard

sample type in routine pathology diagnostic workflows worldwide. Since only a small fraction of preserved material is used for diagnostics, the remaining parts are archived and can also be exploited for research purposes. Consequently, FFPE tissues available for researchers are often associated with clinical records and multiple pathology analyses making them a valuable source for cancer research in general.

Formalin fixation is a chemical process in which multiple covalent bonds between nucleic acid, polysaccharides, and proteins are formed (crosslinks). Tissues are immersed in formaldehyde solution that can penetrate around 20 mm of specimen in 24 h (Fox et al., 1985). In terms of protein crosslinking, formaldehyde has been shown to be reactive towards multiple amino acids including lysine, arginine, histidine and cysteine, resulting in the formation of methylene bridges between them (Thavarajah et al., 2012). Such cross-linked tissues are then dehydrated by sequential immersions in several solutions of increasing ethanol concentrations and subsequently embedded in paraffin.

### 2.1.6. Challenges in FFPE tissue proteomics

Although formalin fixation is compatible with histological analyses of tissues, FFPE samples are challenging for biochemical experiments, in particular for proteomics. One of the major reasons why FFPE tissues were considered unsuitable for mass spectrometry for a long time is the presence of multiple unspecific protein modifications introduced during fixation. Since these sites of modification cannot be entirely predicted, they cannot readily be included in databases, and as such modified peptides are not accessible for MS analysis. In addition, the extraction of proteins from such samples is inefficient. Efficient solubilization of FFPE material can be achieved only under very harsh conditions in the presence of highly concentrated detergents (Shen et al., 2015). These are, however, not compatible with downstream MS analysis. In addition, yields of such extractions are usually low and thus, a large amount of starting material is required.

### 2.1.7. Strategies to enable FFPE proteomics

Crosslinking reversal

Apart from standard histological evaluation based on haematoxylin and eosin (H&E) staining, FFPE tissues are frequently subjected to IHC analysis. As a consequence of chemical crosslinking, however, some epitopes may not be accessible to antibodies. To overcome this

problem, several antigen retrieval approaches have been developed. The most common one, known as Heat Induced Antigen Retrieval (HIAR) is based on the heat-induced reversion of formaldehyde crosslinks. It has been shown that incubation of FFPE tissues at high temperatures significantly improves antibody based staining, without affecting tissue architecture (Shi et al., 1991). The high temperature treatment is believed to reverse formaldehyde crosslinks by breaking methylene bridges (Shi et al., 1991). Although the exact mechanisms as well as the efficiency of this process are not known, HIAR is routinely used for IHC of FFPE tissues. Crosslinking reversal is also beneficial for biochemical applications, including mass spectrometry. Reducing the number of crosslinks enhances sample solubilization prior to enzymatic digestion.

Detergent removal

Presence of detergents in samples can affect MS measurements in multiple ways. Common, non-ionic surfactant (e.g. Triton X-100, NP40, or Tween20) ionize with much higher efficiency than peptides. Because of their polymeric nature, detergents characteristically cause multiple regularly spaced peaks in a single MS spectrum. Due to the higher ionization efficiency, the intensity of such signals is often orders of magnitude higher than the signal derived from peptides, therefore surfactants can completely mask the signals of interest. The presence of ionic detergents, such as SDS, interferes with the ionization processes. This leads to weak MS signals and significantly reduced sensitivity of measurements. Additionally SDS, as a harsh ionic detergent, inhibits the enzymatic activity of trypsin and as such needs to be removed from the sample before the digestion step. Due to the presence of hydrophobic chains, surfactants bind to the C18 resin and frequently elute throughout the entire gradient, affecting the whole analysis. Moreover, they also have a tendency to stick to the tubing system, resulting in a long lasting and difficult-to-remove detergent contamination. Due to surfactant's affinity to the C18 resin, they cannot be removed from the MS samples by standard methods used for the peptide purification. It is therefore recommended to entirely avoid using them for MS sample preparation. On the other hand, the use of detergents is beneficial while analyzing membrane proteins. To enhance their solubilization, MS-compatible surfactants have been developed. An example of such a compound is RapigestSF® (Waters), which undergoes hydrolysis in acidic conditions and therefore can be removed from the sample prior to MS analysis. RapigestSF enhances the solubilization of membrane proteins and does not affect the enzymatic activity of trypsin; therefore it is now routinely used in proteomics workflows. It is, however, considered rather as a mild detergent, and hence is not the primary choice for the extraction of cross-linked proteins from FFPE tissues.

Detergents forming micelles of relatively small sizes (such as SDS) can be removed by filter-aided sample preparation (FASP) (Manza et al., 2005). In this approach, samples are first solubilized in the presence of SDS and then urea is added to a final concentration of 8 M. Samples are then placed on an ultrafiltration membrane with a 30 kDa cut-off. Subsequent steps of concentration by centrifugation followed by dilutions with SDS-free urea solution allow for removal of detergent from samples. The presence of urea prevents the precipitation of SDS-soluble proteins. Finally, the concentration of urea is reduced to enable enzymatic digestion. Detergent free peptides can be eluted by an additional step of centrifugation and desalted using C18 resin.

A Single-Pot Solid-Phase-enhanced Sample Preparation (SP3) protocol has been recently introduced as an alternative approach. It utilizes carboxylate-coated paramagnetic beads that can immobilize proteins and peptides on their hydrophilic surface (Hughes et al., 2014). In SP3, a suspension of beads is first added to an aqueous solution of proteins. Upon the addition of an organic solvent, proteins are trapped on the surface of the hydrophilic beads forming a solvation layer. Protein-coated beads can be washed with a range of organic solvents on a magnetic rack. This way detergents, chaotropes and salts can be removed from the sample. After several wash steps, proteins are eluted again into aqueous buffer for digestion. A similar procedure is performed after digestion on the peptide level to remove components of the digestion buffer. An additional peptide labeling step can be added before peptide cleanup, if needed. It is also possible to include a step of peptide fractionation based on their hydrophobicity. This can be achieved by subsequent washes with buffers of decreasing concentration of organic solvent. SP3 has been shown to outperform FASP especially when it comes to small amount of samples (Sielaff et al., 2017), and it has also been applied to the FFPE tissues successfully (Hughes et al., 2016).

Successful proteomic analyses of FFPE tissues have been previously reported. Nevertheless, in the majority of studies comparisons of bulk tumor versus healthy tissues were presented. Proteomic analysis of FFPE tissues has not yet been used to systematically analyze ITH.

## 2.2. Aims

Although various studies have recently demonstrated the power of mass spectrometry based proteomic to study cancer specimens, the potential of the technology to study ITH has not been fully explored. It thus remains a matter of debate, to which extent the genetic and morphological variations are reflected at the proteome level, and whether an additional layer of heterogeneity can be unrevealed by measuring the abundance of thousands of proteins. The aim of this study was to quantify the tumor heterogeneity on the proteome level using hepatocellular carcinoma as a model system. This was achieved with the following steps:

1. Development of the efficient protocol allowing for the proteomic analysis of limited amount of FFPE material
2. Comparison and selection of the most suitable quantification strategy
3. General characterization of the hepatocellular carcinoma proteome
4. Proteomic analysis of different tumor regions derived from the same tumors.

## 2.3. Results

***PART I: Development and validation of a quantitative proteomic approach for FFPE tissues***

### 2.3.1. Reproducibility of the SP3 sample processing workflow for FFPE tissues

During the SP3 procedure, all sample processing steps are performed in a single test tube to minimize sample loss. The approach is therefore suitable for applications where the amount of starting material is limited. Additionally, it allows for the complete removal of reagents that are non-compatible with MS, including commonly used detergents. Detergent-based solubilization facilitates the extraction of proteins from FFPE specimen. The SP3 protocol was therefore adapted for the processing of small amounts of FFPE material. To test the quality of MS samples derived from FFPE material, two consecutive tissue sections (of 10 µm in thickness) of HCC were processed separately for comparison. 25 $mm^2$ tissue areas of both, HCC and adjacent peritumoral tissue, were extracted from each slide, solubilized and further processed using the SP3 protocol. From each sample, 10% of the total peptide material was analyzed by shotgun proteomics using LFQ approach. The quality of the MS data was assessed by quantification across the two consecutive sections as well as by manual inspection of the chromatogram and comparison of the number of peptide identifications. The MS base peak chromatogram (BPC) of FFPE samples looked comparable to HeLa cell lysate (used as a standard quality control), indicating the successful peptide extraction and removal of contaminants, such as salts and detergent. On average, around 20% of the acquired MS/MS spectra were matched against the human proteome database. For comparison, around 50-60% of MS/MS spectra are usually assigned when non-fixed lysates were analyzed (e.g. HeLa cell lysate). This indicates that peptides with chemical modifications were abundant in the analyzed sample and that the reversal of formaldehyde crosslinks was incomplete. As a consequence, the overall number of identified proteins is lower as compared to non-fixed samples. Nevertheless, a high correlation coefficient (>0.95) was observed for protein abundances across both sections as determined by LFQ, thus confirming the reproducibility of approach (Figure 9).

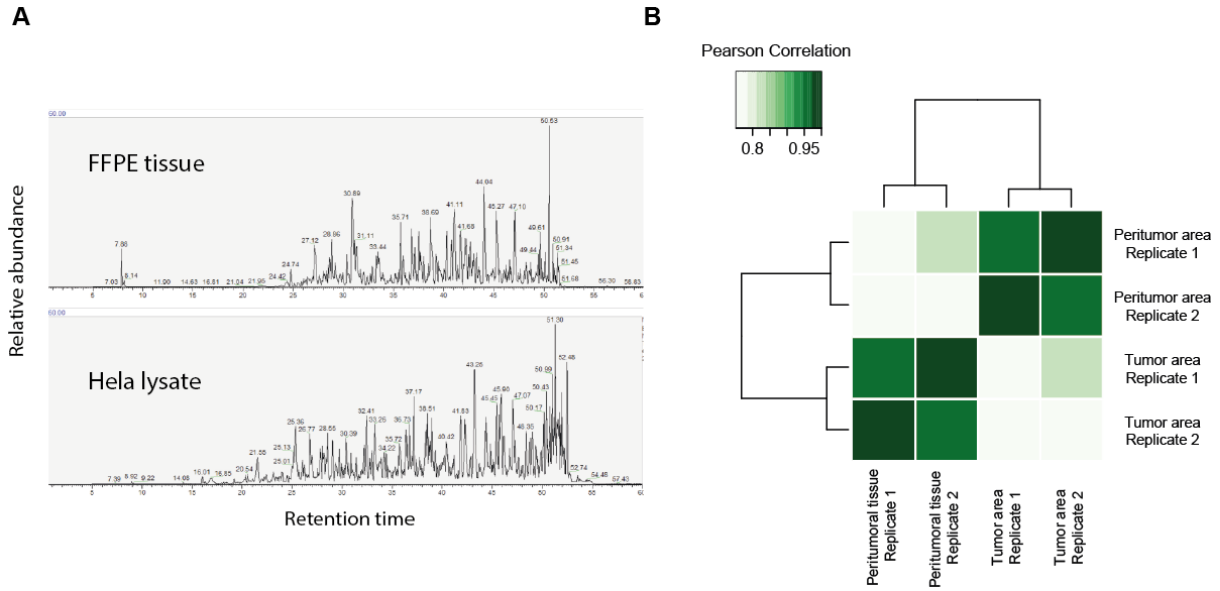



**Figure 9. Quality control of SP3 based FFPE tissue processing**

(A) Base peak chromatograms of peptides extracted from FFPE tissue (top panel) and HeLa cell lysate (bottom panel). (B) Pearson correlation of protein abudances between different FFPE sectors (HCC and peritumoral tissue). High correlation values between replicates of the same tissue area indicate very good reproducibility.

### 2.3.2. Selection of the optimal sample for the analysis of ITH

In order to asses if proteomic ITH goes beyond morphological and genetic heterogeneity, the following approach was used: For several HCC patient samples obtained from the Heidelberg tissue data bank, the level of heterogeneity was assessed based on tissue morphological analysis according to standard diagnostic procedures. This was done by Dr. Stephan Singer, a pathologist from the University of Heidelberg. An HCC specimen with no apparent differences at the morphological level was selected for further analysis as indicated by the H&E stain (Figure 10B). For subsequent proteomic and genetic analysis, half of the encapsulated spherical solid tumor was stained using H&E stain. In total, 5 different tissue sectors were analyzed. These were: (i-iii) three concentric rings within the tumor tissue, (iv) tumor capsule formed by the connective tissue, (v) adjacent peritumoral tissue (Figure10A). The genetic analysis (NGS) was performed by Dr. Volker Endris from the Heidelberg Institute of Pathology. The different tissue areas were subjected to NGS targeting a panel of genes that are frequently mutated in HCC(Guichard et al., 2012). Only three mutations were found in the following genes: *DNA (cytosine-5)-methyltransferase 3A* (*DNMT3A* Y247F), *Myosin Heavy Chain 11* (*MYH11* L1563P) and *Cyclin-dependent kinase 12* (*CKD12* H369R). Nevertheless, the identified mutations showed similar allelic frequency (with the exception of *CDK12*) across all analyzed

tumor sectors indicating that the analyzed specimen is largely homogenous also on the genetic level (supplementary table 1).

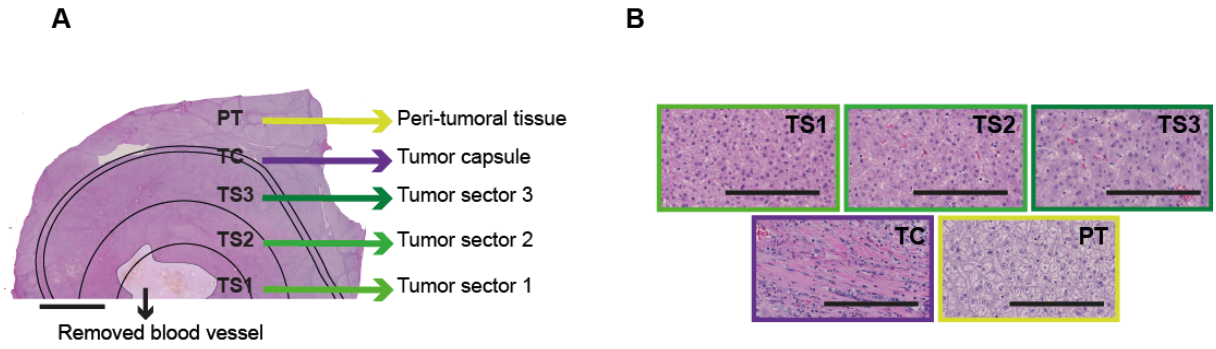



**Figure 10. Morphologically homogenous tumor used for the ITH analysis**

(A) Macroscopic picture of the HE-stained HCC specimen (scale bar = 5 mm). As indicated with the arrows, 5 different tissue sectors were analyzed. These include three tumor sectors, adjacent peritumoral tissue and tumor capsule. A blood vessel localized to the middle of specimen and was removed during LCM procedure (B) Microscopic images (40x, scale bar is 200 µm) of the analyzed sectors. Modified from (Buczak et al., 2018) IHC images were provided by dr. Stephan Singer.

For the proteomic analysis, two consecutive slides of the same tumor sample were subjected to the laser capture microdissection (LCM) and subjected to two different quantitative experiments (TMT and DIA; see materials and methods for detail). The sectors indicated in Figure 10A were collected for the analysis (around 0.4 $mm^3$ of tissue of each sector). To avoid contamination with unrelated cell types, the blood vessel localized in the central part of the specimen was removed.

### 2.3.3. Comparison of different quantification approaches

In this section, I will compare different mass spectrometric quantification experiments, to assess the experimental reproducibility of FFPE tissue analysis but also to comment on which MS workflow permits the most comprehensive analysis of such samples. Each sector was analyzed with both DIA and TMT quantification strategies. For DIA analysis, the generation of a spectral library by label-free DDA (shotgun) analysis is a prerequisite, which also offers the opportunity to compare the DDA and DIA workflows. To this end, I first used a Q-Exactive HF mass spectrometer and analyzed all five sectors by label-free DDA. Specifically, after each full MS1 scan, the 20 most intense ions were isolated for fragmentation. On average 3018 proteins were identified in single DDA run with only 2142 being detected across all measured samples (Figure 11A). Next, the spectra from all runs were combined, resulting in the spectral library of 32681

peptides belonging to 4004 proteins. For DIA analysis, the m/z range of interest (400-1200m/z) was divided into 38 windows. After every full MS1 scan, each m/z window was subsequently isolated and all ions present within a single m/z window were fragmented an analyzed simultaneously in the orbitrap mass analyzer. To minimize the MS2 spectra complexity, the widths of m/z windows were adjusted to the precursor ion density across the m/z range, as reported in (Bruderer et al., 2015). The DIA dataset was almost complete with minimal number of missing values as compared to the spectral library. On average 95% of proteins were identified in single DIA runs and 93% of them were detected across all measured samples (Figure 11B). The DIA workflow thus permits considerable better cross-quantification as compared to label-free DDA.

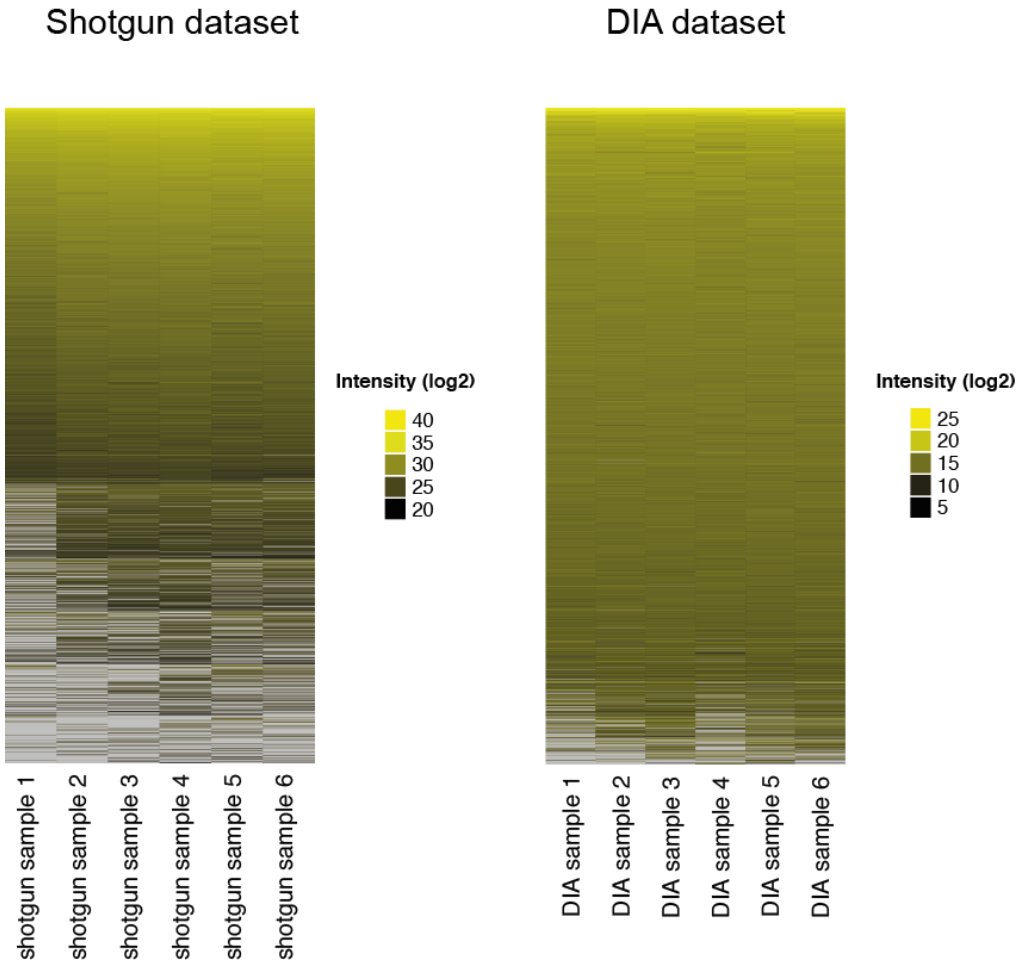



**Figure 11. Shotgun and DIA datasets comparison**

Heatmaps indicating the normalized intensity values for all quantified proteins in (A) shotgun and (B) DIA datasets. White spots represent the missing values. Number of missing values is significantly reduced in DIA dataset when compared to shotgun experiment.

The issue of missing values is to quite some extent addressed by TMT workflows that rely on reporter ions for quantification. To analyze the five sectors for the specimen, a TMT 6-plex reagent was used with the 6$^{th}$ channel containing a repeat sector. To measure the efficiency of labeling, 10% of each sample was analyzed separately by shotgun proteomics with the TMT label being considered as a variable peptide modification. The labeling efficiency was > 95%. For TMT analysis, peptides from all 6 channels were pooled and fractionated using offline high-pH reverse phase fractionation. In total, 16 fractions were collected and analyzed using Orbitrap Fusion mass spectrometer. To minimize the reporter ion ratio distortion resulting from the fragmentation of co-isolated interfering ions, an additional step of fragmentation was used. Specifically, quadrupole-isolated ions were fragmented and their spectrum was acquired in the ion trap. The 8 most intense MS2 fragment ions were isolated using synchronous precursor selection (McAlister et al., 2014) and their MS3 spectrum was acquired using the orbitrap mass analyzer. To enhance the reporter ion signal, LysC was used for enzymatic digestion. LysC cleaves the peptide bonds only after lysine residues but not arginines and, as a consequence, each peptide carries two tags at the primary amines of the N-terminus and the C-terminal lysine residue. In total, 4570 proteins were detected across all analyzed samples, thus outperforming the above-discussed DIA analysis in terms of protein identification.

The DIA and TMT differ in terms of the sample processing after protein extraction, the selection of precursor ions for fragmentation, the quantification methods and the proteolytic digestion. Thus, the directly observed signal for each protein from each sector cannot be directly compared between two datasets. I therefore compared the protein ratios observed across different sectors. A significant positive correlation of $\log_2$-transformed fold changes calculated for different sectors confirms the consistency between DIA and TMT datasets (Figure 12). The application of the two quantitative strategies to the same specimen therefore provided complementary quantitative information and demonstrated the compatibility of presented workflow with both approaches.

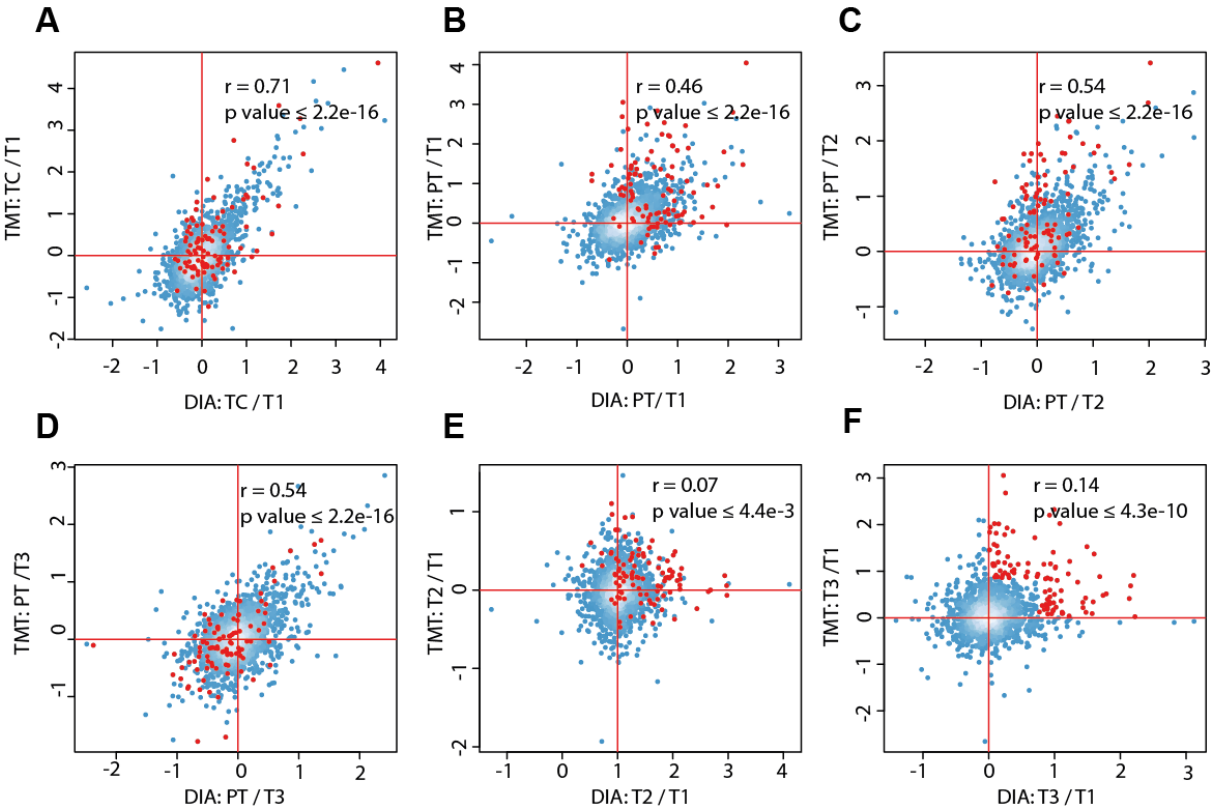



**Figure 12. Comparison of protein fold-changes measured with DIA and TMT approaches**

To compare both datasets, $\log_2$-transformed fold-changes calculated for different sectors were compared. Representative examples are presented (A-F). Differences between non-tumor tissues, either peri-tumor (PT), or tumor capsule (TC), and tumor sectors (TS1-TS3) are more pronounced and show a high degree of linearity (A-D). Differences between tumor sectors (TS1-TS3) are less pronounced. Therefore, the correlation between fold-changes derived from DIA and TMT methods are lower, although still positive (E-F). Proteins identified as differentially expressed between tumor center and its periphery are highlighted in red. Modified from (Buczak et al., 2018)

### 2.3.4. Analysis of proteome profiles of HCC

I first compared the overall proteome profiles across different tissue sectors. For this, I calculated the Pearson correlation of normalized expression values (Figure 13A). High correlation values (>0.9) between tumor sectors and peritumoral tissues were observed. The proteome of tumor capsule formed from the connective tissue can be clearly separated from other sectors mainly due to the high abundance of extracellular matrix proteins, which are absent in other regions. The observed high similarity between the peritumoral tissue and tumor sectors is consistent with the fact that the sample used for the analysis was a well-differentiated HCC and reminiscent of non-tumorous liver tissue in terms of pathological morphology (Figure 10). Nevertheless, soft clustering analysis of protein abundances using the fuzzy c-means algorithm (Kumar and E Futschik, 2007) identified a subset of proteins that clearly separate HCC from the surrounding

non-neoplastic hepatocellular tissue (Figure 13B). I applied the same strategy to identify proteins with heterogeneous expression within the three tumor sectors (other sectors were not included). The soft clustering indicated a prominent subset of proteins with an increased expression level at the tumor periphery when compared to its center (Figure 13B). Taken together this analysis indicated, that despite the overall proteomic similarity, the ITH on the proteome level can be detected with mass spectrometry.

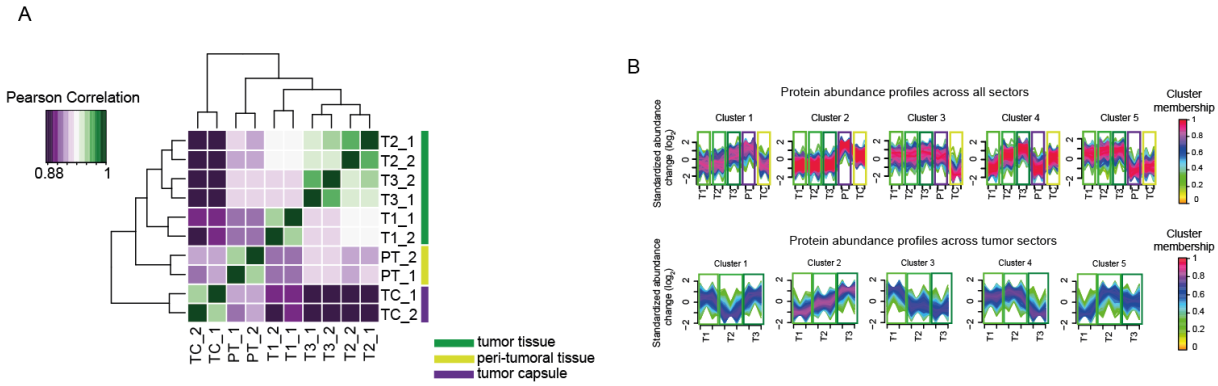



**Figure 13. Proteome profiles of HCC**

(A) Heatmap representing the Pearson correlation coefficient of the analyzed sectors. (B) Soft clustering analysis of HCC spatial proteome using the fuzzy c-means algorithm (Kumar and E Futschik, 2007). The optimal number of clusters was estimated using the "elbow" algorithm (Schwämmle and Jensen, 2010). The upper panel includes all measured sectors (including tumor capsule and peri-tumoral tissue). The lower panel shows clusters calculated only for the three tumor sectors. The DIA dataset was used to create this figure. The TMT data are shown in the supplementary figure 1. Modified from (Buczak et al., 2018)

### 2.3.5. Functional analysis of heterogeneously expressed proteins

In order to assess whether the above identified proteins affected by ITH are functionally related, I performed a network analysis on proteins that displayed heterogeneous expression, as described in the following. As indicated by the proteome profiles (Figure 12B), the most pronounced differences were observed between tumor center (T1) and its periphery (T3). Therefore, T3 vs T1 fold changes were calculated for all the identified proteins. A two component model was then fitted on the centered ratio distribution using fdrtool R package (Strimmer, 2008). Proteins with a q value < 0.2 were considered as differentially expressed between both tumor sectors. TMT and DIA datasets were treated separately (Figure 14).

A B

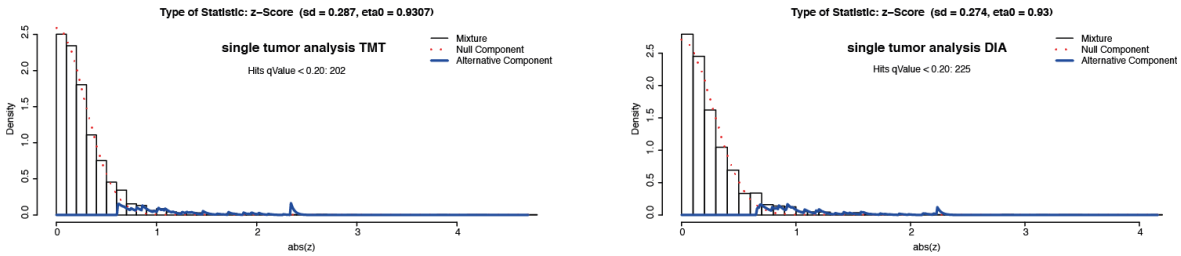



**Figure 14. Statistical models for TMT and DIA datasets**

Histograms and the densities of the fitted two-component models (null component = proteins with no differential expression; alternative component = proteins with differential expression) are shown. Models were fitted on median centered $\log_2$- transformed fold-changes (z). Proteins with q-values < 0.2 were considered as differentially expressed. Modified from (Buczak et al., 2018)

To combine the quantitative information from both datasets, I extracted proteins that were: (i) consistently differentially expressed in both datasets (q value < 0.2), (ii) differentially expressed in only on of the datasets (q value < 0.2) but not detected in the other or (iii) classified as differentially expressed in only one of the datasets (q value < 0.2) with the $\log_2$ of fold change $>|1|$ and the same fold change sign in the other one. In total, 230 proteins were extracted (Figure 12F) and analyzed with the STRING database (Jensen et al., 2009) using high confidence score (>0.7). The thus derived network of protein-protein interactions was further analyzed with cytoscape. Specifically, the MCODE plugin (Bader and Hogue, 2003) was used to extract the sub-network modules that were subsequently subjected to the functional enrichment analysis using ClueGO (Bindea et al., 2009). A list of all the identified proteins was used as a background gene list. A strong enrichment of ribosomal proteins was observed at the tumor periphery, possibly indicating the higher translational activity. Such heterogeneous expression patterns were also detected for proteins involved in the regulation of cell migration. For example, the small GTPases RAC1 and CDC42 that control the formation of lamellipodia and filipodia together with the regulators of actin cytoskeleton dynamics, like Actin-related protein 3 (ACTR3), showed higher expression levels at the tumor periphery. Also proteins that regulate the cell migration at the extracellular level, such as Integrin beta 1 (ITGB1) and Integrin beta 2 (ITGB2), displayed similar expression profiles.

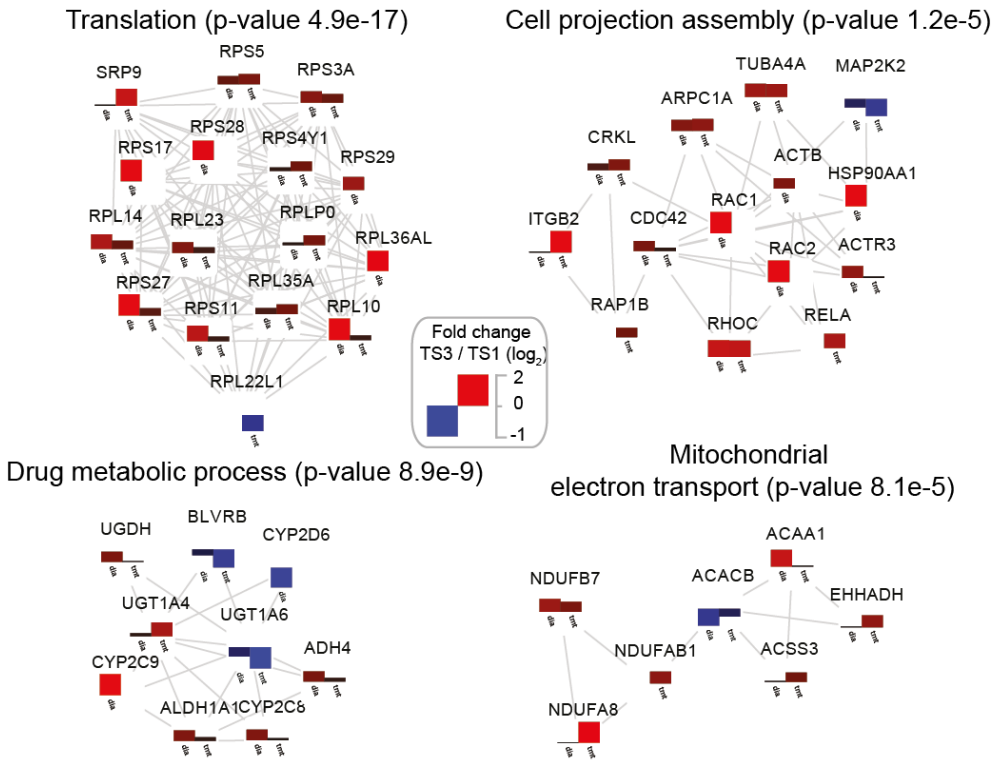



**Figure 15. Network analysis of proteins differentially expressed across the HCC**

Proteins differentially expressed across the HCC specimen were subjected to a network analysis, as described in the main text. The presented network modules were extracted with MCODE. The GO term enrichment within each module was assigned with ClueGO. The barplots displayed next to each network node indicate the calculated fold-change for both, the DIA (left) and TMT dataset (right). Modified from (Buczak et al., 2018)

### 2.3.6. Validation of proteomics data with immunohistochemistry

The IHC analysis presented in this section was performed by the tissue bank of the National Center for Tumor diseases (NCT) Heidelberg.

The differences in the number of ribosomes across the tumor sectors may be a consequence of different proliferation rates in distinct tumor areas. This could potentially affect the proteome profiles. Thus, to rule out this additional source of heterogeneity the expression level of Ki-67, which is a proliferation rate marker protein (Gerdes et al., 1983), was assessed. Since it was not detected in the MS analysis, an immunohistochemical staining (IHC) was performed. As shown in the Figure 16, in all sectors similar low proliferation rates were observed. IHC was also used to validate the mass spectrometry data. A subset of proteins with different expression patterns was selected for this analysis: (i) Hepar-1 antigen, marker of hepatocellular tissues (Butler et al., 2008), as a protein with even expression (ii) Decorin, which was strongly enriched in the tumor capsule, (iii) Rac1, showing a gradient of increased expression towards the periphery. Light

microscopy images presented on the Figure 16 indicate that the IHC is consistent with proteomic data.

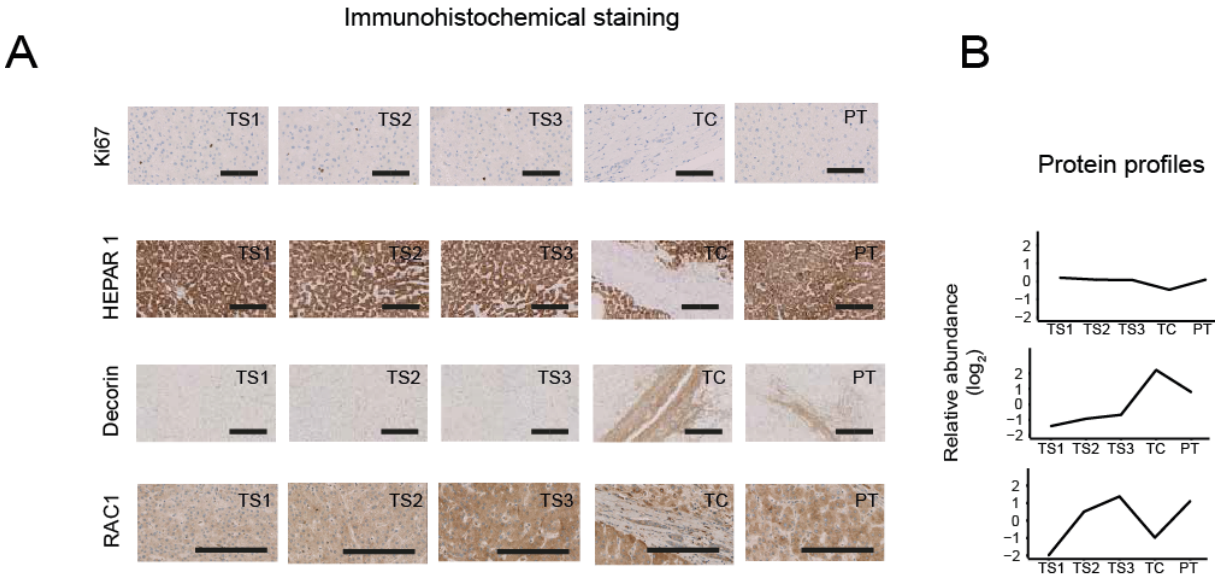



**Figure 16. Immunohistochemical validation of proteomic data**

(A) Consecutive tissue sections of the analyzed specimen were stained with antibodies against Ki67, Hepar-1 antigen, Decorin and RAC1 (scale bars = 200um). IHC images were provided by dr. Stephan Singer (B) Expression profiles based on the proteomic dataset for comparison. Modified from (Buczak et al., 2018).

***PART II: Comprehensive analysis of HCC proteome:***

In the first part of my work I showed that the analysis of very limited amounts of FFPE material with deep proteomic coverage is possible and that the developed approach is sensitive enough to detect proteomic intratumoral heterogeneity. As a next step, I analyzed the HCC proteomes from 5 additional patients (for patient characteristics see supplementary table 2). Like in the previous analysis, only morphologically homogenous tumors were analyzed. In this chapter, I discuss the proteomic analysis of HCC across different patients that includes on one hand tumor versus peritumor comparison and on the other a comparison of different sectors within each sample. For the first part, I will also refer to other datasets that were not generated by me. One of them is the publicly available gene expression data, which was derived from 241 HCC patients (Roessler et al., 2010). I also compared human HCC proteome with proteomes derived from murine HCC models. These models were generated by transposon-based gene transfer of different oncogenes

(N-rasG12V, Myc, and myriostylated Akt1) into wild type mice, as well as into mice with homozygous or heterozygous deletions of the tumor suppressor genes CDKN2aARF and Trp52 (Dauch et al., 2016). In total, 12 fresh frozen tissues of murine livers were provided by Dr. Daniel Dauch and Prof. Lars Zender from the University Hospital Tübingen (11 HCC livers and one control). The proteomic data of murine liver tissues that are mentioned below was obtained by Dr. Alessandro Ori using a label free quantification approach.

### 2.3.7. Proteomic comparison of tumor and surrounding peritumoral tissue

I separated the bulk tumor from the adjacent peritumoral tissue using LCM. To avoid contamination with unrelated cell types, I removed the tumor capsule, fibrous septa, and blood vessels where relevant. 5 tumors and 5 peritumors were combined into a single TMT 10-plex experiment. Peptides were fractionated using high pH reverse phase chromatography and measured on an Orbitrap Fusion Lumos. This analysis quantified 5838 proteins with at least two proteotypic peptides across the respective conditions. Pearson correlation (Figure 17A) indicated that peritumoral tissues are highly similar across individuals and can be distinguished from malignant areas. In contrast, correlation values for the analyzed tumors were generally lower, indicating the high level of inter-patient heterogeneity. One of the samples, labeled as tumor 1, appeared as an outlier and was more similar to the peritumoral tissues of the other specimens. Motivated by this finding, Dr. Stephan Singer inspected the respective clinical records and found that the corresponding patient received transarterial chemoembolization (TACE) treatment before the surgery. Since the impact of TACE treatment on the HCC proteome is not known, this sample was excluded from some of the following analyses (as specified below). Also for the other 4 specimens, a relatively high degree of inter-tumor heterogeneity was observed. I therefore decided quantitatively analyze each tumor-peritumor pair separately.

To investigate which proteins and functional modules are affected by HCC, I performed the gene ontology (GO) enrichment analysis using single ranked list algorithm (Eden et al., 2009). The list of identified proteins was sorted based on the calculated fold-change (tumor vs. peritumor). Enrichment of GO-terms in “biological processes” category was calculated for both up- and down-regulated proteins for each tumor-peritumor pair separately. The redundancy of the identified GO-terms was reduced using REVIGO (Supek et al., 2011). Although for each sample a slightly different set of enriched GO-terms was retrieved, certain similarities were observed. For example, among the proteins up-regulated within the tumors, I found multiple proteins involved in the transcription and RNA processing, suggesting higher rates of transcription in the

cancer cells. Among the down-regulated proteins, I found enrichment for several mitochondria related processes possibly indicating an alteration of the energy metabolism in hepatocellular carcinoma (Figure 17B).

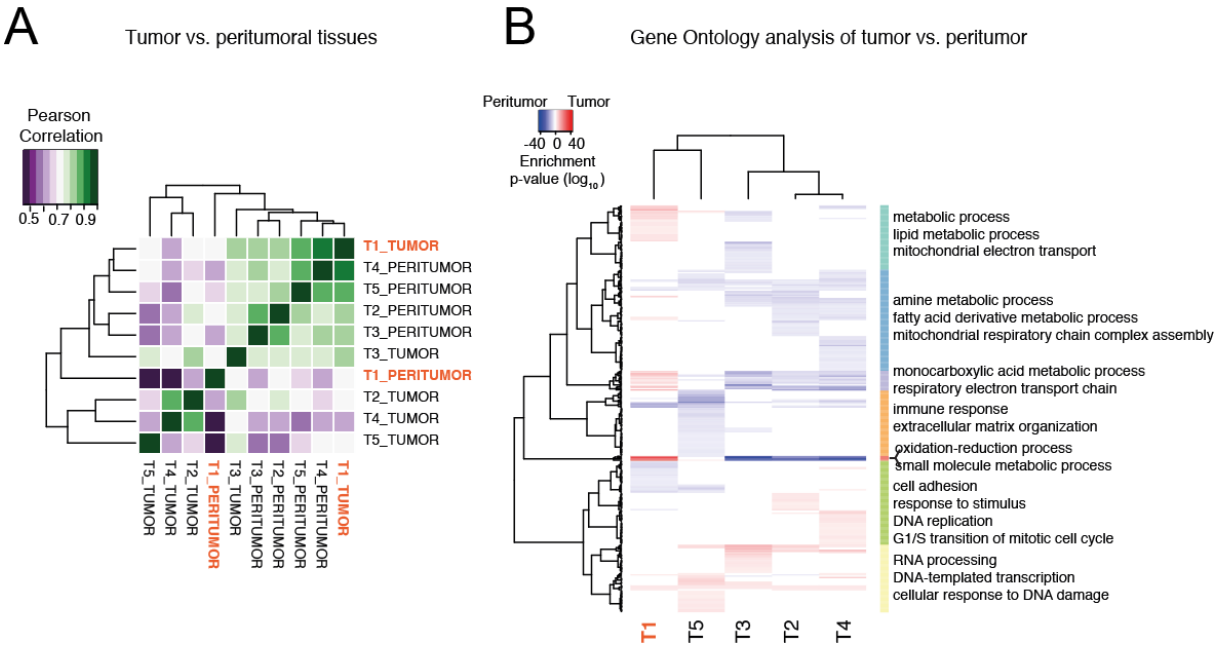



**Figure 17. Tumor vs. peritumor comparison**

(A) Pearson correlation between samples used for neoplastic vs. non-neoplastic tissue comparison. (B) Ranked gene ontology enrichment of proteins differentially expressed in tumor vs. peritumor comparison. The blue color corresponds to proteins that were down-regulated in the tumor. The red color indicates up-regulated proteins. Representative terms from each cluster are displayed next to the heatmap. Modified from (Buczak et al., 2018)

### 2.3.8. Comparison of proteomic and transcriptomic changes in HCC

As a next step, I attempted to derive an expression signature of HCC containing proteins and genes that are differentially expressed in the HCC as compared to non-neoplastic tissue. For this purpose, I integrated publicly available gene expression data (Roessler et al., 2010) with the proteomic data of human samples and the proteomic data of genetically defined HCC mouse models that closely resemble human hepatocarcinogenesis (Dauch et al., 2016), as follows. To define the HCC signature, I combined genes and proteins that showed consistent fold changes in gene expression and protein levels in both: tumor versus peritumor tissues, in both patients and murine models. I did not take into consideration Tumor 1 because it appeared as an outlier in previous analysis. Due to the low coverage of murine proteomes, missing values were allowed for this dataset. In total 755 proteins were selected, out of which some had already been previously linked to the HCC development (Figure 18A). For example, multiple components of

minichromosome maintenance complex (MCM complex) were found to be expressed at higher levels in the analyzed tumors (Figure 18B) and the up-regulation of MCM proteins has been previously linked with multiple cancers including HCC (Das et al., 2014). In fact, one of the subunits, namely MCM6, was proposed as a novel HCC marker (Zheng et al., 2014). Other proteins there were not yet connected to the HCC development, such as fatty acid binding protein (FABP4), which is involved in the lipid transport or Zing finger protein 207 (ZFP207), which is a kinetochore and microtubule binding protein. These findings indicate that the acquired dataset does not only recapitulate multiple known HCC factors, but also identify new proteins that may be involved in the process of liver tumorigenesis. In particular because of the relatively small number of human patient samples included into my study, these however need to be further investigated in the future.

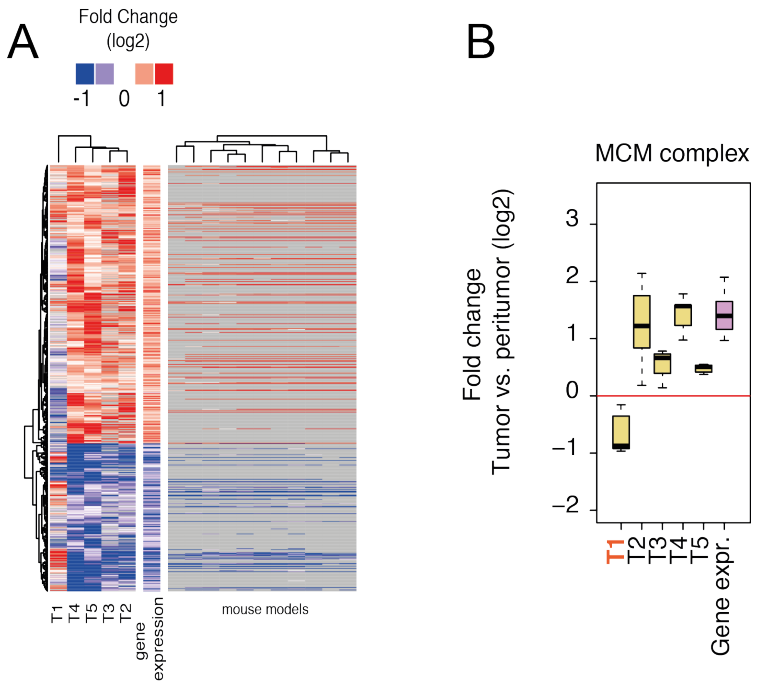



**Figure 18. HCC expression signature**

(A) Heatmap highlighting the expression changes of 755 proteins identified as regulated in HCC. T1 was not considered for this analysis but included in the heatmap for comparison. (B) Boxplot indicating the expression changes of MCM protein complex. All subunits were included. Modified from (Buczak et al., 2018)

Since the gene expression data had been previously explored for its prognostic value (Roessler et al., 2010), I also tested whether the identified signature correlated with a more aggressive tumor phenotype. Since the latency of analyzed murine models was previously defined, I first correlated it with the expression of proteins included in the proteomic signature. A significantly

higher expression of the respective proteins in HCC models of 4 weeks latency in comparison to the moderately aggressive HCCs with the latency of 8-12 weeks was observed (Figure 19A). In the following, it was also tested if the corresponding expression profiles of the transcript levels were also associated with the poor clinical outcome in HCC patients (this part was done by dr. Stephanie Roessler). Survival risk prediction using the number of 755 genes included in the signature derived from the combined transcriptomic and proteomic analysis resulted in a high (n=122) and low (n=119) risk group of patients (permutation p=0.001). Indeed, as the Kaplan-Meier survival curves demonstrate that HCC patients defined as high risk by the proteomic signature had significantly shorter overall survival than low-risk patients (lon-rank p-value < 0.001, Figure 19B)

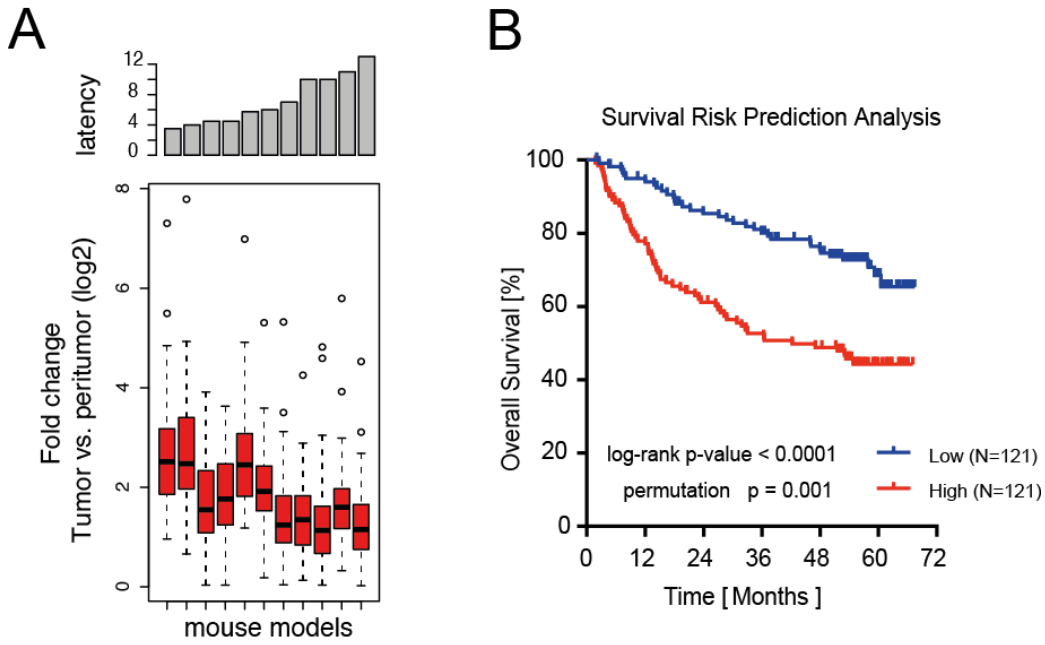



**Figure 19. Survival risk prediction based on the HCC expression signature**

(A) The boxplot (below) shows fold changes of up-regulated proteins from the HCC signature across different murine models. The barplot (above) indicates the tumor latency of corresponding models (given in weeks after cells injection), as proxy for tumor aggressiveness. (B) Survival risk prediction analysis was performed on a cohort of 241 human HCC patients. The cohort was dichotomized into high and low risk patients according to the prognostic index derived from the HCC gene signature. Plot on the panel B was provided by dr. Stephanie Roessler.

### 2.3.9. Analysis of proteomic changes that are not reflected at the gene expression level

Next, I asked whether the proteomic data could reveal biological insights that would not emerge from the solely from the transcriptomic data. This analysis was performed using HCC proteomic data (excluding tumor 1) and previously aforementioned gene expression dataset of large cohort

of HCC patients. In order to define proteins that are consistently differentially expressed, I calculated the average fold change for all quantified proteins together with their statistical significance. Only hits with q value < 0.2 were considered as differentially expressed (supplementary figure 3). From these, I further selected for hits that showed either no or modest change at the gene expression level (with the log2-transformed fold change < |0.5|). In total 148 proteins were extracted (Figure 20). Among them I observed several members of mitochondrial NADH dehydrogenase complex I, which is a part of mitochondrial respiratory chain. Although remaining components of the complex did not pass the significance scores, for a majority of them negative values of fold changes were observed, possibly indicating rearrangements in the inner mitochondrial membrane or general differences in subcellular compartmentalization. In order to verify whether this was a common feature of various types of HCC, I also compared the expression of NADH in murine models. Consistently with the human samples, the majority of NADH dehydrogenase complex I components were expressed at lower level in HCC when compared to the healthy liver.

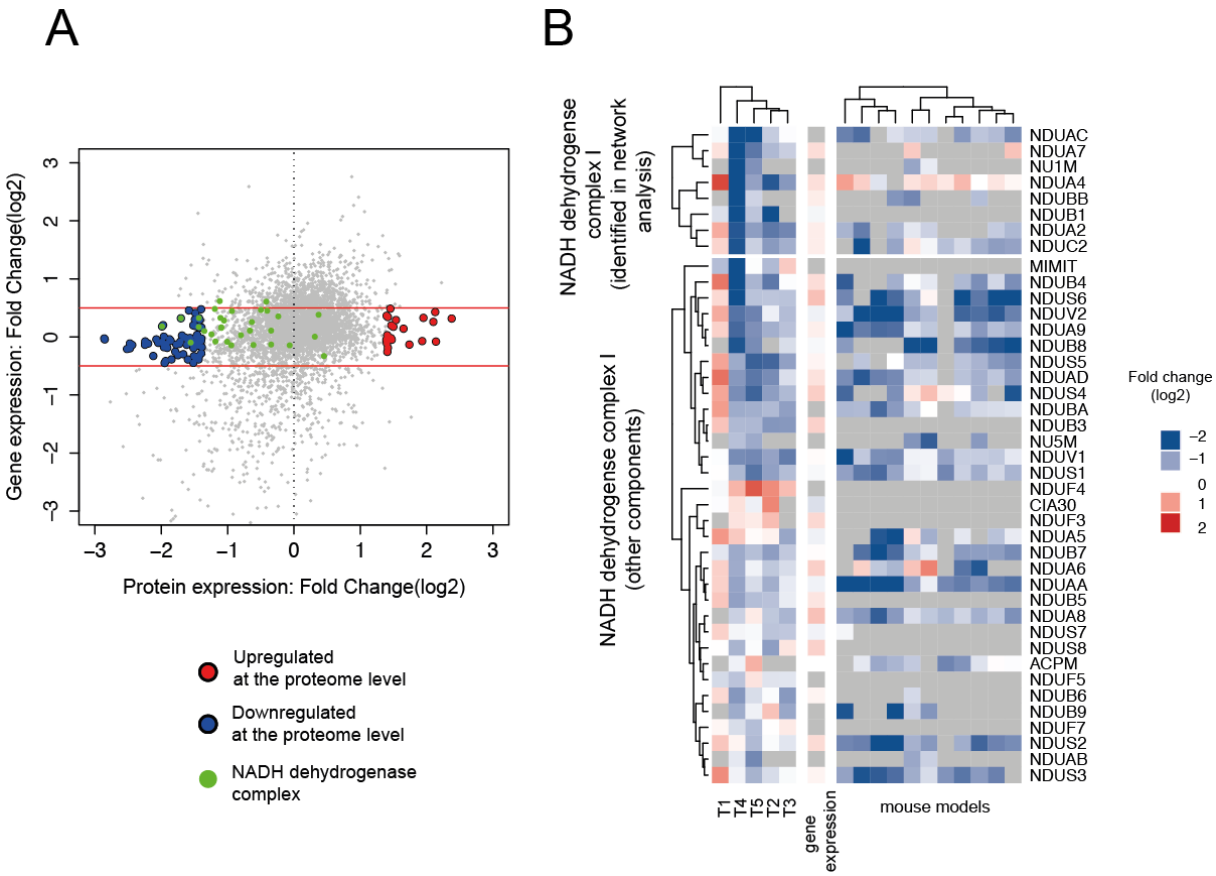



**Figure 20. Proteomic changes not detected by the gene expression analysis**

(A) Comparison of changes of gene expression and protein abundance. The highlighted proteins (blue and red) are either down-regulated (blue) or up-regulated (red) at the proteome level, but their expression is

not affected at the transcriptome level. Green points indicate NADH dehydrogenase complex I components. (B) Heatmap showing the expression of NADH dehydrogenase complex I components. The upper panel includes hits with q value < 0.2. The remaining components of the complex are displayed in the bottom panel. Modified from (Buczak et al., 2018)

### 2.3.10. Analysis of the mitochondrial content in HCC

The proteomic analysis indicated that multiple mitochondrial proteins are expressed at lower levels in tumors as compared to non-neoplastic liver tissue, possibly indicating a reduced number of mitochondria. To check whether other mitochondrial proteins are also affected I analysed on the proteomic changes within each subcellular compartment separately. For this analysis, I focused on four compartments namely nucleus, cytoplasm, mitochondrion and extracellular components. For each quantified protein, I assigned a subcellular location using biomart R package as annotated in uniport (Durinck et al., 2009). To visualize the differences, the distributions of fold changes within different compartments were plotted for each of analyzed specimen (Figure 21). Proteins with multiple annotated localizations were included in each identified compartment. In case of murine HCCs, a clear separation between mitochondrion and remaining compartments was observed indicating that overall amount of mitochondrial protein content is significantly reduced and therefore supporting the hypothesis that tumor cells contain lower number of mitochondria. In case of human HCC the difference is much less pronounced. Nevertheless, in 4 out of 5 cases still significant (as estimated by the Kolmogorov-Smirnov test).

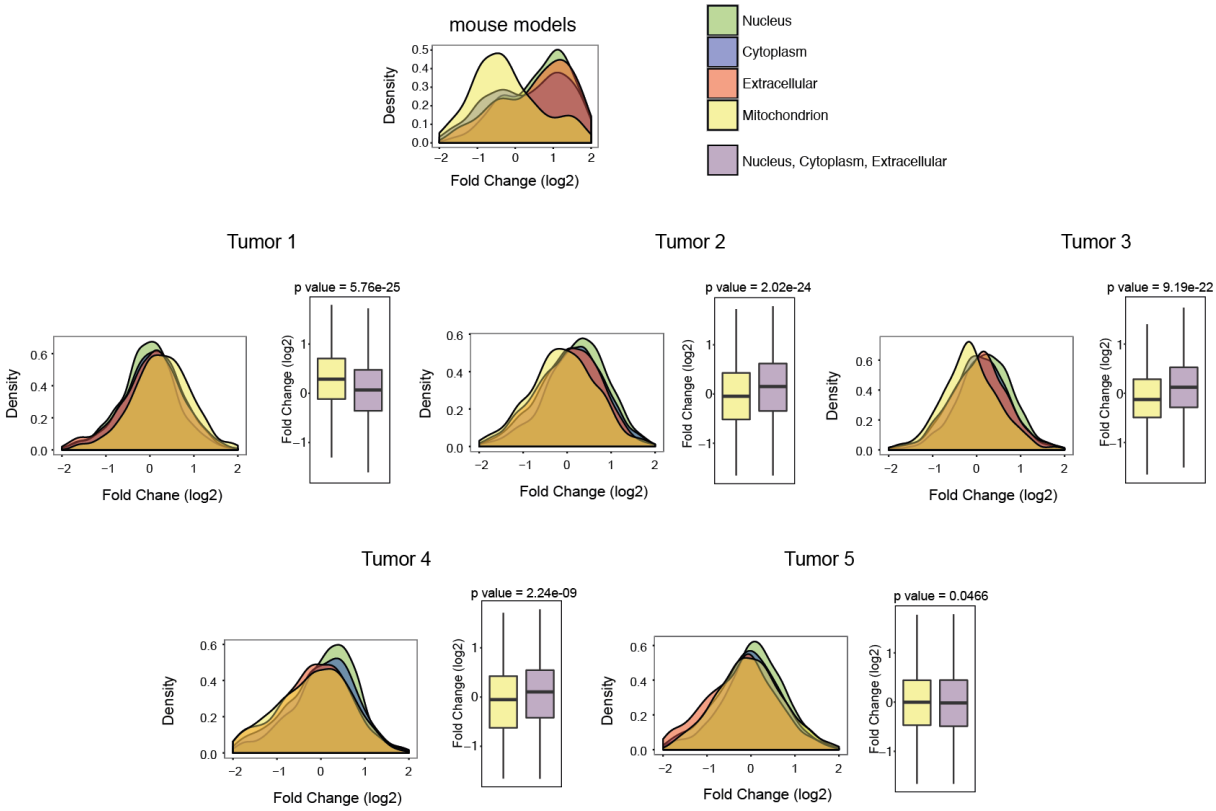



**Figure 21. Distribution of fold-changes for different subcellular compartments**

Density plots show the compartment specific distribution of fold-changes between neoplastic and non-neoplastic tissue. Barplots show the difference in fold-changes of mitochondrial proteins and proteins localized to all other compartments. A representative murine model was selected for comparison. Modified from (Buczak et al., 2018)

In order to validate this finding, I quantified the mitochondrial content in analyzed human HCC specimens with an independent approach. Such quantification is commonly achieved by measuring the activity of either citrate synthase (Mogensen et al., 2006) or the electron transport chain (Picard et al., 2011). Due to fixation such metabolic assays are however not possible in FFPE tissues. I therefore decided to quantify the mitochondrial DNA (mtDNA) content. Although, there are other factors that can affect the abundance of mtDNA, it has been shown that it can be used as a proxy of mitochondrial content (Larsen et al., 2012). The pool of DNA including both mitochondrial and nuclear DNA (mtDNA and nDNA respectively) was isolated from consecutive slides of analyzed FFPE tissues. It was then used as a template for quantitative PCR (qPCR). For mtDNA two genes were selected as targets for qPCR: *MT-TL1* encoding the 12S rRNA and *MT-RNR1* encoding tRNA-leu. The signal was normalized to the amount of nDNA (target gene was *B2M*), which is expected to be constant in all analyzed tissues. The mtDNA/nDNA ratios were compared between tumor and peritumoral tissues. As a control, I performed similar analysis for two renal oncocytoma (RO) specimens – a type of tumor for

which the increased number of mitochondria has been previously reported (Joshi et al., 2015). As expected, an enhanced level of mtDNA was observed in RO. Similarly, the amounts quantified for HCC specimens were in agreement with the proteomic data. For tumor 4, I was not able to perform the qPCR due to the lack of the peritumoral tissue in the remaining FFPE block. Nevertheless for other specimens the constant decrease of around 2-fold was observed. Tumor 1 showed increased level of mtDNA, which is line with the increase of the mitochondrial protein expression observed for that sample (Figure 22).

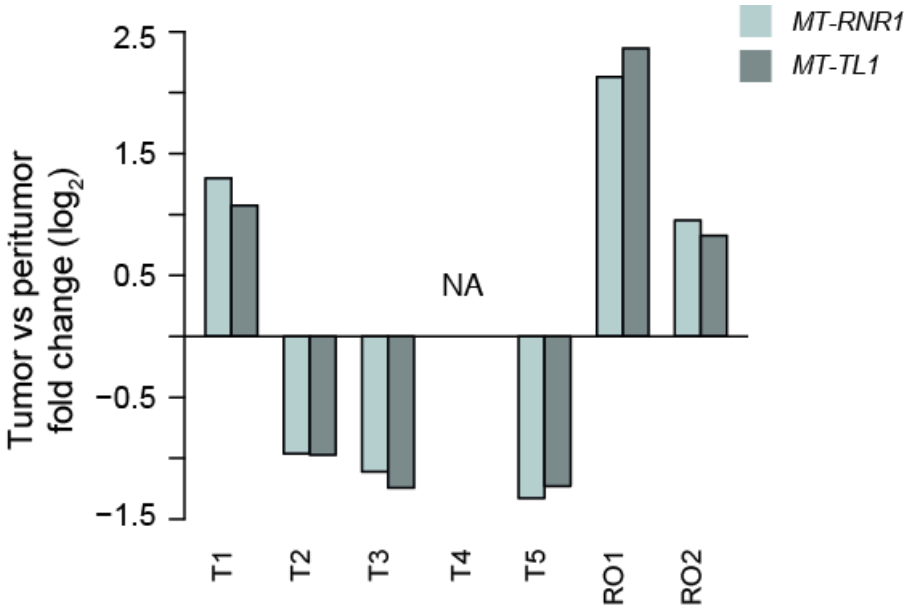



**Figure 22. Quantification of mitochondrial DNA**

The barplot shows the fold change ($log_2$) of mtDNA abundance of tumor in comparison to peritumoral tissues. Two mitochondrial genes (*MT-RNR*1 and *MT-TL1*) were analyzed using qPCR and normalized to the nuclear gene *B2M*. Modified from (Buczak et al., 2018)

***PART III: Proteomic ITH in HCC***

### 2.3.11. Proteomic comparison of tumor center to its periphery

To investigate the intratumoral heterogeneity I used consecutive slices, adjacent to the previously analyzed 5 patient HCC specimens. Using LCM, I separated tumor center and its periphery (10 samples in total) and assembled them into a TMT experiment. In total 5659 proteins were quantified across all analyzed samples with at least two unique peptides. Statistical significance was calculated separately for each sample as before (supplementary figure 4). As expected, differences between patients were more pronounced than these observed within a

single tumor (indicated by the Pearson correlation, Figure 23A). Nevertheless proteomic ITH was still observed, however up to different extent between the analyzed tumors (Figure 23B). Ranked GO enrichment analysis showed that within each tumor, proteins of distinct functions were affected by ITH.

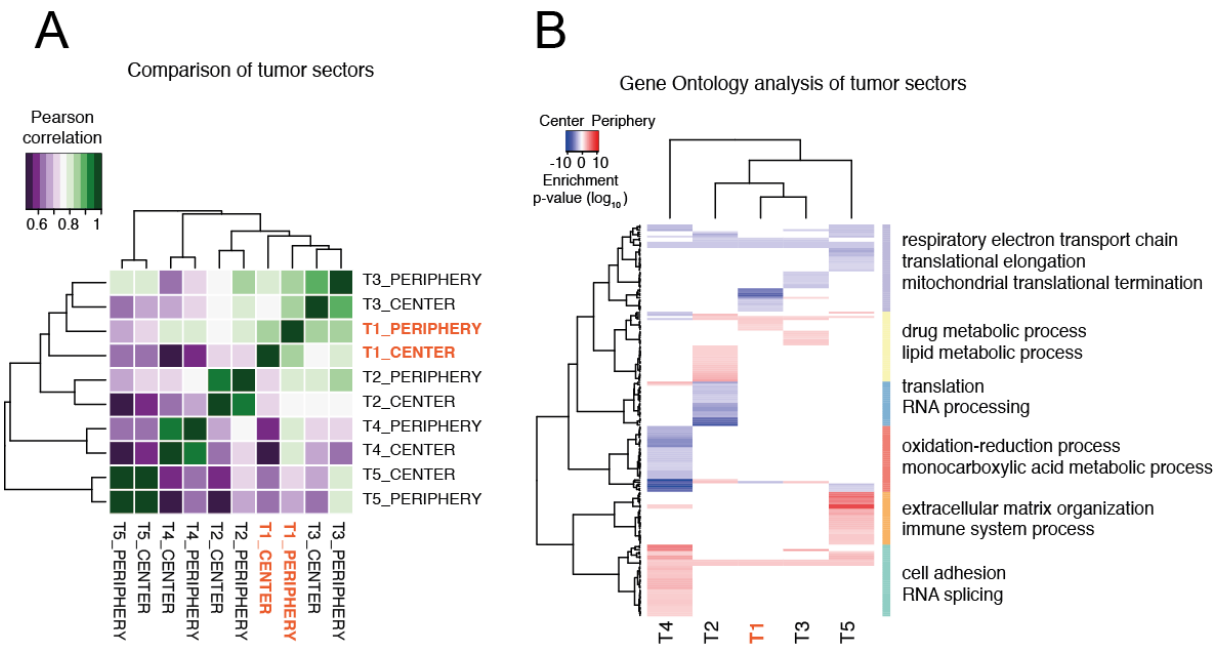



**Figure 23. Overall heterogeneity comparison**

(A) Pearson correlation between samples used for tumor periphery vs. tumor center comparison. (B) Ranked gene ontology enrichment of differentially expressed proteins. The blue color corresponds to proteins that were down-regulated in the tumor. The red color indicates up-regulated proteins. Representative terms from each cluster are displayed next to the heatmap. Modified from (Buczak et al., 2018)

In order to assess, whether proteins affected by ITH are functionally related, I performed a network analysis of significantly regulated proteins. For each tumor, significant hits (q value < 0.2) were mapped to the STRING network using the high confidence score (> 0.7). The resulting networks were analyzed in Cytoscape using MCODE (Bader and Hogue, 2003) and ClueGO plugins (Bindea et al., 2009) as described before. Selected modules of interaction networks are presented in the Figure 24. Among the proteins that were differentially expressed within tumor 2 and 3, I found an enrichment of cytochrome P450 enzymes. Some of them were previously identified as a part of the expression signature of HCC. This finding indicates that markers that are up-regulated in HCC, are not necessarily expressed evenly within a single specimen. The remaining samples displayed different characteristic of ITH. In tumor 4 for example, I observed the increase of proteins involved in the mRNA processing at the center (including multiple

components of the spliceosome). Interestingly, the overexpression of splicing related genes has been previously shown to be common for aggressive types of HCCs (Liu et al., 2014). In tumor 5, I have observed an increase of extracellular matrix proteins, which might indicate that this area was infiltrated with the components of tumor stroma (connective tissue etc.)

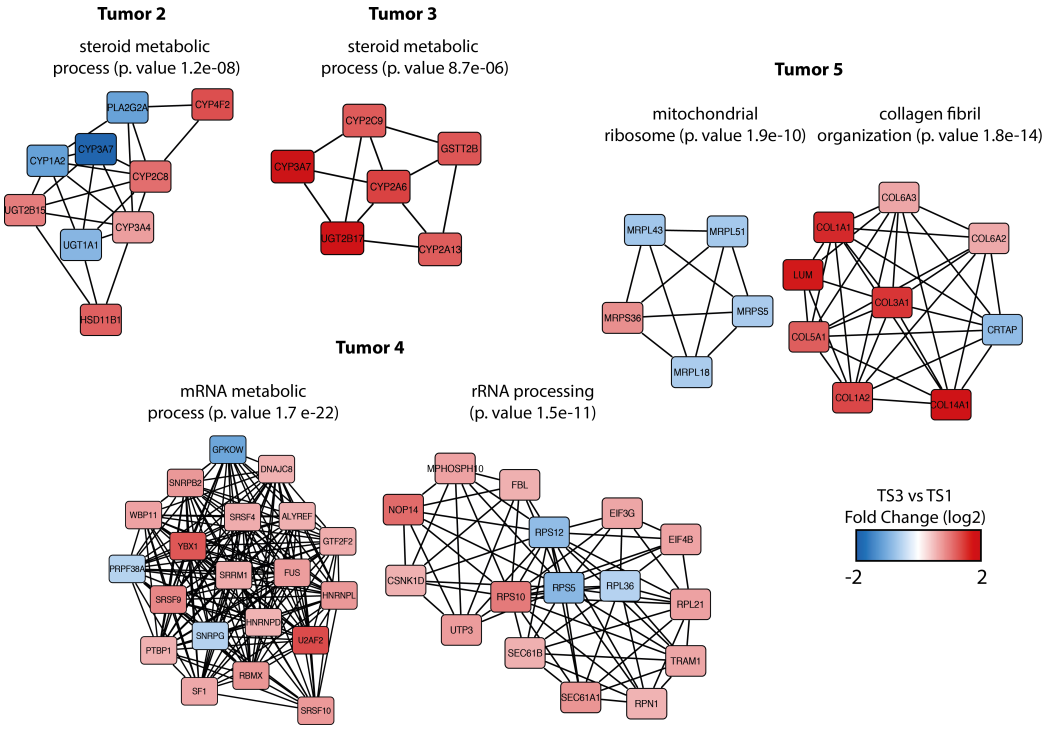



**Figure 24. Network analysis of ITH observed across different specimens**

Proteins with heterogeneous intra-tumor expression (from each sample separately) were extracted and subjected to network analysis. The presented network modules were extracted with MCODE. The GO term enrichment within each module was assigned with ClueGO. The color code of the network nodes indicate the observed fold-change

### 2.3.12. Identification of proteins commonly affected by ITH

Although the observed proteomic patterns seem to be tumor-specific, I attempted to identify proteins that appear to be frequently affected by ITH. For this purpose, I extracted proteins from the entire dataset that were significantly changed in at least two out of five specimens included in the analysis. In total 43 of such cases were identified (Figure 25A). The most striking example was the expression of SerpinB3 (SPB3) and SerpinB4 (SBP4) also known as squamous cell carcinoma antigens (SCCAs) (Figure 25B). Importantly, SCCAs have been previously linked to the liver cancer and even proposed as histological markers of HCC (Zhao et al., 2013). Only in one of the analyzed tumors (T2) a homogenous expression was observed. In the other specimen, I observed strong expression changes (up to 16-fold) between the tumor sectors. SPB3 and SPB4

were quantified with 5 and 3 unique peptides respectively. None of these peptides were detected in the center of the tumor 1 indicating that the protein amount in this sector is below the level of detection. All of these peptides were quantified in the periphery of the same tumor. I thus concluded that also within the tumor 1 a heterogeneous expression of serpins is apparent. In some cases, SCCAs were found to be upregulated at the periphery, while in others I observed an increased expression in the center. This data suggest that the ITH of SSCAs might be independent from the tumor geometry. The spatial cue driving the ITH remains to be characterized.

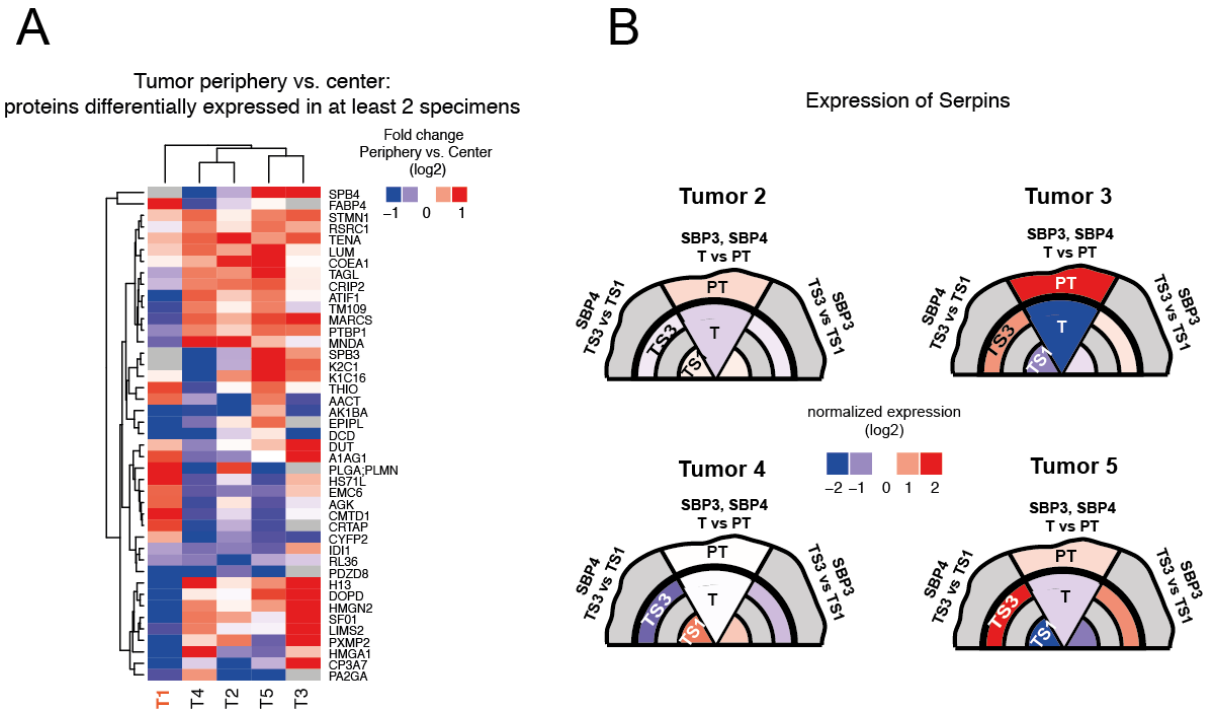



**Figure 25. Proteins frequently affected by ITH in HCC**

(A) Heatmap showing the protein abundance changes between tumor center and its periphery. It includes only proteins that were differentially expressed in at least 2 of the analyzed specimens. For displaying purposes, the fold change values were limited between -1 and 1. (B) Graphical representation of normalized SCCAs expression across tumor sectors. For each tumor, three sections are displayed that correspond to the expression of Serpin B3 (sectors TS1 and TS3) on the left side, Serpin B4 (sector TS1 and TS3) on the right side. The center section is for tumor vs. peritumor comparison and contains information about both of the proteins (both proteins are merged because due to the sequence similarity they were identified as a single protein group in the tumor vs. peritumor TMT experiment). Modified from (Buczak et al., 2018)

Apart from SCCAs, other HCC relevant proteins were differentially expressed in more than one case. For example, two proteins that are required for the cell migration in HCC, namely microtubule-interacting protein stathim (STMN1) (Singer et al., 2007) and myristoylated

alanine-rich C-kinase substrate (MARCS) (Song et al., 2015) were enhanced at the periphery (Figure 26A and B). This is consistent with the section 2.3.5 above, were multiple proteins involved in cell migration were upregulated at the tumor periphery. Also STMN1 has been proposed as negative prognostic marker in the HCC (Yuan et al., 2006). At last, I identified proteins that show a consistent difference between tumor and surrounding non-malignant tissue, but at the same time appear to be evenly expressed in distinct tumor areas. For example, hepatocyte nuclear factor 4-alpha (HNF4A) and quinone oxidoreductase PIG3 (QORX) (Figure 26C and D). This analysis strongly underscores the importance of proteomic studies of ITH, especially during diagnostic marker validation. These findings indicate that the outcome of diagnosis may drastically differ if distinct areas of tumor are inspected during routine clinical analysis.

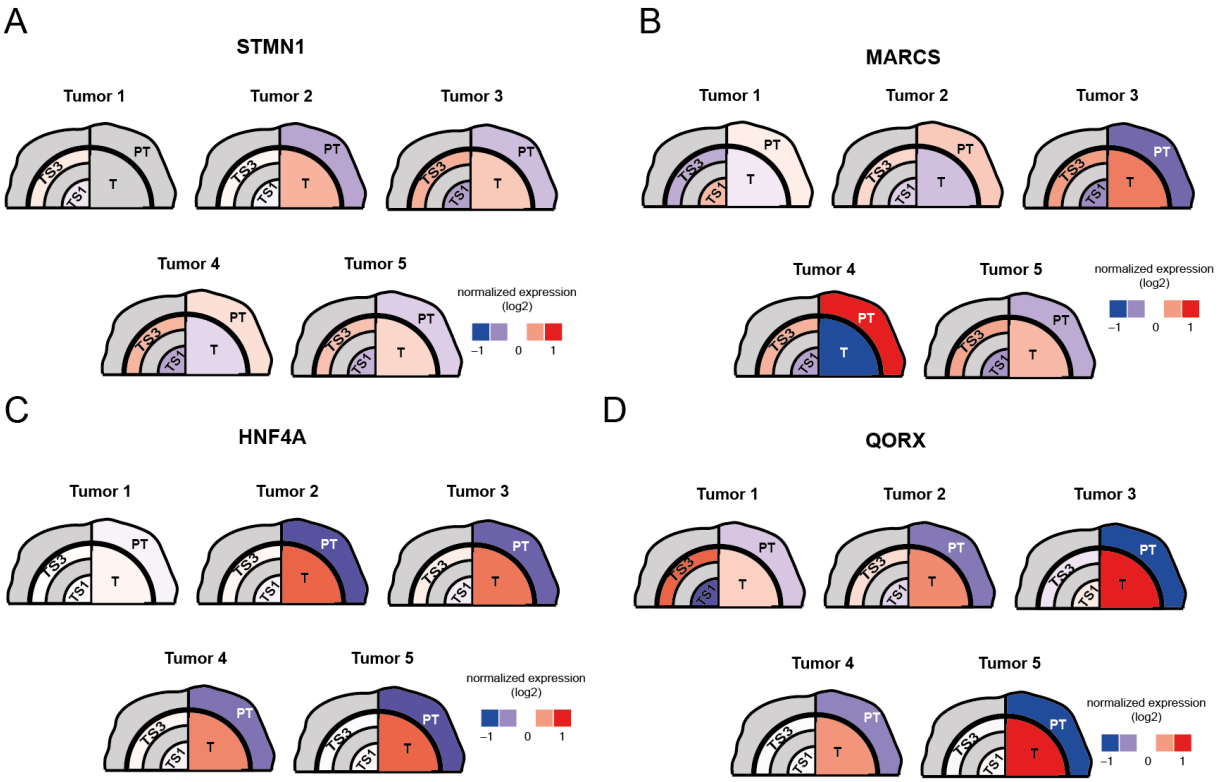



**Figure 26. Expression of clinically relevant markers**

Graphical representation of the normalized expression of (A) Stathmin (STMN1), (B) Myristoylated alaniche-rich C-kinase substrate (MARCS), (C) Hepatocyte nuclear factor 4-alpha (HNF4A) and (D) Quinone oxidoreductase PIG3 (QORX) across different specimen. For each tumor, two sections are displayed that represent the expression within the tumor (TS1 vs. TS3) on the left side, and expression of non-neoplastic vs. neoplastic tissue (T vs. PT) on the right side. Modified from (Buczak et al., 2018)

## 2.4. Discussion

*HCC as a model system to investigate inter- and intra- tumor heterogeneity*

In this work, I focused on the proteomic characterization of hepatocellular carcinoma investigating both overall differences between malignant and non-malignant tissue as well as alterations of protein expression between distinct regions of the same specimen. HCC was selected as a model system for this study for two reasons. First, HCC currently represents the second most deadly cancer. Multiple environmental and genetic factors connected to liver carcinogenesis have been studied (Boyault et al., 2007; Guichard et al., 2012; Schulze et al., 2015), nevertheless the proteomic alteration in HCC are still poorly characterized and require further investigation. Second, HCC is largely homogenous in terms of morphology as compared to other solid tumor types. This makes it a good model for studying proteomic ITH, since it is less likely that the variations in protein expression are simply direct consequences of differences in cellular composition and genetic diversity.

*Development of efficient tool to comprehensively investigate proteomes of limited amount of FFPE material*

In recent years a significant increase of in-depth quantitative analysis has been applied in the field of cancer biology. This is mainly due to the rapid development of the genome sequencing technologies. The genetic analysis provides valuable information about the tumor development and its heterogeneity, nevertheless on its own it is not sufficient to provide a complete picture of the tumor state. Therefore, it should be also supported by the analysis of more dynamic components such as protein expression. Although MS based approaches to study proteomes are well established, for a long time their application to the clinical material such as FFPE tissues was rather limited. Successful attempts to analyze proteomes from fixed specimens have been reported, however the majority of them was based on the analysis of relatively large amount of material that is not always available, depending on specimen size. It has been shown that deep proteomic coverage can be achieved from FFPE tissues (Ostasiewicz et al., 2010; Wiśniewski, 2013) by employing extensive off-line fractionation steps. This approach however significantly increases the total measurement time per sample and as such makes it unsuitable for a clinical setting.

Here, I developed a workflow that enables efficient protein extraction from limited amounts of FFPE tissue. This was achieved by the enhanced protein solublization in the presence of high

concentration of SDS combined with the heat-induced formalin fixation reversal. Protein retrieval was followed by the SP3 based peptide purification allowing for the complete detergent removal with minimal sample loss. The developed workflow is fully compatible with state of art MS-based approaches. In combination with TMT labeling, it allows for the detailed proteomic analysis with coverage close to non-fixed samples. Samples generated with presented approach can also be analyzed by DIA. Due to the lack of off-line fractionation step, protein coverage is slightly reduced, however the required measurement time is significantly shorter. Importantly, for DIA only 10% of obtained peptide material was used, meaning that the amount of starting material can be even further reduced and the number of investigated tissue areas can be increased. In addition, the DIA is not affected by the missing values, even for experiments with large sample sizes, therefore it is suitable tool for clinical studies were hundreds of samples need to be analyzed.

*Significance of spatial proteomics*

The majority of cancer proteome studies do not take the advantage of the excellent preservation of the tissue architecture provided by the formalin fixation. This preservation is valuable not only for the proteomic ITH but also for the general comparison of bulk tumor with the corresponding non-malignant tissues. For example, in liver malignancies fibrous septa (also forming the tumor capsule) are frequently observed within the tumor nodule. While the entire tumor nodule can be relatively easily separated from the surrounding peritumoral tissue by manual macrodissection, the fibrous septa cannot be removed in this way. The presence of fibrotic tissue was also observed in the first HCC specimen presented in this study, but due to the precision of LCM, I was able to remove such regions. As indicated in the Figure 13A, this area has a distinct proteome profile when compared to the analyzed tumor sectors and thus should not be considered as part of the tumor. The same is true for other non-related cell types that can be frequently found within the tumor. This includes infiltrating components of the immune system, blood vessels or necrotic areas. Nevertheless, the major advantage of combining the LCM of FFPE tissue with the comprehensive MS analysis is the possibility to investigate the ITH at the proteome level. Here I demonstrated that within specimens which appear largely homogenous morphologically, another level of yet uncharacterized heterogeneity is observed at the proteome level. The fact that the proteomic ITH affects also clinically relevant marker proteins underscores the importance of characterizing proteomic ITH.

*Robust expression pattern of HCC*

The proteomic analysis of several hepatocellular carcinoma specimens indicates that HCC contain a high level of inter-tumor heterogeneity. While the variation between proteome profiles was observed for the analyzed tumors, the corresponding peritumoral tissues appeared to be relatively similar. As such, the degree of inter-tumor heterogeneity is only partially explained by the fact that the analyzed tumors were of different etiology, including patients with HBV, HCV and steatohepatitis, since the impact of the disease on the overall liver proteome appears minimal. Inter-tumor heterogeneity is thus more likely to be a consequence of exposure to different microenvironmental conditions. Nevertheless, certain features, although manifested up to a different extent, can be identified across several of the analyzed samples. Such features overlap to a large extent with murine models of HCC and transcriptomic data obtained from an independent human patient cohort. By extracting reoccurring features, I was able to derive an expression signature of HCC that contains factors previously linked to HCC as well as some that so far has not been investigated in the context of liver tumorigenesis. Interestingly, the derived signature correlates with the aggressiveness of the murine models, independently of their genetic background. In addition, the expression signature predicts a patient survival in a cohort of human HCCs, thus demonstrating its potential prognostic value.

*Proteomic analysis reveals changes that are not reflected at the transcriptome level*

The aforementioned expression signature of HCC was derived in a way that in contains solely proteins for which changes in abundances were observed at both proteins and transcript levels. As such, it could be potentially applied to predict a patient survival from either transcriptomic or proteomic data. However, there are multiple other regulatory mechanisms that affect the protein abundances regardless of their transcript levels. These can for example be: (i) modulations of translation rates, (ii) regulation of protein stability via post-translational modifications, (iii) activation of protein degradation pathways, and many others. As a consequence, the overall correlation between mRNAs and corresponding protein abundances is usually low (Vogel and Marcotte, 2012) and certain alterations in proteins expression may not be directly reflected at the transcript level. Such cases were also identified in analyzed HCC specimens. Among them I observed an enrichment for members of NADH dehydrogenase complex I, which is a component of mitochondrial electron respiratory chain. Although no changes of mRNA levels of the complex subunits were observed for HCC patient cohort (Roessler et al., 2010), at the protein level consistent down regulation of many of them was detected across nearly all of the analyzed specimen (both human HCC and murine models). In addition, the proteomic data indicated the

general decrease of mitochondrial proteins suggesting that the total mitochondrial volume is lower in tumor cells than in surrounding normal hepatocytes. Aforementioned example indicates that proteomic analysis can point also to the morphological changes on the cellular level (changes in the number or size of particular organelle).

*Mitochondria in hepatocellular carcinoma*

In normal cells, energy is primarily produced via the mitochondrial oxidative phosphorylation. Only in the absence of oxygen, pyruvate – a metabolic product of glycolysis is retained in the cytoplasm and it is converted to the lactic acid. In contrast, in cancer cells, even under the aerobic conditions, glycolysis followed by the lactic acid fermentation is a predominant way of energy production. This phenomenon is widely known as Warburg effect (Warburg, 1956) and has been observed for a vast majority of tumors, including HCC (Sawayama et al., 2014). Therefore, the decrease of NADH dehydrogenase complex I can possibly be a consequence of such metabolic switch in HCC. It can also reflect the presence of mitochondrial dysfunctions in HCC due to the accumulation of mutations in mtDNA. The analysis of mtDNA from 88 HCC individuals revealed that mutation in mitochondrial genes encoding six subunits of NADH complex I are frequently observed in HCC (Li et al., 2017). In addition, the authors of this study observed general decrease of mtDNA in HCC when compared to the healthy liver, which is in agreement with the reduced mtDNA content in HCC specimens analyzed in this work. Finally, the overall decrease of mitochondrial proteins (including NADH dehydrogenase complex I) and mtDNA can be also explained by an increase in the mitophagy activity, which selectively removes dysfunctional mitochondria from HCC. In fact, it has been already reported that enhanced mitophagy is required to promote the hepatocarcinogenesis by controlling the activity of the tumor suppressor protein p53 (Liu et al., 2017).

*Intra-tumor heterogeneity*

A major motivation for this study was to show that the local proteome within tumors intrinsically differs, thus contributing to the overall tumor ITH. By proteomic analysis of different sectors within the same tumor I could show that indeed heterogeneous protein expression is frequently observed in HCC, even in seemingly homogenous specimens. Although the proteomic ITH was observed in all analyzed specimens, for each tumor a distinct set of differentially expressed proteins was identified. Taking into account that the tumor vs. peritumor comparison indicated the large degree of inter-tumor heterogeneity, such outcome was expected. Nevertheless, I identified several proteins that were significantly differentially expressed in multiple HCC

specimens. However, the spatial patterns of these changes were not always the same, suggesting that the proteomic ITH is independent from the tumor geometry. Interestingly, the identified proteomic ITH of HCC is up to some extent functionally related with the proteomic features that distinguish HCC from the normal liver tissue. It therefore highlights the importance of proteomic ITH for biomarker discovery studies and diagnostic applications. For example, ribosomal proteins that were generally upregulated in HCC were also frequently affected by the ITH. Another example is the expression of the fatty acid binding protein-4 (FABP4), which so far has not been investigated in regards of HCC. The expression of FABP4 was generally higher in analyzed tumor when compared to peritumoral tissue, and at the same time it displayed uneven distribution in some of specimens. This could possibly mean that the analyzed tumors comprise of distinct cell populations, out of which some have more aggressive phenotype and some might partially resemble the non-neoplastic tissue of origin.

# 3. Functional characterization of the nucleoporin gp210

## 3.1. Introduction

### 3.1.1. The nucleocytoplasmic transport system

In eukaryotes, the nuclear envelope (NE) constitutes a physical barrier that separates the nucleoplasm and cytoplasm. The eukaryotic cell requires transport of macromolecules between these compartments to function properly. This transport occurs through the cells' largest discrete protein complex – the nuclear pore complex (NPC). The NPC forms a channel across the inner and outer membrane of the NE and is the sole gateway between the two cellular compartments. Molecules smaller than 40 kDa can freely diffuse through the NPC. However, the shuttling of large cargo requires interaction with nuclear transport receptors (NTR) that mediate active transport through the NPC (Panté and Kann, 2002). While the primary role for components of this nucleocytoplasmic transport system (NTS) is the facilitation of macromolecular transport through the NPC, they have been implicated in other cellular processes. These include processes such as chromatin organization and regulation of the gene expression (Akhtar and Gasser, 2007; Brown and Silver, 2007; D'Angelo et al., 2012).

### 3.1.2. NPC composition and architecture

The vertebrate NPC is comprised of approximately 1000 copies of roughly 30 different proteins collectively termed nucleoporins (nups). Nups are organized into small, repeating subcomplex units, each comprised 3-10 individual proteins (Ori et al., 2013). The total molecular weight of the human NPC is approximately 112 MDa (Schwartz, 2016). It is a highly organized macromolecular machine, which exhibits an eight-fold rotational symmetry about the axis of transport. The NPC occupies sites in the NE where the inner and outer nuclear membranes fuse and forms a channel of ~60nm in diameter. The structural scaffold of the NPC is built from three major elements, (i) the inner-ring which is sandwiched between two outer-rings, (ii) the cytoplasmic and (iii) nucleoplasmic rings. The principle component of both outer rings (indicated in dark orange and purple in Figure 27) is the Nup107-subcomplex and these rings

thus display a highly similar architecture(Bui et al., 2013). This ten-protein subcomplex has a characteristic Y shape and is frequently called the Y-complex. In humans, 32 copies of the Nup107 complex manifest the cytoplasmic and nucleoplasmic rings. Each ring is composed of eight pairs of (inner- and outer-) Y-complexes. They are organized into two reticulated rings with 16 copies each and exhibit eight-fold rotational symmetry about the axis of transport (Bui et al., 2013). The inner-ring is comprised of the five-member Nup93 and three-member Nup62 subcomplexes (shown in blue in Figure 27). The architectural principles of the inner-ring and two outer-rings are similar. The four core modules of the inner-ring (constituted from members of the Nup93 and Nup62 subcomplexes) are repeated 8 times around the rotational axis. Two of which, are stacked together in a C2-symmetric fashion and form the inner-ring, yielding an overall assembly with a total of 32 copies of each subcomplex (Kosinski et al., 2016). In addition to this structural scaffold, peripheral nucleoporins are attached to both cytoplasmic and nuclear sides of the NPC. These are known respectively as cytoplasmic filaments and the nuclear basket. Furthermore, the NPC central channel is filled with nucleoporins of intrinsically disordered regions that are enriched in phenylalanine and glycine repeats and called FG-nups. Although, the exact organization of FG-nups within the central channel is not known, a disordered meshwork with liquid-liquid like phase separation characteristics is proposed to constitute the permeability barrier of the NPC (Lemke, 2016). FG-nups interact in a transient manner with nuclear transport receptors (NTRs), and mediate the transport of cargo bound to the receptor across the NPC. Finally, the NPC contains three transmembrane nucleoporins: NDC1, Pom121 and Gp210.

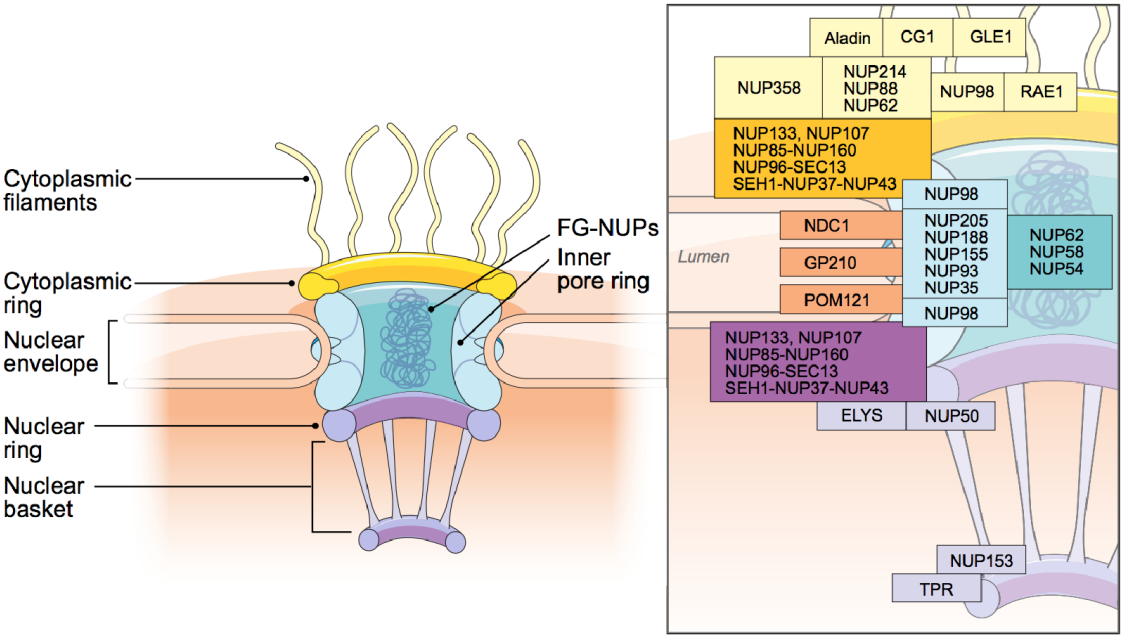



**Figure 27. Schematic representation of the Nuclear Pore Complex (NPC)**

The left panel illustrates the NPC located within the nuclear envelope. Cytoplasmic face (in yellow) consists of the filaments and the cytoplasmic ring. FG-nucleoporins bound to the inner ring (in blue) localize in the middle part of the pore. The nucleoplasmic face (in purple) consists of the nuclear ring connected to the nuclear basket. The right panel indicates the nucleoporins involved in the formation of different subcomplexes. Modified from (Beck et al., 2017)

### 3.1.3. Dynamic composition of the nuclear pore complex

The NPC is a highly dynamic macromolecular assembly with respect to its composition. Measurements of the residence time for different nucleoporins at the NPC indicate that while some nups constitutively associate with the NPC through the cells entire lifespan, others are constantly replaced with new molecules. The residence time of nucleoporins at the NPC ranges from seconds to hours. For example, Nup50 that interacts with cargo at the nuclear basket has a residence time of approximately 20 s, while the inner ring component, Nup93 maintains its association with the pore for around 70 h (Rabut et al., 2004). In general, peripheral Nups tend to associate with the NPC for a shorter time. In contrast, the scaffolding nucleoporins are more stably associated with the NPC (Rabut et al., 2004). In at least some cell types, they exhibit extremely slow turnover rates (Toyama et al., 2013). Targeted proteomic analysis of NPC

components across different cancer cell lines indicates that NPC composition is cell type specific. While structural components are expressed at comparable levels across cell types, significant variation in was observed in the expression of peripheral Nups (Ori et al., 2013) (Figure 28). Together these data suggest that dynamic rearrangements of NPC stoichiometry can be adjusted to the context specific requirements. However, regulatory mechanisms and functional consequences of changes in the NPC composition remain largely elusive.

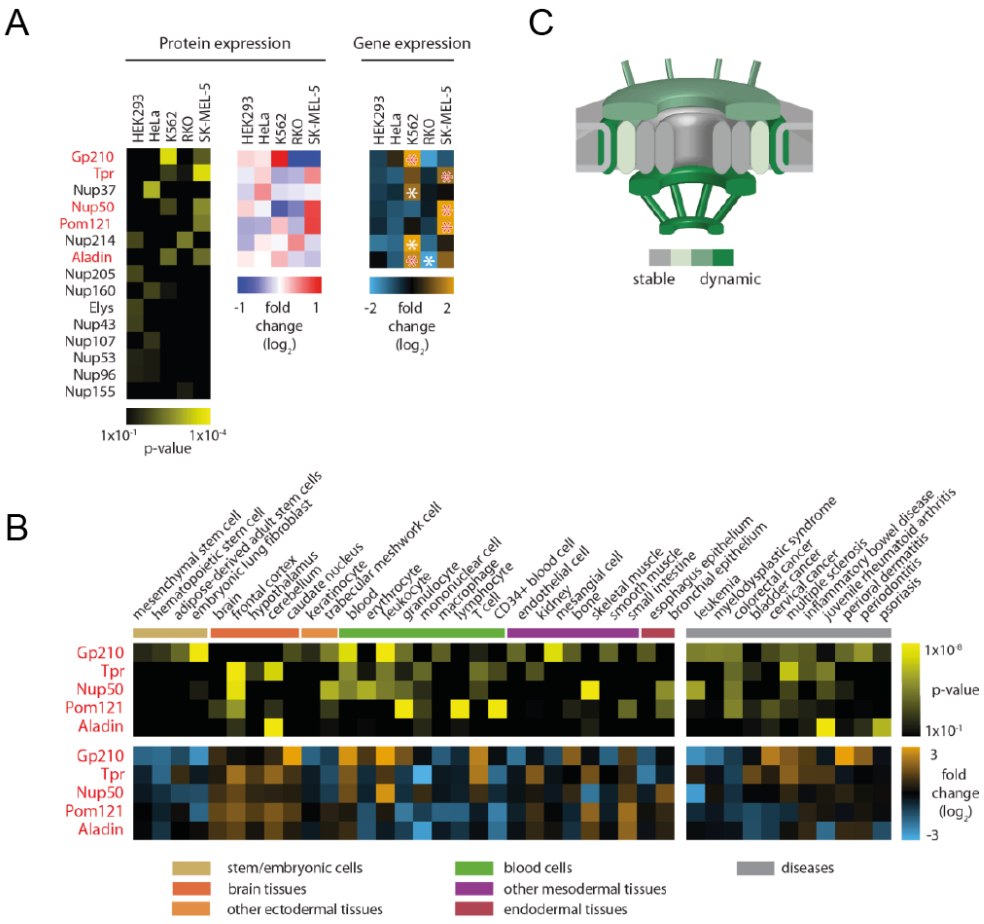



**Figure 28. Cell type specific expression of NPC components**

(A) Targeted proteomics measurement of nucleoporin abundance across different cell lines. Proteins highlighted in red, showed similar pattern in the mRNA expression analysis (B) mRNA expression pattern of selected nucleoporin across various tissues. (C) Components of the NPC are colored according to expression. Scaffold nups are more stable than transmembrane or peripheral nucleoporins. Modified with permission from (Ori et al., 2013)

### 3.1.4. The nucleocytoplasmic transport system in cancer

Multiple studies suggested that cancer driver pathways impact the NTS. For example, the eukaryotic translation initiation factor 4E (eIF4E), classified as an oncogene, can reprogram the NPC to promote mRNA export (Culjkovic-Kraljacic et al., 2012). In this study, it was shown

that overexpression of eIF4E results in alteration of the NPC's cytoplasmic face by a ~3-fold reduction of Nup358, which leads to increased eIF4E-dependent mRNA export. Overexpression of Nup358 was sufficient to inhibit the oncogenic transformation by eIF4E. Another striking example of nucleoporin involvement in cancer is the overexpression of Nup88. First found to be overexpressed in 75% of ovarian tumors, (Martínez et al., 1999), subsequent analysis revealed that Nup88 is also upregulated in a variety of other cancers (Gould et al., 2000). Although the functional link between Nup88 overexpression and cancer development is still unclear, some mechanisms have been suggested. For example, it has been shown that the loss of Nup88 inhibits nuclear accumulation of NF-κB upon TNF (Tumor Necrosis Factor) stimulation suggesting that Nup88 is required for the transport of NF-κB (Takahashi et al., 2008). Nuclear accumulation of NF-κB is a recurring feature in multiple cancers, therefore it has been proposed that overexpression of Nup88 may cause an increase of NF-κB import to the nucleus leading to the upregulation of its target genes.

In HCC, overexpression of several NTRs namely exportin-1, exportin-2, importin-α1 and importin-α5 have been observed. The functional characterization of exportin-2 and importin-α1 suggest that they act as anti-apoptotic factors and whose expression is p53-dependent and thus indicate that its upregulation supports the survival of tumor cells (Winkler et al., 2014). Other NTRs components, like exportin-4 and Nup98 have been shown to have a tumor suppressive role in liver cancer and consistently found to be downregulated in HCC. The targeted proteomic analysis of liver cancer related cell lines (Figure 29A) indicate that NPC composition adapts upon malignant transformation, even among cells of the same origin (Ori and Singer, unpublished data). One such nucleoporin, with altered expression among liver cancer cell lines is gp210. For instance, in Hep3B, HuH7 and HuH6 cells, gp210 expression was significantly greater than in other common liver cell lines, such as Sk-Hep1 and HLF. More significantly, the proteomic analysis of an HCC cohort of 48 patients indicates that the expression of gp210 can vary extremely between the individuals (Ori and Singer, unpublished data). In some cases, the abundance of gp210 was drastically higher (up to 14-fold) than in the surrounding healthy tissue, while in others, a decrease in the expression was observed (Figure 29B). Analysis of the NPC stoichiometry suggests that within a single pore there are 32 copies of gp210 (Ori et al., 2013). It is therefore unlikely, that in cases of strong gp210 overexpression all molecules can be accommodated within NPCs. This hints at a role for excess gp210 apart from the NPC.

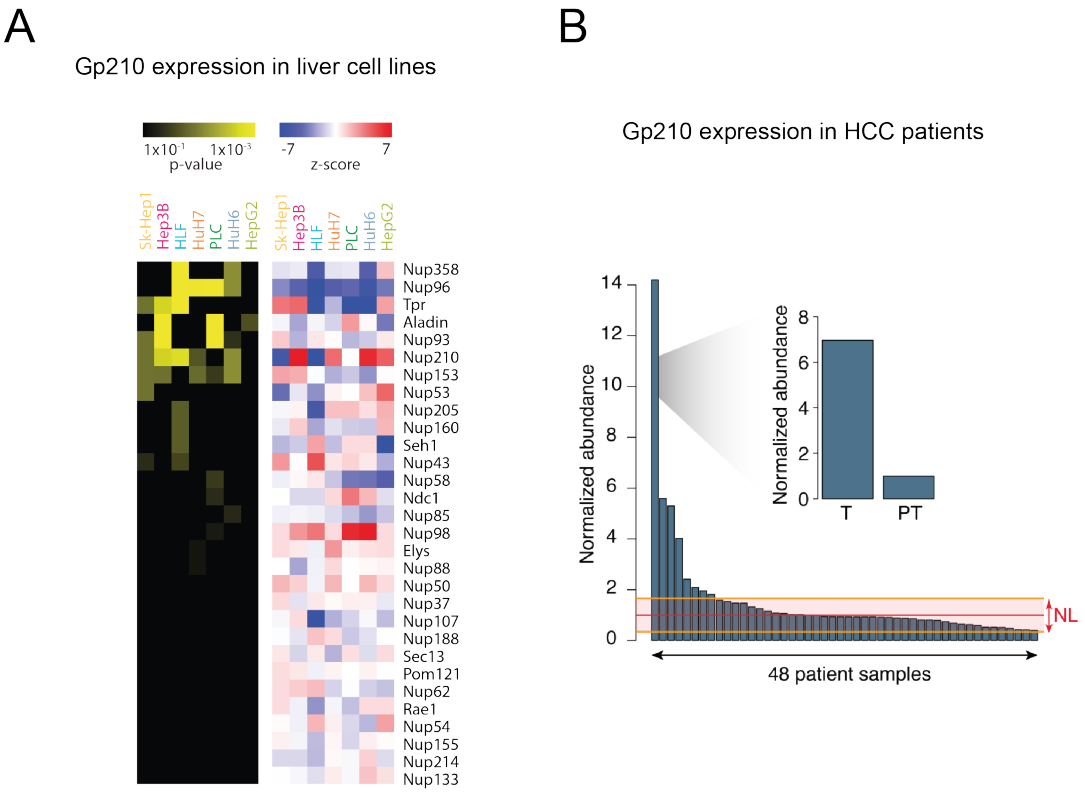



**Figure 29. Gp210 expression in liver cell lines and HCC**

(A) Nucleoporin abundance measured across seven liver cell lines using targeted mass spectrometry. (B) Gp210 abundance for the 48 HCC patients cohort. Barplot indicates the abundance in the tumor normalized to the surrounding peritumoral (PT) tissue.

### 3.1.5. Structural analysis of gp210

Gp210 is a 210 kDa, type I integral membrane glycoprotein (Wozniak et al., 1989). A large part of the protein is localized to the perinuclear space (Greber et al., 1990) and its not required for the targeting to the NPC (Gomez-Cavazos and Hetzer, 2015). Its transmembrane domain is close to the C-terminus, and leaves only a short (58 amino acid) region extending out of the membrane to interact with other NPC components. In yeast, the presumed homolog of gp210 is pore membrane protein 152 (pom152) (Wozniak et al., 1994). It also contains the large luminal domain and short NPC-associated region, however unlike in vertebrates, the C-terminus localizes to the perinuclear space and the N-terminal domain faces the cytosolic side thus interacting with other NPC components. The structural analysis of pom152 revealed that its' luminal region is an approximately 40 nm long, flexible strand of nine Ig-like domains (Figure 30) (Upla et al., 2017). The authors of this work propose a model in which pom152 dimerizes in an anti-parallel fashion and 8 copies of such dimer form a ring around the NPC. Although the structure of gp210 is not known to date, the sequence-based predictions suggest that its

architecture resembles the fold of pom152 (Gomez-Cavazos and Hetzer, 2015; Upla et al., 2017).

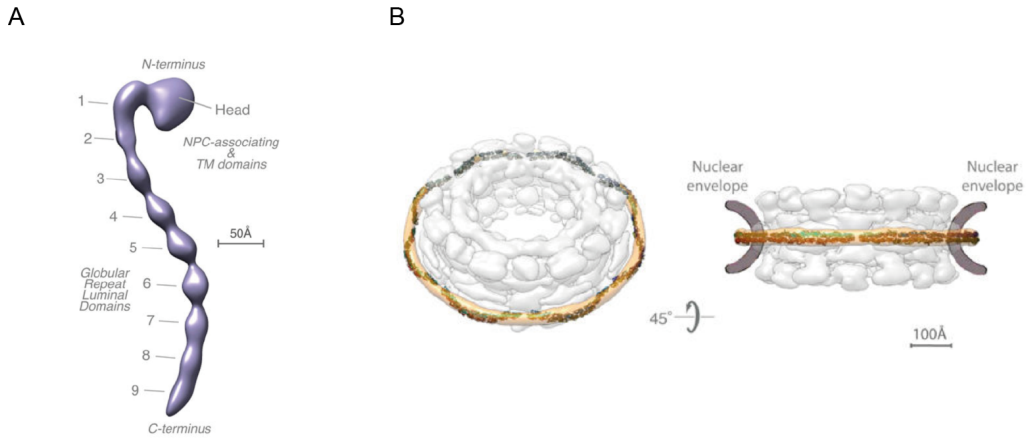



**Figure 30. Structural analysis of pom152. The yeast homologue of gp210**

(A) Negative-stain electron density map of pom152, (B) A model in which 16 copies of Pom152 were fit into the yeast NPC map (Alber et al., 2007) in an anti-parallel fashion. Reproduced from (Upla et al., 2017)

### 3.1.6. Gp210 and NPC assembly

Historically, gp210 was proposed to be involved in NPC formation and in the anchoring of the pore to the nuclear membrane (Wozniak et al., 1989). This hypothesis was supported by a study in which *Xenopus* extract was used to investigate NPC assembly (Drummond and Wilson, 2002). The authors showed that blocking the C-terminal domain (either by recombination or by using an antibody against the C-terminus of the protein) resulted in the arrest of the NPC formation at an early intermediate step. Later, it was shown that depletion of gp210 affects HeLa cell viability and NPC assembly in *C. elegans* (Cohen et al., 2003). However, other studies seem to contradict this hypothesis. Analysis of gp210 content in mice embryos revealed that it is absent in some of the tissues (Olsson et al., 2004). Other groups have shown that depletion of gp210 does not affect NPC assembly (Antonin et al., 2005; Eriksson et al., 2004; Stavru et al., 2006). These findings highlight the possibility that gp210 may be required only in certain cell types and not others. Nevertheless, its exact function at the NPC and beyond remains unclear.

### 3.1.7. Gp210 in differentiation

There is strong evidence indicating important roles for gp210 in non-NPC related contexts. For example, there is evidence to suggest that gp210 is important for myogenic differentiation

(D'Angelo et al., 2012). Using the C2C12 cell line – a well-established murine myogenic model, it was shown that gp210 is absent in myoblasts, however induction of the expression is observed during myotube formation. Depletion of gp210 inhibits the differentiation process by downregulation of genes related to muscle differentiation. In contrast, the overexpression of gp210 results in the acceleration of myotubes formation (D'Angelo et al., 2012). Further studies indicate that localization to the NPC is not required for this process to be rescued as expression of gp210 lacking its' pore-targeting domain was sufficient to restore differentiation in gp210-depleted cells. The reduction of unfolded protein response and blocking the ER-stress related apoptotic pathway with a chemical chaperone - tauroursodeoxycholic acid also restored myotubes formation in gp210-depleted cells suggesting that the pathway for gp210 control of the mygonic differentiation involves regulating ER stress (Gomez-Cavazos and Hetzer, 2015). On the other hand, analysis of muscle differentiation in *Danio Rerio* suggests that gp210 is required at the NPC to enable the assembly of Myocyte-specific enhancer factor 2C (MefC2) complex, which is needed for efficient expression of genes that control muscle structure (Raices et al., 2017).

## 3.2. Aims

Based on the expression analysis and measurement of NPC residence time, gp210 can be classified as one of the most dynamic component of the NPC. The function of this nucleoporin and whether it is required for the proper functioning of the NPC remains unclear. Several lines of evidence indicate that gp201 has non-NPC related functions in development, however the details remain to be characterized. Further, gp210 has been linked to liver disorders. Its expression drastically varies in HCC, but besides this it has been also linked to the primary biliary cirrhosis – an autoimmune liver disease.

In order to further characterize gp210 function and to establish it's links with other cellular processes, I used the following two approaches:

1. Proximity labeling combined with mass spectrometry using BioID system in order to understand cellular interactions of gp210
2. Proteomic analysis of gp210 depleted cell lines to understand its impact onto different functional modules of the cell

## 3.3. Results

### 3.3.1. Analysis of the gp210 localization

To get further insights into potential roles of gp210 away from the NE, I first analyzed the localization of gp210 in different cell lines using the immunofluorescence microscopy. The following cell lines were selected for the analysis: Sk-Hep1 and HLF as liver cell lines with low expression of gp210, Hep3B and HuH6 as liver cell lines with high abundance of gp210 and finally HeLa and Hek293 cell lines with the moderate level of gp210 expression (Figure 31). In case of HeLa, the signal was observed mainly at the nuclear envelope, indicating that majority of the gp210 molecules are localized to NPCs. A similar pattern was observed for Hek293, nevertheless some signal was also observed in other compartments (ER). For cells with low expression level (Sk-Hep1 and HLF) a weak signal was observed around the nucleus. For cells with high expression level of gp210 (Hep3B and HuH6) in addition to the NE, a strong ER signal was also observed. These results indeed show that gp210 also localizes away from the NPC.

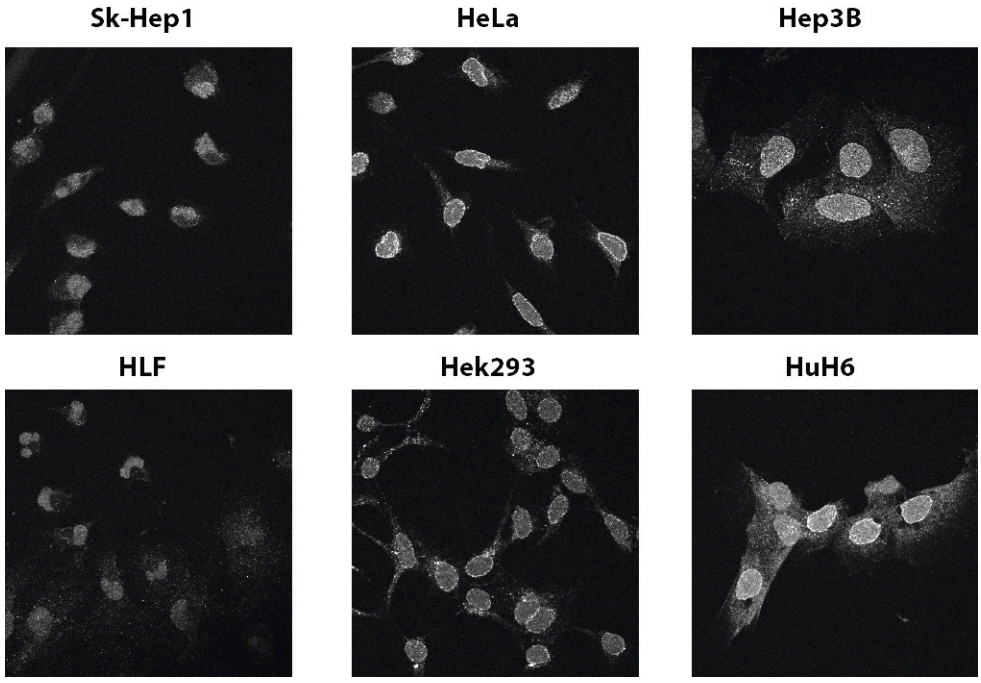



**Figure 31. Immunofluorescence analysis of gp210 in different cell lines**
Cells were stained with anti-gp210 antibody.

### 3.3.2. Generation of cell lines expressing BirA*-tagged gp210

In order to map the gp210 interacting proteins I used the BioID approach (Roux et al., 2012). This system utilizes the biotin ligase (BirA) from *Escherichia coli*, which biotinylates the biotin carboxyl carrier protein (BCCP), a subunit of acetyl-CoA carboxylase (Chapman-Smith and Cronan Jr, 1999) that has been genetically modified to become promiscuous. The biotiniylation of the target protein requires two steps: (i) activation of biotin molecule with ATP to form biotinoyl-5'-AMP and (ii) release of the active biotin in the presence of target sequence. It has been shown that BirA with R118G mutation (BirA*) has two orders of magnitude lower affinity to the biotinoyl-5'-AMP than the wild type enzyme and therefore the active biotin is released from the BirA even without the presence of the target sequence (Kwon et al., 2000). Therefore free biotinoyl-5'-AMP can react with primary amines that are in the close proximity. The mutated BirA is fused to a protein of interest. Once the culture medium is supplemented with biotin, the biotin ligase is provided with its substrate and neighboring proteins are covalently labeled with biotin. Those are subsequently purified using the streptavidin beads and analyzed with mass spectrometry. Such approach has been successfully used in the past to study the interactions of different range of proteins, for example to map specific cargos of nuclear transport receptors (Mackmull et al., 2017).

Using Hek293 Flp-In T-REx cells, I generated stable a cell line with the inducible expression of BirA*-gp210. To enable the proper localization of the fusion protein, BirA* was inserted into the luminal part of gp210, right after the ER localization signal peptide. The radius of biotin labeling with BirA* has been estimated to be around 10 nm (Kim et al., 2014). Therefore, assuming that human gp210 has similar architecture to the yeast homolog pom152 (Upla et al., 2017), it is possible that using N-terminally fused BirA*, the C-terminal interactors may be missed. To obtain a better coverage and information about the domain-specific interactions I also designed constructs in which BirA* was inserted at different positions within the gp210 sequence. Additionally, the sequence between signal peptide and the insertion site was removed. To design the BirA insertion sites, the secondary structure of gp210 was predicted (done by Dr. Jan Kosinski). Based on these predictions, putative IgG-like domains were identified within the gp210 sequences. For BirA insertion, only sites in linkers between domains were selected. In addition, a FLAG-tag was introduced into each construct to enable monitoring of the protein expression and the proper localization. In total 5 different cell lines were generated (summarized in the Table 1)

**Table 1. Generated BirA*-gp210 constructs**

| Construct name | Residues removed | Length of construct |
|---|---|---|
| **Full length** | - | 1886 aa |
| **Asn 287** | 27-287 | 1625 aa |
| **Ser 869** | 27-869 | 1043 aa |
| **Ser 1589** | 27-1589 | 323 aa |
| **Arg 1789** | 27-1789 | 123 aa |

To confirm the expression and proper localization of these constructs, each cell line was analyzed by immunofluorescence. In all the constructs, a similar level of BirA* expression was observed. As expected, the recombinant gp210 was primarily localized to the nuclear envelope; nevertheless a subset of it was retained in the ER. Representative examples are shown on the Figure 32.

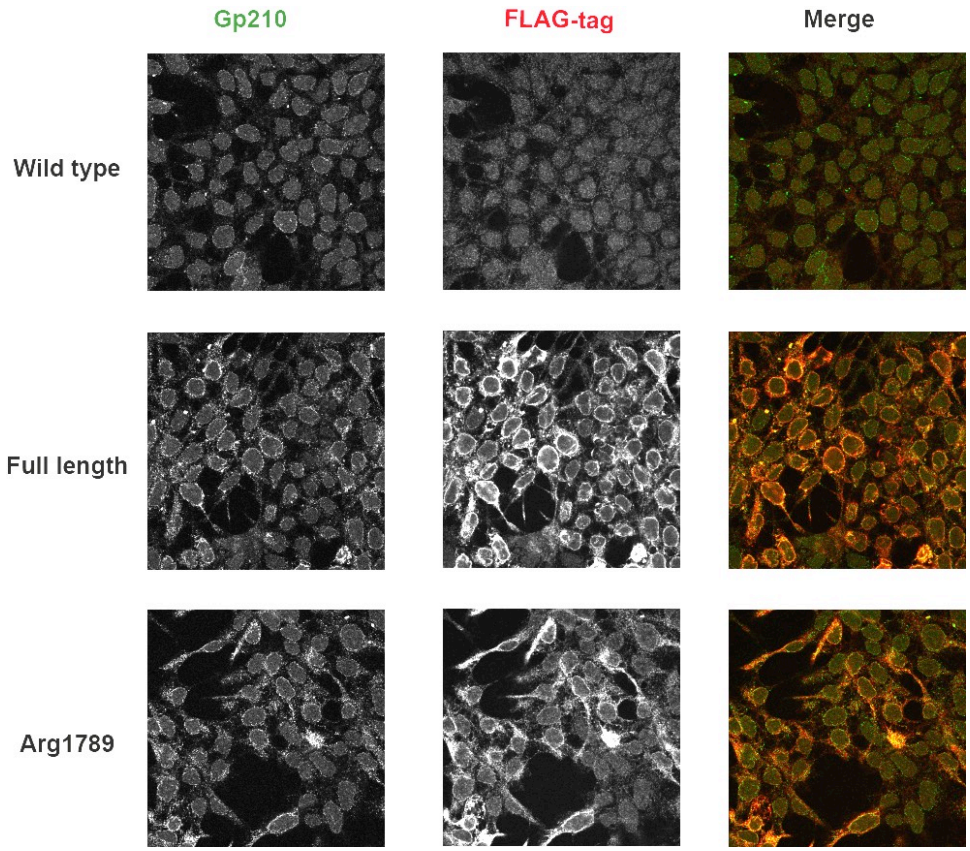



**Figure 32. Immunofluorescence of BirA*-gp210 constructs**

Cells were stained with anti-gp210 antibody (green) and anti-flag (red). For comparison full-length and arg1789 constructs were selected.

I checked the distribution of proteins that were biotinylated by BirA*. As shown in the Figure 33, biotinylated proteins were dispersed throughout the ER and also localized to the nuclear envelope. This observation indicates that either gp210 interactors are mobile within the ER or alternatively gp210 is not constitutively associated with NPCs, as previously suggested (Rabut et al., 2004).

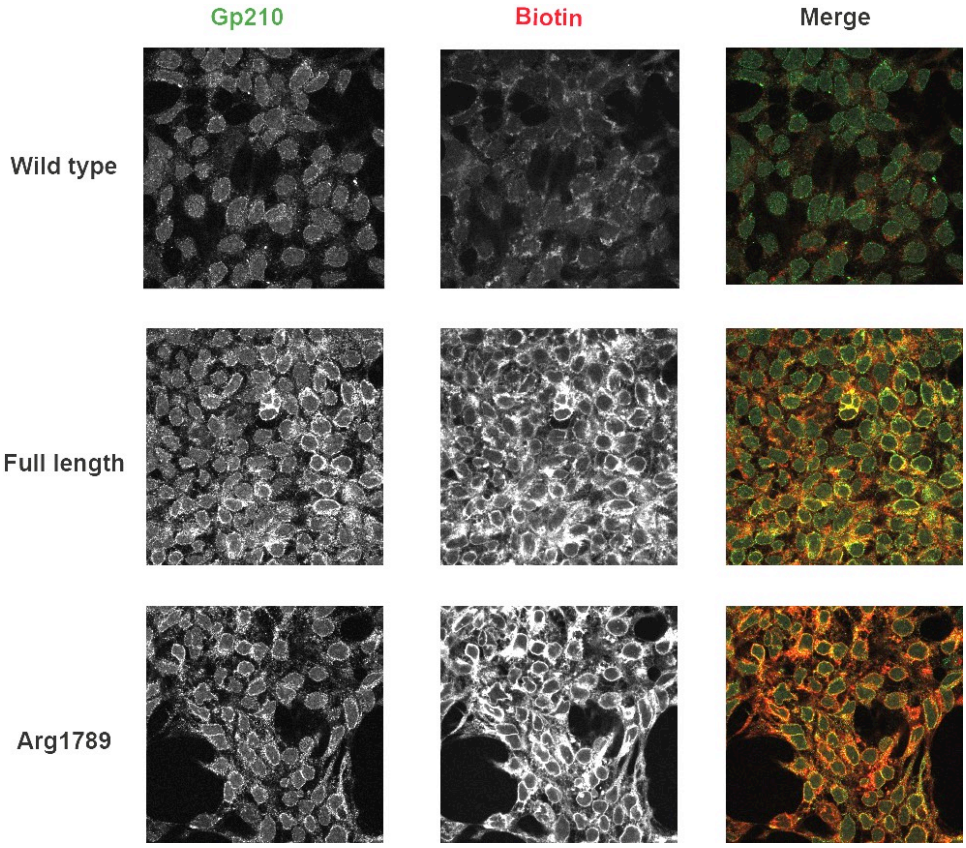



**Figure 33. Distribution of biotinylated proteins**

Cells were stained with anti-gp210 antibody (green) and streptavidin conjugated to alexa647 (red). For comparison only full-length and arg1789 constructs were selected.

### 3.3.3. Identification of gp210 interactome

For the BioID experiments, the expression of the BirA*-gp210 fusion proteins was induced for 48h and subsequently biotin was added to the cell culture medium to proceed with the biotin labeling for another 48h. Cells were collected and biotinylated proteins were purified using streptavidin beads. Due to the extremely strong interaction between streptavidin and biotin (Green, 1975), the protein digestion was performed directly on the beads. To obtain the information about biotinylation sites, peptides remaining on the beads were eluted under very harsh condition (80% acetonitrile, 20% formic acid) as previously suggested (Mackmull et al., 2017). To identify non-specific hits, such as naturally biotinylated proteins or proteins that bind

to the beads in an unspecific manner, non-modified Hek293 cells were used as a control. Three independent biological replicates of each experiment were performed. As interactome of gp210, I considered proteins that were: (i) identified in at least 2 biological replicates of at least one construct and never detected in the control samples or (ii) enriched in the BirA cell line sample in comparison to the control (log2 of fold change > 2, adj. p.value < 0.05). The entire dataset was comprised of 3531 proteins out of which 854 were classified as specific hits according to aforementioned criteria. This rather large number might indicate that many ER proteins are labeled by gp210. These proteins were subjected to the gene ontology analysis. GO term enrichment was calculated for three categories: "Cellular Component", "Biological Process" and "Molecular Function" (Figure 34). As background proteome I used the list of proteins that are usually identified in the shotgun analysis of Hek293 cell lysate (in total 3540). In terms of the cellular components, a strong enrichment of ER resident proteins was observed, which agrees with the subcellular localization of gp210 and the immunofluorescence analysis of biotinylated proteins. With respect to GO terms related to biological process, an enrichment of proteins involved in the protein folding and glycosylation was observed. These proteins are ER-resident as well and likely to be involved in the processing of the gp210 polypeptide chain in the ER lumen.

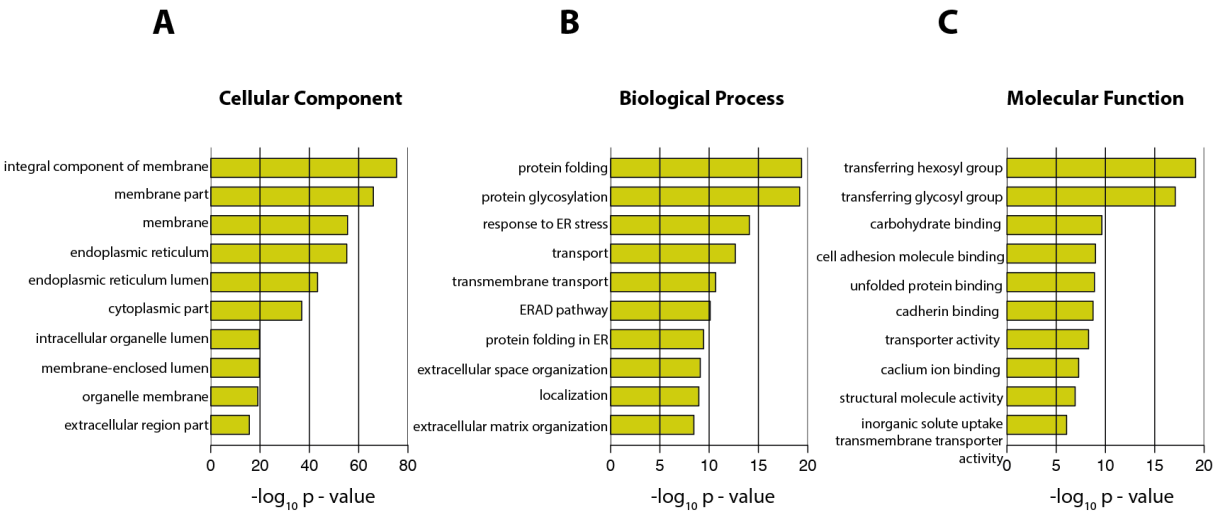



**Figure 34. GO term enrichment analysis of gp210 interactome**

GO-enrichment for (A) Cellular Component, (B) Biological process and (C) Molecular function, were calculated for the identified gp210 interactome.

I identified a number of proteins that are involved in the ER stress regulation and ER-associated degradation (ERAD) pathway, which is a part of unfolded protein response (UPR). Since the BirA*-gp210 fusion protein was not expressed at the endogenous level, this could indicate that

gp210 overexpression on its own may trigger ER stress and identified proteins are not specific gp210 interactors. In order to rule out such a scenario, I checked whether the UPR is triggered upon gp210 expression. For this I used the X box-binding protein 1 (XBP1) mRNA splicing assay (van Schadewijk et al., 2012). Accumulation of unfolded proteins in the ER triggers the phosphorylation of Inositol-requiring enzyme 1 (IRE1). Phosphorylated IRE1 then triggers the unconventional splicing of XBP1 (Yoshida et al., 2001). The XBP1 protein from alternatively spliced mRNA is then transferred to the nucleus and induces the transcription of ER-associated protein degradation (ERAD) pathway components. XBP1 mRNA splice variants differ in size, and therefore can be resolved on agarose gels (figure 35A). The presence of a lower molecular weight fragment indicates the activation of UPR. To investigate, whether the expression of gp210 triggers the UPR, cells with the full-length BirA*-gp210 were induced with tetracycline. After 48 h and 72 h (these times correspond to the addition of biotin and cell collection in BioID experiments), RNA was extracted and cDNA was generated by reverse transcription. In the following, the cDNA was used as a template for a PCR in which XBP1 was amplified. PCR products were analyzed on agarose gel. As a positive control, cells were treated with thapsigargin – an inhibitor of the sarco/endoplasmic reticulum $Ca^{2+}$ ATPase (SERCA) that is frequently used to induce the UPR and ER stress (Lytton et al., 1991). As shown in the Figure 35B, in case of thapsigargin treatment, the low molecular splice variant was detected indicating the induction of UPR as expected. For the mRNA isolated from cells expressing gp210, only higher molecular splice variant was detected, suggesting that the gp210 overexpression does not trigger the UPR.

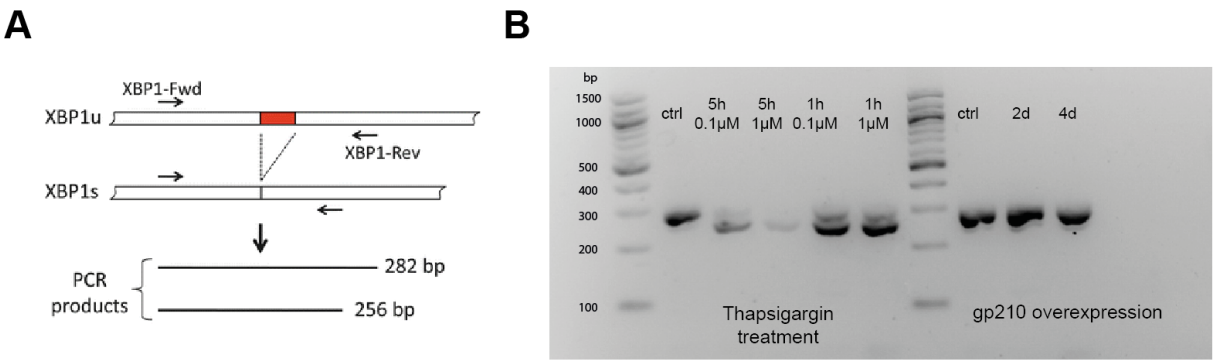



**Figure 35. ER stress control**

(A) Schematic representation of the XBP1 assay. Modified from (Kennedy et al., 2015). (B) The XBP1 assay was performed on cells either treated with thapsigargin (left panel) or with induced overexpression of gp210 (right panel).

### 3.3.4. Functional analysis of gp210 interactors

Identified gp210 interactome can be arbitrarily classified into three groups: (i) proteins involved in the gp210 polypeptide processing and (ii) proteins interacting with gp210 because of its function and (iii) non-interacting proteomic neighborhood of gp210.

The first group contains factors involved in the processing of the polypeptide chain such as translation, folding, post-translational modifications, proper localization of gp210 or its degradation if failed to fold properly. Proteins identified within this group presumably do not provide meaningful information about the function of gp210. Nevertheless, they indicate which pathways are involved in its processing (Figure 36). For example, I identified multiple ribosome components together with several signal recognition particle subunits and other proteins involved in the processing of ER targeting sequence. These are likely to be early interactions: i.e, taking place soon after the translation of BirA. Similarly, the presence of multiple folding chaperones, among all identified proteins, can be explained. Another identified network module contains the most of the members of the endoplasmic reticulum membrane complex (EMC) components. Although the exact function of this complex remains uncharacterized, it has been recently proposed that its role is to insert certain type of membrane proteins (with moderately hydrophobic membrane domain) into the ER membrane(Guna et al., 2018; Shurtleff et al., 2018). High enrichment of EMC components in gp210 interactome indicates that EMC may be also responsible for proper insertion of C-terminal trans-membrane domain of gp210. Finally, several components of endoplasmic reticulum associated degradation (ERAD) pathways components were identified. As shown before, gp210 overexpression on its own does not induce ER stress due to the increased accumulation of unfolded proteins, therefore most of these interactions are probably due to the components involved in degradation of naturally occurring, improperly folded proteins.

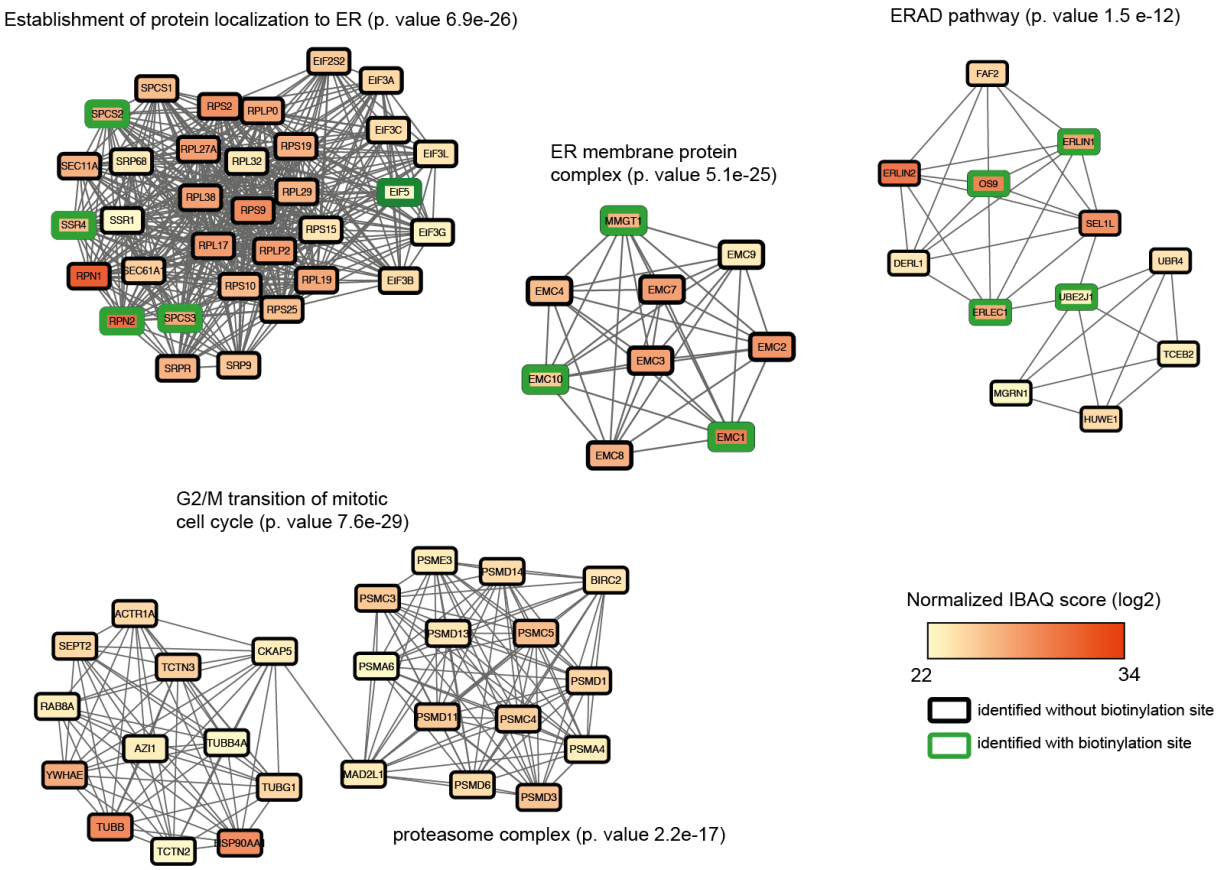



**Figure 36. Proteins involved in the processing of gp210 polypeptide chain**

An interaction network among gp210 interacting proteins was generated. Sub-networks of proteins that are involved in the processing of gp210 polypeptide chain are displayed. Proteins with biotinylation sites identified are marked with the green frame.

The second group consists of proteins that are likely to interact with gp210 because of its function (Figure 37). Three functionally related clusters can be distinguished within this group:

I) Nuclear envelope related proteins:

Several inner nuclear membrane (INM) proteins were identified such as Lemd3, Emerin, Lamin B-receptor and Lamina associated polypeptide proteins. These are known to interact with the nuclear lamina. Within this group, I also identified several components of the nuclear pore, which was expected since gp210 is a member of the NPC.

II) Collagen biosynthesis pathway:

The collagen biosynthesis pathway can be subdivided into three major steps: i) translation of the pre-collagen into ER, ii) hydroxylation of the proline and lysine residues by prolyl and lysyl hydroxylases, iii) glycosylation of hydroxylated lysine residues by procollagen galactosyltransferases. Several procollagen chains together with enzymes involved in both collagen hydroxylation and glycosylation were identified in the BioID experiment. Multiple

biotinylation sites identified for different proteins involved in the pathway strongly indicate that there is a connection of gp210 with collagen biosynthesis pathway.

III) GPI anchored proteins

Another group of functionally related proteins that have been identified in the BioID experiment is comprised of proteins involved in glycosylphosphatidylinositol (GPI) anchor processing. The gp210 interactome contains several proteins involved in the assembly of the GPI-anchor. In the ER, the C-terminal GPI attachment of the signal peptide is cleaved and replaced with the preassembled GPI. This process is mediated by the complex called GPI-anchor transamidase (Fraering et al., 2001). All 5 components of the complex were also identified in the BioID experiment.

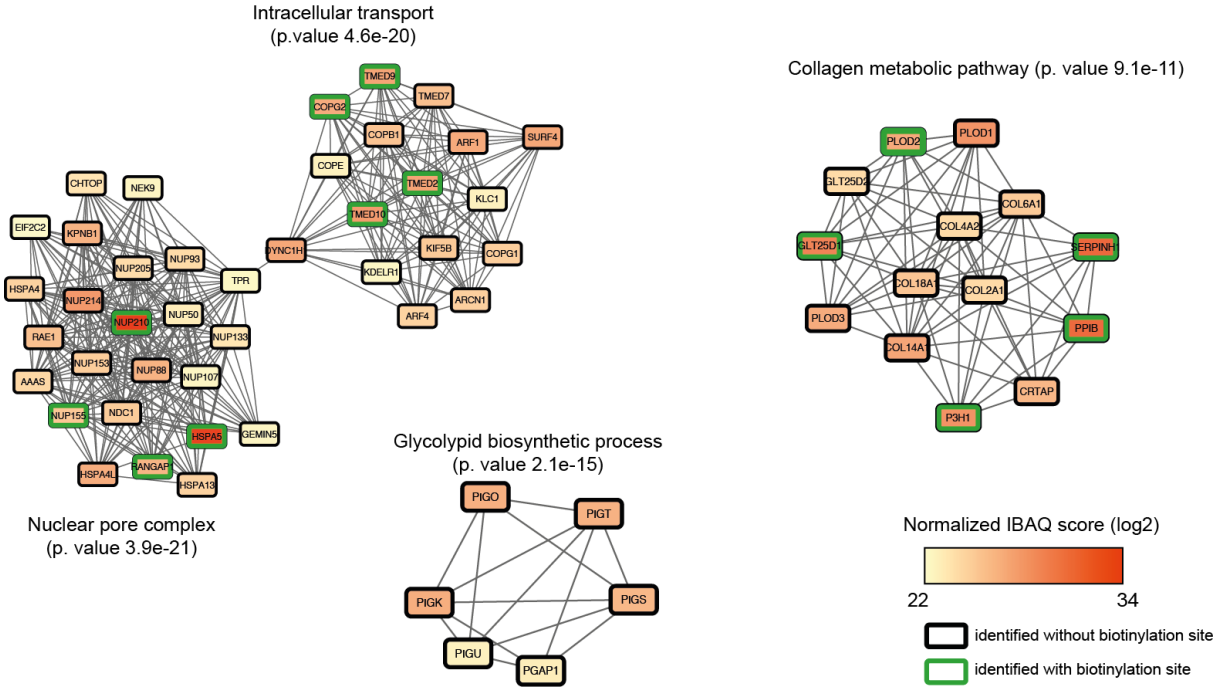



**Figure 37. Ptoteins potentially interacting with gp210**

A network among gp210 interacting proteins was generated. Sub-networks of proteins interacting with gp210 based on their function are displayed. Proteins with biotinylation sites identified are marked with the green frame.

Finally, the third group contains proteins that are in the close proximity of gp210, nevertheless these are unlikely to be its specific interactor. This group contains mainly highly abundant ER resident proteins, for example Calumenin and Calreticulin, which are found throughout the entire ER. At this point, it is important to mention that some of the proteins included in the previous groups may also be unspecific, nevertheless due to the functional relations to other proteins, they were classified as potential interactors of gp210.

### 3.3.5. Analysis of domain specific interactions

Next I checked whether there are any differences in interactomes of distinct regions of gp210. For this, I calculated the Pearson correlation of protein abundances observed in all samples. As shown in the Figure 38A, high correlation values were observed for proteome profiles of longer constructs (full length, Asn287, Ser869). The construct lacking the luminal domain of gp210 (Arg1789) seems to have different set of interactors as compared to remaining cell lines. Nevertheless, the majority of previously mentioned proteins (components of the EMC complex, or members of the collagen synthesis and GPI-anchor pathways) were also identified as part of the C-terminal domain interactors. The analysis of biotinylation sites of gp210 indicates that the BirA fusion proteins interact with their endogenous counterpart. For example, in the cell line expressing BirA and the C-terminal domain, biotin was still detected at the N-terminal part of the protein. It is therefore possible, that some of the interactions are identified because of the close proximity or oligomerization with the endogenous full-length gp210. A pairwise comparison between the Arg1789 and full-length constructs indicates that C-terminal part of gp210 mainly interacts with proteins involved in the nuclear envelope organization (Lemd3, Emerin, Lamin B-receptor, Lamina associated polypeptide-2 and SUN1 – a member of LINC complex). Although all of these proteins were identified in on-bead digestion in each BirA*-gp210 cell line, the biotinylated peptides were only detected for shorter constructs.

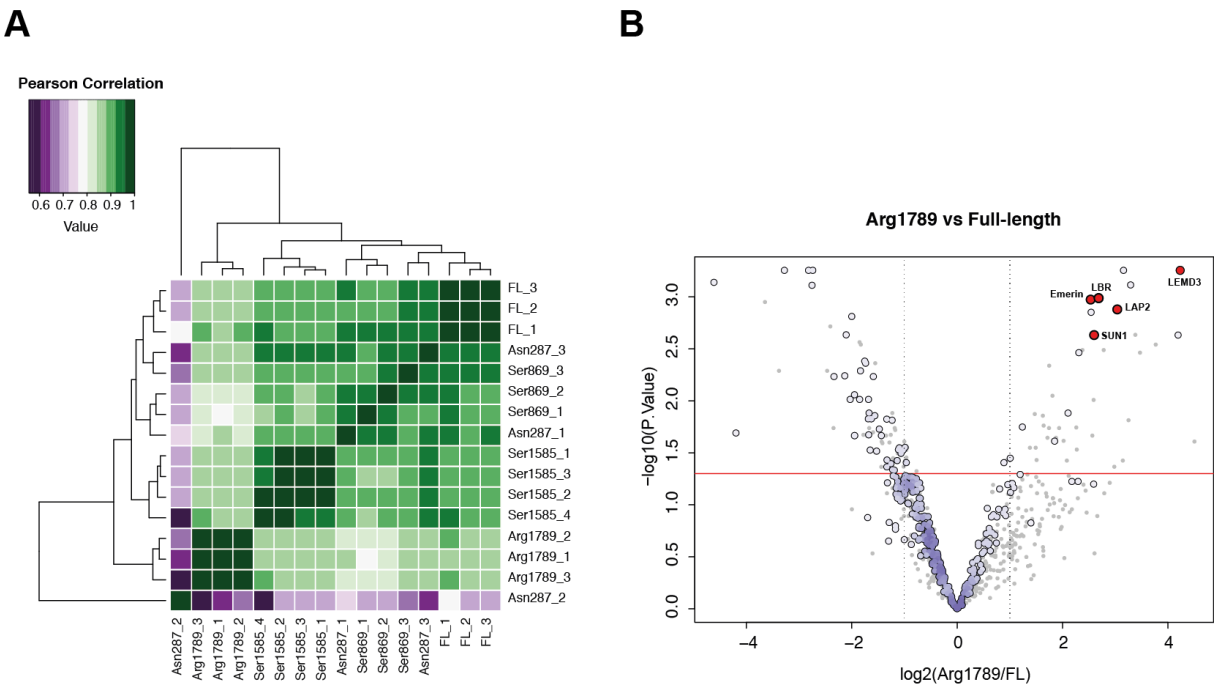



**Figure 38. Comparison of cell lines with different length of gp210**

(A) Heatmap depicting Pearson correlation coefficients of different gp210 constructs. One replicate of Asn287 cell line appears as an outlier. This is possibly a consequence of inefficient biotinylation. (B) Pairwise comparison of Arg1789 and full length constructs. Proteins with large fold changes (highlighted in red) are involved in the organization of the nuclear envelope.

### 3.3.6. Generation of CRISPR-Cas9 knock out of gp210

In order to investigate the impact of gp210 depletion, I used the CRISPR-Cas9 system to generate a gp210 knockout cell line. The target sequences were designed using the DESKGEN cloud online software tool (Hough et al., 2016). Sequences with top specificity and efficacy scores were further tested using an in-vitro cleavage assay to select the most prominent target. Hek293 cells were transfected with preassembled gRNA (guide RNA) -Cas9 complex to allow genome editing. To achieve homogenous cell populations, cells were subjected to the single cell sorting using flow cytometry 48 hours after transfection. After clonal expansion of isolated cells, the genomic DNA was extracted and DNA fragments containing the targeted sequence were amplified and analyzed by sequencing. Around 20% of isolated clones carried a mutation in the target region. Based on the sequencing results, 3 clones were selected for further analysis. In 2 of them heterozygous DNA was observed, suggesting that NHEJ (non-homologous end joining) introduced different mutations in different chromosomes. Nevertheless, deconvolution of sequencing chromatogram traces by TIDE (Brinkman et al., 2018) suggested that none of the chromosomes remained unmodified. In case of the third clone, a double nucleotide deletion was observed in both chromosomes.

To confirm the successful knockout of gp210, the generated cell lines were analyzed with immunofluorescence microscpy. Cells were stained with anti-gp210 antibody that targets its C-terminal domain and, as a control, with monoclonal antibody 414 (mab414) that recognizes FG (phenyloalanine-glycine) rich nucleoporins. As shown on the Figure 39, in case of heterozygous cell lines, a decrease in gp210 signal was observed, however the protein still could be detected. This possibly indicates that despite the mutations introduced, one of the copies still leads to the synthesis of functional protein. In case of homozygous cell line with a frame shift, no signal for gp210 was observed, providing an additional source of evidence for successful knockout. Additionally, cells with gp210 knockout (KO) had an aberrant morphology with frequently large and amorphous nuclei. Another feature observed for the gp210 KO cells, was the differential distribution of FG-nucleoporins. While in wild-type cells, these are mainly localized to the nuclear envelope, with only few spots observed outside of the NE, in the KO apart from NE

stain, the signal was strongly dispersed in KO cells. Although more detailed evaluation of this phenotype is required, it may indicate imperfect NPC assembly in these cells.

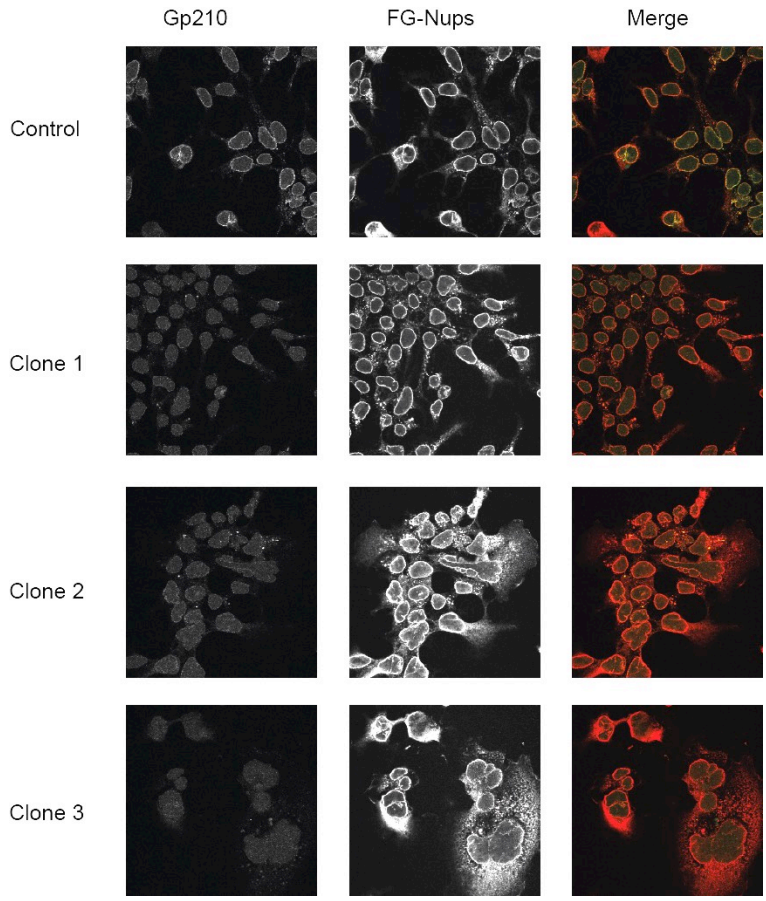



**Figure 39. Immunofluorescence of gp210 knockout cells**
Cells were stained with anti-gp210 antibody (green) and mab414 targeting FG-nups (red).

### 3.3.7. Proteomic characterization of gp210 knock out cell line

In order to investigate proteomic changes induced by the depletion of gp210, I collected wild-type (WT) and the KO cells from three consecutive passages. After protein extraction and digestion, peptides were labeled with TMT reagents, fractionated using high pH fractionation and analyzed using Orbitrap Fusion Lumos. In total 6301 proteins with at least two unique peptides were quantified. A summary of the quantitative data is displayed in the Figure 40. Although the overall proteome profiles of these cell lines were mostly similar, some significant differences were still detectable. Interestingly, multiple proteins previously identified as gp210 interactors were affected by the gp210 knockout further confirming the functional link between gp210 and non-NPC related processes. Among differentially expressed proteins I observed an enrichment of proteins involved in GPI anchor biosynthesis. In total, 10 out of 22 proteins

involved in GPI-anchor biosynthesis pathway were identified in this experiment. The abundance of 7 of them was significantly affected, with the majority of them being down regulated. All 5 components of GPI-transamidase complex were expressed at lower level in gp210-depleted cell line in comparison to the wild type condition.

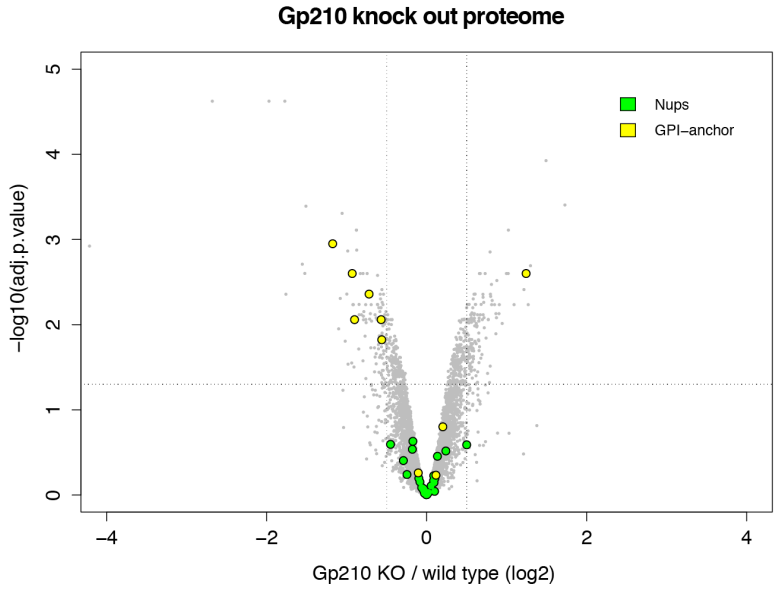



**Figure 40. Pairwise comparison of gp210 KO cell line and non-modified Hek293 cell line**

Proteins highlighted in green correspond to the quantified nucleoporins. Hits labelled in yellow are involved in the GPI-anchor processing.

## 3.4. Discussion

*BioID as a tool to map the protein interactions*

Several lines of evidence suggest that gp210 acts also away from the NPC, nevertheless it remains unclear how and in which other cellular pathways it is involved. To better understand the non-NPC related function of gp210, I attempted to identify its interactome. Traditionally, this is done by the affinity purification, followed by the MS analysis. However, gp210 is a transmembrane protein and because of that its native purification is challenging. In addition, it does not allow for the identification of transient interactions, which may by highly informative with regards of the function in case of such a dynamic protein. Instead, I used the BioID system, in which proteins that are in the close proximity of bait are biotinylated and therefore can be purified under more stringent conditions. Furthermore, neither direct nor indirect interactions between BirA* and proteins that become biotinylated are required. As such non-interacting but proximal proteins can also be detected providing the additional information about the proteomic neighborhood of the target. At the same time, this can be considered as a major drawback of the approach as it can lead to the number of unspecific identifications, which includes also proteins that are further than theoretical biotinylation radius. For example, if the number of exposed primary amines from truly interacting protein is low, or they are already modified, the free active biotinoyl-5'-AMP might diffuse further away, thus biotinylating surrounding non-specific proteins. This can for example happen if the labeling time in cell culture is too long. Nevertheless, even though the proximity labeling based approach does not ultimately prove the physical interaction, it indicates potential functional links to other proteins and therefore it is a valuable tool to characterize the proteins of elusive functions.

*Interaction of gp210 with the nuclear envelope proteins*

The NPC-associated portion of gp210 is in close proximity to proteins that specifically localize to the inner nuclear membrane. Several of such proteins were identified in the BioID experiment, including lamina interacting proteins such as Lemd3, Emerin, Lamin B-receptor and Lamina associated polypeptide-2. Interestingly, these proteins were significantly enriched in pull-down experiments of constructs that contain only the C-terminal domain of gp210. This suggests that the C- and N- terminal ends of gp210 are distant to each other. In the model proposed by Upla et al., 2017, the yeast homolog of gp210 (pom152) forms a head-to-tail dimer that oligomerizes as a ring around the INM at the site of NPC. If the same is true for human gp210, a similar protein

neighborhood would be expected for its C- and N- terminal ends. This, however does not seem to be the case and possibly suggests a different conformation of the dimer. Due to the presence of multiple IG-like fold repeats, gp210 is structurally similar to cadherins – a type of cell adhesion molecule, which dimerizes in a head-to-head manner (Troyanovsky et al., 2003). Therefore, it could be speculated that gp210 forms a similar type of dimer, and is not parallel to the membrane as proposed for yeast (Upla et al., 2017). Nevertheless, this hypothesis requires further validation.

*Gp210 and ER stress*

Among the identified interactome of gp210 I observed an enrichment of proteins involved in ER stress and the UPR including the key regulatory components of the pathway. During either physiological or pathological stimuli the proper ER homeostasis may be impaired resulting in the accumulation of misfolded proteins in the ER causing the cellular stress. Upon the ER stress three parallel signaling pathways can be activated initiating UPR. Each of them is mediated via the interaction with Binding immunoglobulin protein (BiP, also known as GRP78 or HSP70), which is a molecular chaperone located in the lumen of the ER (Dorner et al., 1987; Gething, 1999). In non-stress conditions, BiP is bound to the IRE1, PRKR-like endoplasmic reticulum kinase (PERK) and Activation transcription factor 6-alpha (ATF-6α) preventing their activation (Marcu et al., 2002; Wang and Kaufman, 2014). Upon accumulation of misfolded proteins, BiP is sequestered by newly synthetized polypeptides thus leading to the activation of aforementioned factors. BiP free IRE1 forms a homodimer and via autophosphorylation it activates its endoribonuclease activity. Activated IRE1 cleaves *Xbp1* mRNA initiating its alternative splicing, which in turn leads to the synthesis of transcription factors that regulates the expression of UPR related genes in the nucleus (Yoshida et al., 2001). Similarly, ER retention of BiP triggers the dimerization and autophosphorylation of PERK that can then phosphorylate Eukaryotic translation initiation factor 2A (EIF2A), which in consequence attenuate translation (Scheuner et al., 2001). Finally, dissociation of BiP from the ATF-6α releases it from the ER membrane. ATF-6α is transported to the Golgi apparatus, where it is cleaved by site-1-proteoase (S1P) protease to form an active transcription factor that induces the expression of genes involved for the ERAD pathway (Ye et al., 2000). Both BiP as well as its binding partners (IRE1, PERK, and ATF-6α) were identified in the BioID assay, however only for BiP biotinylated peptides were observed. Since BiP is an abundant ER chaperone involved in the folding of newly synthetized ER proteins, such interaction is expected. Therefore it is also

possible that the remaining factors were identified only because of the strong interaction with biotinylated BiP rather than due to the proximity to gp210 and thus it is not yet a convincing evidence for a role of gp210 in ER stress. The biotinylation sites (indicating the close proximity) were however identified also for two other proteins that seem to be involved in the negative regulation of the UPR, namely reticulocalbin 1 (RCN1) and nucleobindin 1 (NUCB2). RCN1 has been shown to suppress the ER stress induced apoptosis by affecting the PERK signaling (Xu et al., 2017), while NUCB2 inhibits the signal cascade triggered by the ATF-6α activation (Tsukumo et al., 2007). Taken together, the outcome of the BioID experiment points to the link between gp210 and regulation of the ER stress in the investigated experimental system. The overexpression of luminal proteins may on its own trigger the high folding demand and as such trigger ER stress. Such scenario was however excluded by the analysis of the *Xbp1* splice variants upon the gp210 overexpression, indicating that gp210 indeed may be involved in the ER stress. As a matter of fact, a previous study already has suggested that gp210 may regulate the ER stress during muscle cell differentiation, however no mechanistic details were provided (Gomez-Cavazos and Hetzer, 2015) and remain to be further investigated in the future.

*Gp210 and GPI-anchor biosynthesis pathway*

Among proteins classified as members of the gp210 interactome, I identified a group of functionally related proteins, which contains components of the ER resident GPI transamidase complex. This is an enzymatic complex that mediates the attachment of GPI moiety to the proteins, which later are going to be anchored to the outer leaflet of the plasma membrane(Kinoshita, 2014). GPI-anchored proteins are usually associated with the membrane rafts and they are proposed to be involved in the signal transduction(Saha et al., 2016). Interestingly, the abundance of components of the GPI transamidase complex as well as other proteins involved in the GPI-anchor biosynthesis pathway was significantly affected by the depletion of gp210 in Hek293 cell line. Taken together, these independent experiments, for the first time suggest that gp210 may be involved in the regulation of the GPI-anchor attachment pathway. However, its role in the process remains elusive and difficult to explain with the current state of knowledge. To draft any conclusions, further characterization of this link is required.

# 4. Materials and Methods

## 4.1. FFPE tissue proteomics

### 4.1.1. Source of tissue specimens

Formalin-fixed and paraffin-embedded tumor tissues were provided by the tissue bank of the National Center for Tumor Diseases (NCT, Heidelberg, Germany) in accordance with the regulations of the tissue bank and with the approval of the ethics committee of Heidelberg University and the European Molecular Biology Laboratory. Only tissue specimens of high quality (high tumor cell content, lack of significant necrotic and fibrotic changes, etc.) as judged by a trained pathologist were included and each selected tumor was re-evaluated regarding its grading. Patient characteristics (age, sex, pT-stage, tumor grading and etiology) are provided in the supplementary table 2. Fresh frozen tissue samples of murine HCCs were generated by Dr. Daniel Dauch and Prof. Lars Zender.

### 4.1.2. Laser microdissection of human HCC specimens

The specimens were cut on a microtome into 10 µm thick sections and processed as follows: sections were mounted on membrane slides (PEN-membrane, 1 mm glass, Carl Zeiss MicroImaging GmbH), deparaffinized in 2x xylene for 3 minutes, rehydrated in 2x 100% ethanol for 2 minutes and then washed in 90% (v/v), 70% (v/v) and 50% (v/v) ethanol. Finally, the sections were stained for 15 s in cresyl violet acetate (1 % (w/v) in ACS-grade ethanol (Sigma-Aldrich)). Subsequently, the slides were washed in 50% (v/v), 70% (v/v), 90% (v/v) and 100% ethanol and incubated for 5 minutes in xylene. After air-drying the slides were mounted onto the stage of an inverted microscope of a Microbeam LMPC System (Carl Zeiss MicroImaging GmbH). The RoboLPC method was employed to microdissect and capture the different tumor sectors, capsule and non-tumorous tissue. For each sector I collected ~40 $mm^2$ of tissue (400 nL).

### 4.1.3. Immunohistochemistry staining and evaluation

“Immunohistochemical staining was performed with an automated immunostaining instrument (BenchMark ULTRA IHC/ISH Staining module, Ventana Medical Systems). The OptiView DAB IHC Detection Kit (Ventana Medical Systems) was used based on the manufacturer’s protocol. The procedure included the following steps: 4 min deparaffinization at 62°C, rinsing

with EZ Prep (Ventana Medical Systems) and incubation with Cell Conditioner No. 1 (Ventana Medical Systems) for 40 min at 90°C. Primary antibody treatment was performed using the following antibodies: RAC1 (GeneTex) diluted 1:25, Decorin (Thermo Scientific) diluted 1:300, HEPAR 1 (Cell marque) and Ki67 (clone MIB1, DAKO) diluted 1:200 for 24 min at 36°C. This was followed by exposure to Optiview Peroxidase Inhibitor for 4 min, 12 min incubation with Optiview HQ Universal Linker, 12 min treatment in Optiview HRP Multimer, 8 min incubation with a mixture of Optiview $H_2O_2$ and DAB, 4 min exposure to Optiview copper The samples were counterstained with Haematoxylin for 12 min and incubated for 4 min with Bluing Reagent. These incubation steps were followed by multiple rinsing steps in reaction buffer manufactured by Ventana Medical Systems. Dehydration of each FFPE slide was performed as follows: 1 x 5 min 70% (v/v) ethanol, 1 x 5 min 96% (v/v) ethanol, 2 x 5 min 100% ethanol, 1 x 5 min Xylene by using the Leica autostainer XL. Finally the slides were mounted with cover slips (Leica CV5030)". Provided by dr. Stephan Singer

### 4.1.4. Protein solubilization for FFPE samples

Tissue sections were collected in PCR tubes containing 100 µL of protein solubilization buffer (80 µM Tris pH 8.0, 80 µM DTT and 4% SDS) and processed directly. Samples were sonicated using a Bioruptor (Diagenode) for 22.5 minutes (15 cycles: 1 min on, 30 sec off) at the highest settings, and then heated for 1 h at 99°C. Sonication followed by heating was performed twice. Cysteine residues were alkylated by adding 200 mM iodoacetamide to a final concentration of 15 mM (incubated for 30 min at room temperature in the dark). The reaction was quenched by addition of 10 µL 200 mM DTT.

### 4.1.5. Protein purification, digestion and peptide desalting of FFPE samples

Sera-Mag Speed Beads (#45152105050250 and #65152105050250, Thermo Scientific) were mixed 1:1, rinsed with water and stored as a 40 µg/µL stock solution at 4°C as described in (Hughes et al., 2014). 4 µL of beads stock was added to the reaction tube and mixed by pipetting, then 100% acetonitrile (ACN) was added to a final concentration of 50% (v/v). Samples were incubated for 8 minutes at room temperature to allow protein binding to the beads. Next, the tubes were placed into a magnetic rack. Supernatant was removed and discarded. The beads were washed twice with 180 µL of 70% (v/v) ethanol and once with 180 µL of 100% ACN. After removal of ACN, the beads were air-dried for 60 seconds and then resuspended in 7 µL of digestion buffer consisting of either 6 µL 4 M urea in 100 mM ammonium bicarbonate or 100 mM HEPES pH 8.5 for the TMT experiment and 1 µL of 0.1 µg/µL of LysC (Wako).

Samples were sonicated for 5 minutes in a water bath sonicator, incubated for 5 minutes at 37°C and then mixed by pipetting. Digestion was allowed to proceed for 4 h at 37°C. After the first step of digestion beads were resuspended by pipetting, urea was diluted to a final concentration of 1.5 M and 1 µL of 1 µg/µL of sequencing grade trypsin (Promega) (1 µg/µL of LysC for TMT-6plex experiment) was added to the samples. Digestion was performed for 12 h at 37°C. After digestion the beads were resuspended by pipetting. 100% ACN was added to a final concentration of 95% (v/v) and samples were incubated for 8 min at room temperature. The tubes were placed into a magnetic rack and washed twice with 100% ACN. The supernatant was removed and beads air-dried and reconstituted in 9 µL of 2% DMSO followed by 5 min of sonication in a water bath. Samples were resuspended by pipetting and placed into a magnetic rack. The supernatant containing the peptides was transferred into a fresh tube and acidified with 1 µL of 1% (v/v) formic acid.

### 4.1.6. TMT labeling

TMT-6plex (for the initial experiment, Thermo Scientific) or TMT-10plex (for the 5 tumor analysis) reagents were reconstituted in 100% ACN according to the manufacturer's instructions. 1 µL of 1 M HEPES pH 8.5 was added to 9 µL of digested and purified peptides. TMT labeling was performed by addition of 1 µL of the TMT reagent. After 30 minutes of incubation at room temperature a second portion of TMT reagent (1 µL) was added and the samples were incubated for another 30 minutes. The reaction was quenched with 1 µL of 20 mM lysine in 100 mM ammonium bicarbonate. 4 µL of beads stock solution was added to the sample. Peptides were bound to the beads, washed and eluted as described in the section 4.1.5. The labeled peptides were pooled together and fractionated.

### 4.1.7. High pH peptide fractionation for TMT labeled samples

Offline high pH reverse phase fractionation was performed using an Agilent 1200 Infinity HPLC System equipped with a quaternary pump, degasser, variable wavelength UV detector (set to 254 nm), peltier-cooled autosampler and fraction collector (both set to 10°C for all samples). The column used was a Gemini C18 column (3 µm, 110 Å, 100 x 1.0 mm, Phenomenex) with a Gemini C18, 4 x 2.0 mm SecurityGuard (Phenomenex) cartridge as a guard column. The solvent system consisted of 20 mM ammonium formate (pH 10.0) as mobile phase (A) and 100% ACN as mobile phase (B). The separation was accomplished at a mobile phase flow rate of 0.1 mL/min using the following linear gradient for the TMT-6plex experiment: 99% A for 2 min,

from 99% A to 37.5% B in 61 min, to 85% B in a further 1 min and held at 85% B for an additional 5 min, before returning to 99% A and re-equlibration for 18 min. Thirty seven fractions were collected along with the LC separation that were subsequently pooled into 16 fractions. A slightly modified gradient was used for the TMT-10plex experiment, whereby the LC separation time was 100 min (from 10% to 40% B (column reconditioning to 85% B for 5 mins and re-equilibration for 18 minutes (99%A) followed the gradient separation)) and 48 fractions were collected over this separation time, which were subsequently pooled again into 16 fractions. Pooled fractions were dried in a vacuum evaporator and then stored at -80°C until LC-MS/MS analysis.

### 4.1.8. Data acquisition of TMT labeled samples

For TMT 6-plex, samples were analyzed using nanoAcquity UPLC system (Waters) directly coupled to an Orbitrap Fusion (Thermo Fisher Scientific) using a Proxeon nanospray source. Samples were separated on the BEH C18 (2.5 µm, 75 µm x 500 mm) nanoAcquity UPLC column (Waters) using a stepwise gradient. Solvent A was water, 0.1% (v/v) formic acid and solvent B was ACN, 0.1% (v/v) formic acid. Peptides were separated at a constant flow of 0.3 µL/min at 55°C. During the elution step, the percentage of solvent B increased in a linear fashion from 5% to 7% in 10 minutes, then from 7% B to 30% B in a further 105 min and finally to 45% B by 130 min. Full scan MS spectra with mass range 300-1500 m/z were acquired in profile mode in the Orbitrap with resolution of 60000 FMWH. The filling time was set at maximum of 50 ms with an AGC target of $4x10^5$ ions. The instrument was operated in the top20 mode. The most intense ions from the full scan MS were selected for MS2 using quadrupole isolation (the isolation window was 1.6 Da). CID fragmentation was performed in the ion trap with normalized collision energy of 35% and an intensity threshold of $5x10^3$. A maximum fill time of 70 ms for each precursor ion was set, with an AGC target of $1x10^4$. MS2 data were acquired in centroid mode. Only multiply charged (2+, 7+) precursor ions were selected for MS2. The properties of the dynamic exclusion list were as follows: a maximum retention period of 40 s, a relative mass window of 7 ppm and exclusion of isotopes. For MS3 the precursor selection window was set to a range of 400-1300 m/z with an exclusion width of 30 m/z (high) and 5 m/z (low). Isobaric tag loss exclusion was set to Reagent=TMT. The most intense fragments from the MS2 experiment were co-isolated (isolation window 2Da, using Synchronus Precursor Selection = 10) and fragmented by HCD (collision energy, 65%). MS3 spectra were acquired in an Orbitrap mass analyzer over a mass range of 100-200 m/z with the resolution set to 30000. The maximum injection time was set to 100 ms with an AGC target of $1x10^5$ ions and 1 microscan. Data were

acquired in profile mode and the instrument was allowed to inject ions for all available parallelizable times.

A similar strategy was used for the acquisition of TMT-10plex experiment, with the following exceptions: The analytical column used for the LC separation was 250 mm wide and MS data acquisition took place on an Orbitrap Fusion Lumos (Thermo Fisher). Full scan MS spectra with a mass range of 375-1500 m/z were acquired in profile mode in the Orbitrap with a resolution of 60000 using quadrupole isolation. The filling time was set at maximum of 100 ms. HCD fragmentation was performed with collision energy of 35%. A maximum fill time of 50 ms for each precursor ion was set. MS2 data were acquired with a fixed first mass of 120 m/z. For the MS3, the precursor selection window was set to a range of 400-2000 m/z, with an exclusion width of 18 m/z (high) and 5 m/z (low). The most intense fragments from the MS2 experiment were co-isolated (using Synchronus Precursor Selection = 8). MS3 spectra were acquired in an Orbitrap mass analyzer over the mass range of 100-1000 m/z and resolution set to 50000. The maximum injection time was set to 105 ms and the instrument was set not to inject ions for all available parallelizable times.

### 4.1.9. TMT data processing

TMT-6plex data were processed using Proteome Discoverer v1.4 (Thermo Fisher Scientific). Data were searched against the Uniprot Human FASTA database (release 2014_07, 20230 entries) using Mascot v2.2.7 (Matrix Science) with the following settings: Enzyme was set to LysC, with up to 1 missed cleavage. MS1 mass tolerance was set to 10 ppm and MS2 to 0.5 Da. Carbamidomethyl cysteine was set as fixed modification and oxidation of Methionine as variable. Other modifications included the TMT-6plex modification from the quan method used. The quan method was set to reporter ions quantification with HCD and MS3 (mass tolerance, 20 ppm). The false discovery rate for peptide-spectrum matches (PSMs) was set to 0.01 using Percolator (Brosch et al., 2009).

TMT-10plex data were processed using Proteome Discoverer v2.0 (Thermo Fisher Scientific). Data were searched against the Swissprot Human FASTA database (release 2016_11, 20211 entries) using Mascot v2.5.1 (Matrix Science) with the following settings: Enzyme was set to trypsin, with up to 1 missed cleavage. Other settings were as for TMT-6plex search data, with the exception of the modifications from the quan method, which was set to TMT10 and Acetyl (Protein N-term) as a variable modification.

Reporter ion intensity values for the filtered PSMs were exported and processed using R scripts (written in-house) to remove common contaminants and decoy hits. Additionally only PSMs having reporter ion intensities above $1x10^3$ in all relevant TMT channels were retained for quantitative analysis.

### 4.1.10. Data acquisition of DIA samples

For DIA, samples were analyzed using EASY nano-LC directly coupled to a Q-Exactive HF mass spectrometer using the nanoelectrospray ion source (all from Thermo Fisher Scientific). Samples were separated on a RP-HPLC column (75 µm x 500 mm) packed in-house with 1.9 µm C18 resin (Reprosil-AQ Pur) using a linear gradient. Solvent A was water, 0.15% (v/v) formic acid, 2% (v/v) ACN and solvent B was ACN 0.15% (v/v) formic acid, 2% (v/v) water. Peptides were separated at a constant flow of 0.2 µL/min at 60°C. During the elution step the percentage of solvent B increased in a linear fashion from 5% to 30% solvent B over 120 min. For spectral library generation (DDA) full scan MS spectra with a mass range of 350-1600 m/z were acquired in profile mode in an Orbitrap with a resolution of 120000 FWMH (at 200 m/z). The filling time was set to a maximum of 100ms with an AGC target of $3x10^6$ ions. The instrument was operated in top20 mode. The most intense ions from full scan MS were selected for MS2, using quadrupole isolation (the isolation window was 1.4 Da). HCD fragmentation was performed in the HCD cell with normalized collision energy of 28% and an intensity threshold of $2x10^3$. A maximum fill time of 50ms for each precursor ion was set with an AGC target of $1x10^5$. MS2 data were acquired in centroid mode using an Orbitrap mass analyzer with a resolution of 15000 FWHM (at 200 m/z). Only multiply charged (2+, 7+) precursor ions were selected for MS2. The dynamic exclusion list was with a maximum retention period of 30 sec.

For data-independent acquisition (DIA) analysis the same LC-MS platform and settings with a few modifications were employed. Specifically, a survey scan at a resolution of 120,000 FWHM (at 200 m/z) using a maximum of $5x10^6$ ions and 100 ms injection time was followed by 38 DIA mass windows acquired at a resolution of 30,000 FWHM (at 200 m/z) accumulating a maximum of $3x10^6$ ions and using an automated injection time. The mass range scanned was from 400 to 1,220 m/z and a stepped normalized collision energy (22.5, 25 and 27.5) was employed. 38 overlapping mass windows (Bruderer et al., 2015) splitting each mass window into equal halves were employed to cover the mass range of interest from 400 to 1,200 m/z.

#### 4.1.11. DIA data processing

A spectral library was generated by acquiring 5 shotgun runs (one for each tumor sector). Raw files were processed using MaxQuant (version 1.5.2.8) (Cox and Mann, 2008) The search was performed against the human UniProt FASTA database (release 2014_07, 20230 entries) using the Andromeda search engine (Cox et al., 2011) and the following search criteria: enzyme was set to trypsin with up to 2 missed cleavages; Carbamidomethylation (C) as a fixed modification; oxidation (M) and acetylation (protein N-term) were set as variable modifications; mass tolerance of 10ppm (precursor ions) and 0.02 Da (fragment ions); and the minimal peptide length to 7 amino acids. The false discovery rate was set to < 0.01. A spectral library was generated in Spectronaut (Biognosys AG) using the default settings.

DIA files were searched in Spectronaut against the generated spectral library using the default settings. For quantification only peptides with a q-value < 0.01 and a signal to noise ratio > 20 were selected and exported.

#### 4.1.12. Quantitative analysis of FFPE samples

Both TMT and DIA data were analysed using the same R procedures based on the MSnbase package (Gatto and Lilley, 2011). Reporter ion (TMT) and peptide (DIA) intensities were log2-transformed and normalized using the vsn package (Huber et al., 2002). Peptide-level data were summarized into their respective protein groups by taking the median value. For differential protein expression, each patient-sample was treated individually. Protein ratios were calculated for all the protein groups quantified with at least 2 peptides. The R-package ‘‘fdrtool’’ (Strimmer, 2008) was used to fit a two components model on the median centered log2 ratio distributions using the statistic ‘normal’. Protein groups with a ratio belonging to the alternative component (q-value < 0.2) were considered as differentially expressed between the conditions tested.

#### 4.1.13. Protein solubilization, digestion and peptide desalting for freshly frozen murine HCC

Freshly frozen tissue samples of murine HCCs (~60 mg per sample) were homogenized by bead beating in ice-cold PBS using a Precellys24 homogenizer (6,000 rpm, 30 s, repeated twice). Tissue debris were removed be a quick spin, and the supernatant was transferred to a fresh tube. To solubilize proteins 0.2 volumes of 2% Rapigest (Waters) in 10 mM AmBic was added. Samples were sonicated for 3x30 s (1 minute on ice in between the cycles). Subsequently 0.8

volumes of 10 M Urea in 250 mM AmBic were added to a final concentration of 4 M Urea. Samples were sonicated in the same way as before. To reduce the samples, DTT was added to a final concentration of 10 mM and samples were incubated 30min at 37°C. Cysteine residues were alkylated by addition of iodoacetamide to a final concentration of 15 mM (incubated 30 min at RT in the dark). For protein digestion, LysC (Wako) was added in 1:50 w/w (enzyme to protein) concentration and incubated for 4 h at 37°C with 600 rpm shaking. Urea was then diluted to 1.5 M and sequencing grade trypsin (Promega) was added in 1:100 w/w concentration. Samples were incubated overnight at 37°C with 600 rpm shaking. After digestion, samples were acidified with 10% trifluoroacetic acid (TFA, to ~pH 2) and incubated for 30 min at 37°C to allow Rapigest cleavage. Peptides were clarified by centrifugation at 17000 g for 5 min and transferred to fresh tubes.

Peptides were purified using C18 macro-spin colums (Harvard Apparatus). The resin was activated with 100% methanol and washed twice with wash buffer (5% ACN, 0.1% formic acid). Samples were loaded onto the column and washed four times with wash buffer. Peptides were eluted twice with elution buffer (50% ACN, 0.1% formic acid). The solvent was removed using vacuum evaporator. Peptides were reconstituted in MS buffer (5% ACN, 0.1% formic acid).

#### 4.1.14. Data acquisition for LFQ

For shotgun analysis samples were analyzed using a nanoAcquity UPLC system (Waters) directly coupled to an LTQ-Orbitrap Velos pro instrument (Thermo Fisher Scientific) using a Proxeon nanospray source. Samples were separated on a BEH C18 (1.7 µm, 75 µm x 500 mm) nanoAcquity UPLC column (Waters) using a stepwise gradient. Solvent A was 0.1% (v/v) formic acid in water and solvent B was 0.1% (v/v) formic acid in acetonitrile. Peptides were separated at a constant flow of 0.3 µL/min at 55°C. During the elution step the percentage of solvent B increased in a linear fashion from 3% to 7% in 10 minutes, from 7% B to 25% B by 110 min and from 25% to 40% by 120 min. Full scan MS spectra with mass range 375-1600 m/z were acquired in profile mode in an Orbitrap mass analyzer with a resolution of 30000 FWMH (at 200 m/z). The instrument was operated in top20 mode. The top 20 most intense ions were CID fragmented in the ion trap with a normalized collision energy of 35% and analyzed in the LTQ. A maximum fill time of 50 ms for each precursor ion was set with an AGC target of $1x10^6$. Only multiply charged (2+, 7+) precursor ions were selected for MS2. The dynamic exclusion list was with a maximum retention period of 60 sec.

### 4.1.15. Data processing for LFQ

Raw files were processed using MaxQuant (version 1.3.0.5) (Cox and Mann, 2008). The search was performed against the mouse Ensembl database (GRCm38.70, 50879 entries) using the Andromeda search engine (Cox et al., 2011) with the following search criteria: enzyme was set to trypsin with up to 2 missed cleavages; Carbamidomethylation (C) as a fixed modification; oxidation (M) and acetylation (protein N-term) were set as a variable modifications; mass tolerance of 20 ppm (precursor ions) and 0.5 Da (fragment ions); minimal peptide length of 7 amino acids. The reversed sequences of the target database were used as a decoy database. Peptide and protein hits were filtered at a false discovery rate of 1%. Protein quantification was performed using the label-free quantification (LFQ) function of MaxQuant and the match between run option was selected using a time window of 2 min. LFQ values were extracted from the protein group table, log2 transformed and normalized by quantile normalization using the preprocessCore library (Gentleman et al., 2004). For each murine HCC protein fold changes were calculated against an average LFQ value measured from independent normal liver samples obtained from three different mice.

### 4.1.16. Quantification of mtDNA level by qPCR analysis

Genomic DNA (including mtDNA) was isolated with the QIAamp DNA FFPE tissue kit. (Qiagen). A total of 20 ng was used as template for qPCR with SYBR Green PCR Master Mix (Applied Biosystems). qPCR reaction was performed according to the following protocol: 1× 95 °C – 10 min (DNA denaturation and polymerase activation); 40 alternations of 95 °C − 15 s (melting) and 60 °C – 1 min (annealing/extension/read fluorescence). Mitochondrial DNA abundance was estimated based on the mitochondrial genes *MT-RNR1* and *MT-TL1* and normalized to a gene localized in the nucleus, B2M. Each qPCR reaction was performed twice to control for experimental errors. CT values were averaged from two technical replicates. The primers used for the analysis are listed in the Table 2.

**Table 2. Primers used for the mtDNA content analysis**

| Gene name | Forward primer | Reverse primer |
|---|---|---|
| *MT-RNR1* | CCACGGGAAACAGCAGTGAT | CTATTGACTTGGGTTAATCGTGTGA |
| *MT-TL1* | CACCCAAGAACAGGGTTTGT | TGGCCATGGGTATGTTGTTA |
| *B2M* | TGGCCATGGGTATGTTGTTA | TGCTGTCTCCATGTTTGATGTATCT |

### 4.1.17. Survival risk prediction

"Survival risk prediction (SRP) was performed in a cohort of 247 HCC patients mainly with a history of hepatitis B virus (HBV) infection or HBV-related liver cirrhosis from the Liver Cancer Institute and Zhongshan Hospital (Fudan University, Shanghai, China). Disease-free survival and gene expression profiles were available at the GEO repository with accession number GSE14520 for 241 patients (Roessler et al. 2010). SRP of the gene expression data was performed with the BRB-Array Tools software (http://linus.nci.nih.gov/BRB-ArrayTools.html; Version 4.5.0) as described previously (Roessler et al. 2010). Briefly, genes whose expression was significantly related to survival were identified by applying univariate Cox proportional hazards regression followed by principal component analysis. This resulted in a regression coefficient (weight) related to survival time based on two principal components. Next, to compute a prognostic index, the weighted average of the principal component values was calculated, using the regression coefficients derived from the Cox regression, described above. Finally, samples were split into two groups (n=199 and n=122) by the median of the prognostic index calculated from the expression of the genes identified by the proteomic analysis. Thereby, a high value of the prognostic index corresponded to a high value of hazard of death (high risk), and consequently a relatively poor predicted survival. Kaplan-Meier survival curves for the predicted cases to have above average risk and the cases predicted to have below average risk were plotted. In order to evaluate the predictive value of the method, 10-fold cross-validation with 1000-fold random permutation of the Cox-Mantel log-rank test was performed. Kaplan-Meier survival analysis was performed using GraphPad Prism software 6 (GraphPad Software, San Diego, CA) and the statistical p values were generated by log-rank test." Provided by dr. Stephanie Roessler

### 4.1.18. Next generation sequencing

"The multiplex PCR-based Ion Torrent AmpliSeqTM technology (Life Technologies) with the Comprehensive Cancer Panel (IonTorrent / Thermo Fisher Scientific, Waltham, USA), covering more than 400 cancer-relevant genes and a modified HCC-specific panel (including 29 genes), were used for library preparation. Amplicon library preparation was performed with the Ion AmpliSeq Library Kit v2.0 using approximately 40 ng of DNA. Briefly, 10 ng DNA were mixed with each of the 4 primer pools, containing all primers for generating ~16.000 amplicons, and the AmpliSeq HiFi Master Mix and were transferred to a PCR cycler (BioRad). After completion

of the PCR reaction primer end sequences were partially digested using FuPa reagent, followed by ligation of barcoded sequencing adapters (Ion Xpress Barcode Adapters, Life Technologies). Each individual primer pool was purified using AMPure XP magnetic beads (Beckman Coulter) and quantified using qPCR (Ion Library Quantitation Kit, Thermo Fisher Scientific) on a StepOne qPCR machine (Thermo Fisher Scientific). The individual library pools were diluted to a final concentration of 100 pM. In total 6 to 8 samples were pooled and processed for library amplification on Ion Spheres using an Ion PI™ Hi-Q OT2 200 Kit. Un-enriched libraries were quality-controlled using Ion Sphere quality control measurement on a Qubit instrument (Life technologies). After library enrichment (Ion OneTouch ES) the library was processed for sequencing using the Ion Torrent Hi-Q sequencing 200 chemistry and the barcoded libraries were loaded onto a PI v3 chip and sequenced on an IonTorrent Proton instrument." Provided by Dr. Volker Endris

#### 4.1.19. Variant calling and annotation

"Data analysis was performed using Ion Torrent Suite Software (version 4.4.3). After base calling the reads were aligned against the human genome (hg19) using the TMAP algorithm within the Torrent Suite. Variant calling was performed with the variant caller plugin within the Torrent Suite Software and the IonReporter package using a corresponding bed-file containing the coordinates of the amplified regions. Only variants with an allele frequency > 5% and minimum coverage > 100 reads were taken into account. Variant annotation was performed using ANNOVAR (Wang et al., 2010). Annotations included information about nucleotide and amino acid changes of RefSeq annotated genes, COSMIC and dbSNP entries, as well as detection of possible splice site mutations. IGV browser was used to visualize the aligned reads for data interpretation and verification (Broad Institute) (Bader and Hogue, 2003) "Provided by Dr. Volker Endris

### 4.2. Functional characterization of gp210

#### 4.2.1. Insertion of the BirA* sequence into the pDONR221-gp210 vector

Using restriction free cloning (Bond and Naus, 2012) the pDONR221-gp210 vector (previously generated in the lab by Dr. Amparo Andres-Pons) was modified to introduce the BirA* sequence after the signal peptide of gp210. In the first step a so-called megaprimer was generated by amplifying the BirA*-FLAG-tag sequence (available in the lab) with overhangs that are complementary to the sequences flanking the insertion site on the pDONR221-gp210 primer.

The reaction mix was prepared by mixing: 1 µL of plasmid containing the BirA*-gp20 sequence (50 ng/µL), 2.5 µL of 10 µM forward primer, 2.5 µL of 10 µM reverse primer, 25 µL of 2x Phusion High-Fidelity PCR Master Mix with GC buffer (Thermo Fisher) and 19 µL of water. Primers used for the PCR are listed in the Table 3. The following PCR program was used: 30 seconds of initial denaturation at 98°C, 35 cycles of denaturation (98°C, 10 seconds), annealing (57°C, 20 seconds) and extension (72°C, 30 seconds), followed by a final extension step (72°C, 5 min).

**Table 3. Primers used to insert BirA* sequence into gp210**

| Primer | Sequence |
|---|---|
| BirA*_forward | GGCCCCTCCGCCGCTGCGGCCAAGGACAACACCGTGCCC |
| BirA*_ reverse | CAGCACTTTGGGGATGTTGAGCTTTTTATCGTCATCGTCTTTGTAGTCT |

The PCR product was purified using the QIAquick PCR Purification Kit (Qiagen). Next, the megaprimer was incorporated into the pDONR221-gp210 vector in the secondary step PCR. The reaction was prepared by mixing 100 ng of pDONR-gp210 and 250 ng of purified megaprimer. Reaction volume was adjusted to 10 µL with water and 10 µL of 2x Phusion High-Fidelity PCR Master Mix with GC buffer (Thermo Fisher) was added. The following PCR program was used: 30 seconds of initial denaturation at 98°C, 15 cycles of denaturation (98°C, 8 seconds), annealing (61°C, 20 seconds) and extension (72°C, 10 minutes), final extension (72°C, 20 minutes). After PCR the template DNA was digested by adding 1 µL of restriction enzyme DpnI. The reaction was allowed to proceed for 4 h at 37°C. Subsequently the enzyme was heat inactivated for 20 min at 80°C. Chemically competent *E. coli* DH5α cells were transformed with the generated pDONR-BirA*-gp210 vector and plated on LB-Agar containing kanamycin. A single colony was expanded in liquid culture and plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen).

As a next step an LR recombination reaction was performed using pDONR-BirA*-gp210 clone and pDEST-TFG vector (Flp-In T-Rex compatible) according to the manufacturer's protocol. Chemically competent *E. coli* DH5α cells were transformed with the generated pDEST-TFG-gp210 vector and plated on the LB-Agar containing ampicillin. A single colony was expanded in liquid culture and plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen).

### 4.2.2. Generation of truncated pDEST-BirA*-gp210 constructs

Q5-site directed mutagenesis (New England Biolabs) was used to generate the truncated constructs. Primers used for the PCR are listed in the Table 4. Reaction mix was prepared by mixing 0.1 µL of pDEST-BirA*-gp210 (270 ng/µl), 1.25 µL of 10 µM forward and reverse primers (see Table 4), 12.5 µL of 2x Phusion High-Fidelity PCR Master Mix with GC buffer (Thermo Fisher) and 9 µL of water. The following PCR program was used: 30 seconds of initial denaturation at 98°C, 30 cycles of denaturation (98°C, 10 seconds), annealing (60°C, 20 seconds) and extension (72°C, 6.5 min), final extension (72°C, 20 min). The KLD reaction was prepared by mixing: 1 µL of PCR product, 5 µL of KLD reaction buffer (2X), 1 µL of KLD enzyme mix (10x) and 3 µL of nuclease free water. The reaction was incubated for 5 minutes at room temperature. Half of the reaction was used for transformation of chemically competent *E. coli* DH5α cells. A single colony was expanded in liquid culture and plasmids were purified using QIAprep Spin Miniprep Kit (Qiagen).

**Table 4. Primers used for generation of the truncated gp210 constructs**

| Construct | Forward Primer | Reverse Primer |
|---|---|---|
| Asn 287 | AGCATCCCGGGCCCCGAAGGA | TTTATCGTCATCGTCTTTGTAGTCTGGC |
| Ser 869 | TCTGCCAGAACAAAGCAGCCGCA | TTTATCGTCATCGTCTTTGTAGTCTGGC |
| Ser 1589 | TCTAACCTGAGAGGCGAGTGCA | TTTATCGTCATCGTCTTTGTAGTCTGGC |
| Arg 1789 | CGTGGGCCCGGTCCTTATGGA | TTTATCGTCATCGTCTTTGTAGTCTGGC |

### 4.2.3. *E. coli* transformation

For all transformation reactions NEB® 5-alpha Competent *E. coli* (High Efficiency, New England Biolabs) cells were used. 15 µL of cell stock was used for the reaction. Cells were thawed on ice for 10 minutes. 2-3 µL of plasmid was added to the cells and mixed by gentle flicking. The reaction was incubated for 30 minutes on ice. For heat shock cells were placed in a water bath (temperature set to 42°C) for 30 seconds and subsequently placed on ice for 5 minutes. 700 µL of SOC medium at room temperature was added to the reaction and cells were incubated for 1 h at 37°C (with 250 rpm shaking). Cells were spun down at room temperature (3 min, 3000 g) and the supernatant was removed. The pellet was resuspended in 50 µL of sterile water and placed on plates with the corresponding antibiotics. Plates were incubated overnight at 37°C.

### 4.2.4. Mammalian cell culture

Modified human embryonic kidney cells 293 (HEK Flp-In™ T-Rex™ 293 Cell Line, Life Technologies) designed for rapid generation of stably transfected cell lines with tetracycline inducible expression system were used. Cells were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 5 g/L glucose and 10% heat inactivated fetal bovine serum (FBS, Sigma-Aldrich). Cells at 90% confluency were trypsinized with 0.25% trypsin with EDTA (Life technologies). Every second cell passage the following selection antibiotics were added to the culture medium: Blasticidin (15 µg/mL) and zeocin (100 µg/mL) for parental cell lines, blasticidin (15 µg/mL) and hygromycinB (100 µg/mL) after transfection. Cells were grown at 37°C in 5% $CO_2$.

### 4.2.5. Stable cell line generation

24h prior to transfection, $2x10^4$ cells were seeded per well on a 6-well plates (no selection antibiotics were used at this stage). For transfection 100 µL of DMEM (without FBS), 3 µL of X-tremeGENE9 transfection reagent (Roche), 100 ng of pDEST-TFG and 900 ng of pOG44 plasmid (encoding a recombinase allowing for the integration of the plasmid to a modified site of genomic DNA, Life Technologies) were mixed, incubated 15min at room temperature and added to cells. Reaction without any pDEST-TFG vector was used as negative control. 24h later cells were trypsinized onto 150 mm diameter dishes. On the following day selective antibiotics were added (blasticidin, hygromycinB). Twice a week the medium was replaced and fresh antibiotics were added. Selection was carried out until there were no surviving colonies in the negative control dish.

### 4.2.6. Immunofluorescence

All steps of staining were performed at RT. In between each step cells were washed 3 times with PBS. First, cells were fixed with 2% paraformaldehyde (PFA) in PBS for 15 minutes, washed and then permeabilized with 0.2% triton x-100 in PBS for 15 minutes. As a next step, blocking in blocking buffer (2% BSA, 2% FBS in PBS) was performed for 1h. Cells were stained with primary and subsequently secondary antibodies diluted in blocking buffer for 1h each. Primary antibodies used were mouse mAb414 (1:2500, Covance) and rabbit anti-gp210 (1:500, Glavy lab, Stevens Institute of Technology). Secondary antibodies used were anti-mouse Alexa Fluor 647, anti-rabbit Alexa Fluor 488 (both 1:1000, Life Technologies). To visualize biotinylated proteins streptavidin conjugated with Alexa Fluor 647 (1:1000, Thermo Fisher Scientific) in 0.1% BSA in PBS was used. Labeling was performed for 10 minutes. To preserve the stained

cells cover slides were mounted in ProLong™ Diamond Antifade Mountant (Thermo Fisher Scientific). Mounted coverslips were dried for 24 h at RT and then stored at -20°C until imaging.

### 4.2.7. BioID pull down experiments

$2x10^6$ cells were seeded per 500$cm^2$ cell culture dish. On the following day tetracycline (Sigma-Aldrich) was added (final concentration 1 µg/mL) to induce the expression of BirA*-gp210 constructs. After 48 h biotin (Sigma-Aldrich) was added to a final concentration of 50 µM. Biotin labeling was allowed to proceed for 48 h. After this time cells were collected ($4x10^7$ per pull-down), snap frozen in liquid nitrogen and stored at -80°C until further processing. Cell pellets containing $4x10^7$ cells were resuspended in 9.5 mL of lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% triton x-100, 10 µg/mL Aprotinin, 10 µg/mL Leupeptin, 25 U/mL Turbonuclease (Accelagen), 1 mM PMSF and 0.1% SDS). Cells were lysed for 1 h at 4°C rotating at 15 rpm. After this time samples were sonicated 5x 30 s on/off using the Bioruptor (Diagenode) at 4°C. To remove cellular debris cells were spun at 17000 g for 30 min at 4°C. Supernatant was collected and 80 µL of streptavidin-sepharose slurry beads (GE Healthcare) were added to the lysates. Beads were previously equilibrated with lysis buffer. To allow biotin binding to the beads samples were incubated for 3 h at 4°C, rotating at 15 rpm. After this time beads were spun down at 2000 g for 5 min. Supernatant was removed and beads were transferred to Snap Cap Pierce Spin Columns (Thermo Fisher Scientific). Beads were first washed with 800 µL of lysis buffer and then 5 times with 800 µL of 50 mM AmBic, pH 8.3. Using 3x 300µL of 50 mM AmBic beads were transferred to the 2 mL tubes. Samples were spun down at 2000 g for 5 min and around 700 µL of supernatant was removed. 1 µg of trypsin (mass spectrometry grade, Promega) was added and samples were incubated for 16 h at 37°C, shaking at 500 rpm. After this time an additional 0.5 µg of trypsin was added and digestion was allowed to proceed for another 2 h. Samples were then transferred to new Snap Cap Pierce Spin Columns (Thermo Fisher Scientific) and supernatant was collected. Beads were washed with 2x 150 µL of 50 mM AmBic (each time the eluted liquid was collected). To elute the biotinylated peptides columns were transferred to new collection tubes and 2x 150 µL of 80% ACN and 20% TFA was added to the beads. Beads were pipetted up and down 5 times and the eluted liquid was collected. Samples were then dried using speed-vac concentrator. Peptides were reconstituted with 200 µL of 5% ACN and 0.1% TFA in water and desalted as described in the section 4.1.13 (in this case micro-spin columns were used).

### 4.2.8. BioID shotgun proteomics

For on-bead digestion data were acquired with the same acquisition parameters as described in section 2.5.14. In case of peptides eluted from the beads the acquisition parameter were the same with the exception of the gradient length (60 instead of 120 min).

### 4.2.9. BioID data analysis

Raw data were processed in MaxQuant using the same strategy as described in section 2.5.15. with the following exceptions: (i) Human Swiss-Prot entries of the Uniprot KB (database release 2016_09, 19594 entries) was used, (ii) carbamidomethylation (C) as fixed modification was not set, (iii) N-term or (K) biotinylation was set as variable modification. Only proteins identified with at least 2 unique peptides were retained. For quantitative analysis of the on-bead digestion experiment a proteinGroups.txt output file was used. LFQ values were extracted and subjected to quantile normalization. For pairwise comparison the significance of differential expression was assessed using the Limma package in R (Smyth, 2005). To define the gp210 interactome proteins that were (i) identified in at least 2 biological replicates of at least one construct and never detected in the control samples or (ii) enriched in BirA cell line in comparison to the control (log2 of fold change > 2, adj. p.value < 0.05) were extracted.

### 4.2.10. XBP1 mRNA splicing assay

5 days prior the RNA extraction, $5x10^2$ cells with inducible expression of BirA*-gp210 (full-length) were seeded per well on a 6-well plate. Expression of recombinant protein was induced with 1 µg/mL tetracycline (Sigma-Aldrich) for either 48 h or 72 h (these times correspond to the addition of biotin and cell collection in BioID experiments). WT cell line was used as negative control. For positive controls, WT cells were treated for 1 h and 5 h with Thapsigargin (Sigma). Two concentrations were tested – 1 µM and 0.1 µM. Cells from all conditions were harvested at the same time. The total RNA was isolated RNAEasy Mini Kit (Qiagen) according to the manufacturers protocol. The RNA was reversely transcribed using QuantiTect Reverse Transcription Kit (Qiagen) following the manufacturers protocol using 1 µg of RNA as starting material. Obtained cDNA was used as a template for PCR reaction. Reaction mix was prepared as follow: 1 µL of cDNA template, 1.25 µL of 10 µM forward primer, 1.25 µL of 10 µM reverse primer, 6.25 µL of 2x Phusion High-Fidelity PCR Master Mix with GC buffer (Thermo Fisher) and 3.25 µL of $H_2O$. Primers used for the PCR are listed in the Table 5. The following PCR program was used: 30 s of initial denaturation at 98°C, 30 cycles of denaturation (98°C, 10

s), annealing (58°C, 20 s) and extension (72°C, 15 s), final extension (72°C, 1 min). PCR products were separated on 2.5% agarose gel.

**Table 5. Primers used for the XBP1 splicing assay**

| Primer | Sequence |
|---|---|
| XBP1_splicing_assay_forward | TTACGAGAGAAAACTCATGGCC |
| XBP1_splicing_assay_reverse | GGGTCCAAGTTGTCCAGAATGC |

### 4.2.11. Design of gRNA sequences for the CRISPR-Cas9 genome editing experiment

The DESKGEN Cloud (Desktop Genetics) was used to design a target sequence for the CRISPR-Cas9 genome editing experiment. The following target sequence was selected based on the efficiency and specificity scores: ACCGCGGATGAGGAATGTGA (exon 22). For each target primers allowing for the assembly of the gRNA DNA template were designed as shown in the Table 6:

**Table 6. Primers used to generate DNA template for gRNA in-vitro transcription**

| Primer | Sequence |
|---|---|
| Target_gRNA_Forward | TAATACGACTCACTATAGACCGCGGATGAGGAATGTGA |
| Target_gRNA:Reverse | TTCTAGCTCTAAAACTCACATTCCTCATCCGCGGT |

### 4.2.12. Synthesis of gRNA

Synthesis and purification of gRNA was performed using the GeneArt™ Precision gRNA Synthesis Kit (Thermo Fisher Scientific). First, the gRNA template was assembled by PCR reaction. The reaction mix was prepared by mixing 1 µL of 0.3 uM target oligos, 1 µL of Tracr Fragment containing T7 Primer Mix and 12.5 µL Phusion High-Fidelity PCR Master mix (2X). Nuclease-free water was added to 25 µL. The following two-step PCR program was used for the reaction: 10 seconds of initial denaturation at 98°C, 32 cycles of denaturation (98°C, 5 seconds) and annealing and extension (55°C, 15 seconds), final extension (72°C, 1 min). 6 µL of PCR reaction product was mixed with 8 µL of 100 mM NTP mix, 4 µL of 5X TranscriptAid™ Reaction buffer and 2 µL of TranscriptAid™ Enzyme mix. *In vitro* transcription was allowed to proceed for 2 h at 37°C. After this time 1 µL of DNase I was added into the reaction and incubated for 15 min at 37°C. *In vitro* transcribed gRNA was purified using GeneJET RNA Purification Micro Column (Thermo Fisher Scientific).

### 4.2.13. Transfection of Hek293 cell line with gRNA-Cas9 complex

First $2x10^6$ HEK Flp-In™ T-Rex™ 293 cells were seeded in a 10cm tissue culture dish. Cells were allowed to attach for 24 h hours. To assemble the Cas9-gRNA RNP complex, 500 µL of Opti-MEM (Invitrogen) was mixed with 15 µg of TrueCut Cas9 Protein v2 (Invitrogen) and 3.5 µg of previously purified gRNA. In a second tube 500 µL of Opti-MEM (Invitrogen) was mixed with 43.5 µL of Lipofectamine RNAiMAX transfection reagent (Invitrogen). Reaction tubes were incubated for 5 minutes at RT. Then assembled RNP complexes and diluted transfection reagent were mixed by gentle pipetting. Reaction was incubated for 10 min at RT. 1 mL of Cas9/gRNA/transfection reagent mix was added dropwise to the cell culture dish. To allow genome editing cells were incubated for 48h at 37°C.

### 4.2.14. Single cell sorting

48h after transfection cells were detached using 1 mL of TrypLE™ Express Enzyme (Gibco) without phenol red. Cells were incubated for 1 min at 37°C and the reaction was terminated by diluting the enzyme in 10 mL of PBS. Cells were washed twice with PBS and resuspended in sorting buffer (0.5 mM EDTA, 0.1% FBS in PBS). Single cell sorting was performed in the Flow Cytometry Core Facility at EMBL using the BD FACSMelody cell sorter (BD Biosciences). In total 960 single cells were isolated (each into a separate well on 96-well plates) and expanded for the following 2 weeks.

### 4.2.15. Screening for the successful mutation

After clonal expansion cells were trypsinized and transferred to new 96-well plates. To isolate genomic DNA 10% of cells from each clone were transferred to QuickExtractTM buffer (Lucigen) and vortexed for 15 s. Reactions were incubated for 6 min at 65°C, mixed by vortexing for 15 s and incubated 2 min at 98°C. The 435bp long fragment of gp210 gene containing the target sequence was amplified by PCR using purified DNA as template. Reaction mix was prepared as follow: 1 µL of genomic DNA template, 1.25 µL of 10 µM forward primer, 1.25 µL of 10 µM reverse primer, 12.5 µL of 2x Phusion High-Fidelity PCR Master Mix with GC buffer (Thermo Fisher) and 9 µL of $H_2O$. The following PCR program was used: 30 seconds of initial denaturation at 98°C, 30 cycles of denaturation (98°C, 10 s), annealing (68°C, 20 s) and extension (72°C, 15 s), final extension (72°C, 1 min). PCR products were purified using

QIAquick PCR Purification Kit (Qiagen) and analyzed by sequencing. Primers used for the PCR reaction are listed in the Table 7.

**Table 7. Primers used for the CRISPR mutation screening**

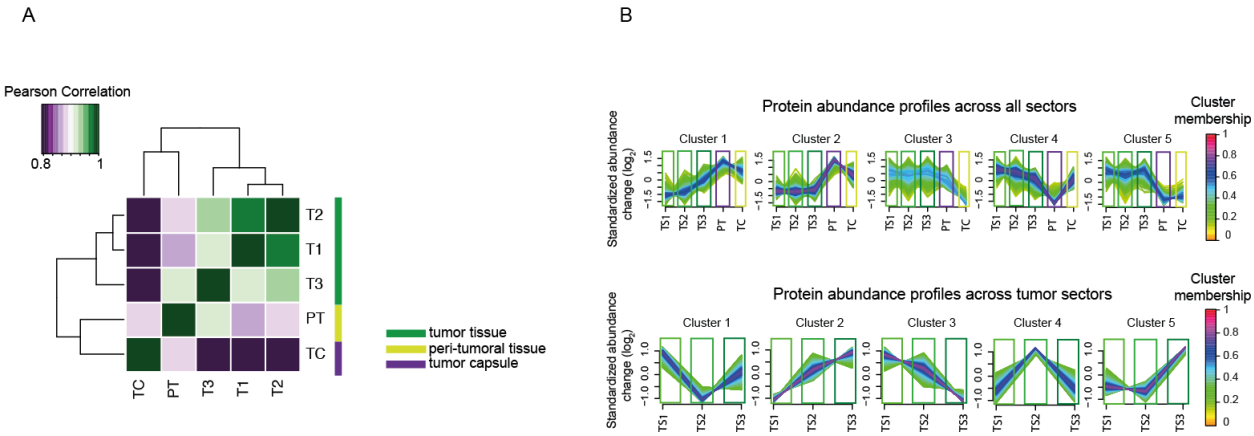
| Primer | Sequence |
|---|---|
| CRISPR_screen_forward | TACTTCCCCTTTATGGACCTGAAGC |
| CRISPR_screen_reverse | ATGAGAGGACATTGTGGGACACAG |

### 4.2.16. Proteomic characterization of the gp210 knockout cell line

For each replicate $1x10^6$ cells in 100 µL PBS were collected. Solubilization, digestion and purification were performed according to the protocol described in section 4.1.13. Purified peptides were labeled with 6-plex TMT reagent according to the protocol described in section 4.1.6. with the following exceptions: (i) 10 µg of peptides were first dried and then reconstituted in 10 µL of 100 mM Hepes (pH 8.5), (ii) after quenching peptides were pooled together and purified using C18 macro spin column (for details see section 4.1.13). Data were acquired as described before in the section 4.1.8.

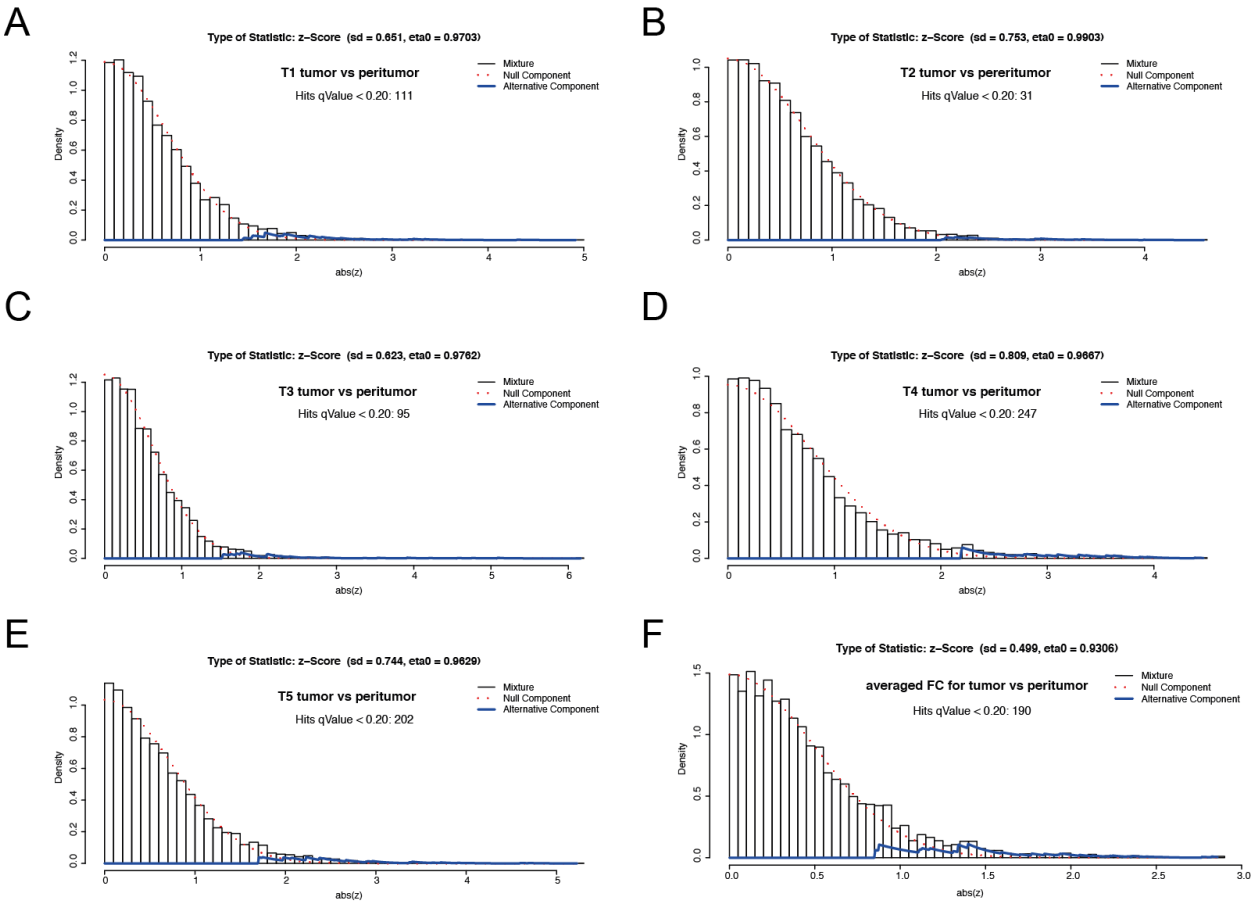
# 5. Appendices

## 5.1. Supplementary figures



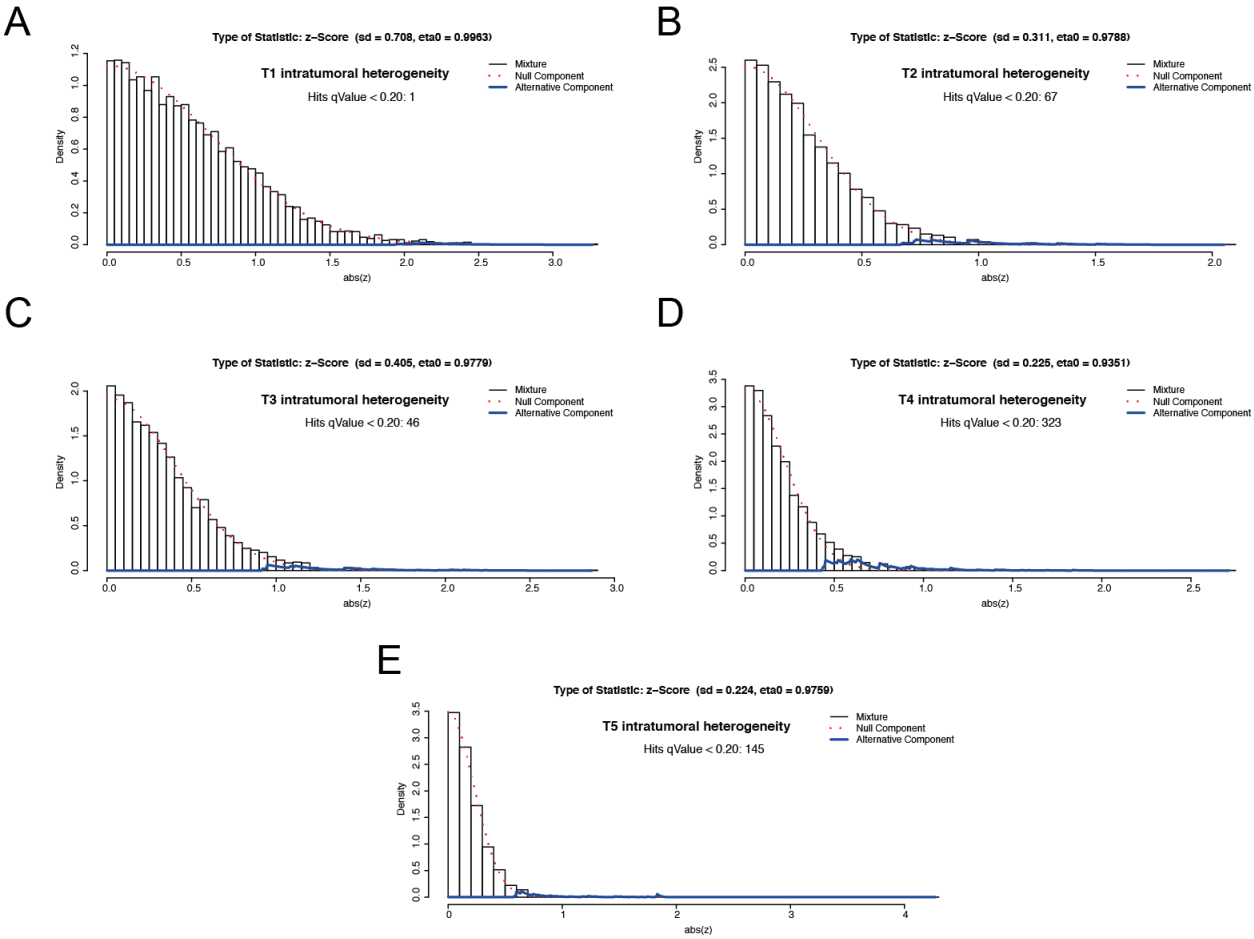

**Supplementary Figure 1. Proteome profiles of HCC (TMT dataset)**

(A) Heatmap representing the Pearson correlation coefficient of the analyzed sectors. (B) Soft clustering analysis of HCC spatial proteome using the fuzzy c-means algorithm (Kumar and E Futschik, 2007). The optimal number of clusters was estimated using the "elbow" algorithm (Schwämmle and Jensen, 2010). The upper panel includes all measured sectors (including tumor capsule and peri-tumoral tissue). The lower panel shows clusters calculated only for the three tumor sectors. The TMT dataset was used to create this figure. The DIA data are shown in the Figure 13. Modified from (Buczak et al., 2018)




**Supplementary Figure 2. Statistical models for tumor vs. peritumor comparison**

Histograms and the densities of the fitted two-component models (null component = proteins with no differential expression; alternative component = proteins with differential expression) are shown. Models were fitted on median centered $\log_2$- transformed fold-changes (z). Proteins with the q-values < 0.2 were considered as differentially expressed.. (A) T1 tumor vs peritumor, (B) T2 tumor vs peritumor, (C) T3 tumor vs peritumor, (D) T4 tumor vs peritumor, (E) T5 tumor vs peritumor. (F) averaged fold changes (averages were calculated without T1). Modified from (Buczak et al., 2018)




**Supplementary Figure 3. Statistical models for periphery vs. center comparisons**

Histograms and the densities of the fitted two-component models (null component = proteins with no differential expression; alternative component = proteins with differential expression) are shown. Models were fitted on median centered $\log_2$- transformed fold-changes (z). Proteins with the q-values < 0.2 were considered as differentially expressed. (A) T1 intratumoral heterogeneity, (B) T2 intratumoral heterogeneity, (C) T3 intratumoral heterogeneity, (D) T4 intratumoral heterogeneity, (E) T5 intratumoral heterogeneity Modified from (Buczak et al., 2018)

## 5.2. Supplementary tables

### Supplementary Table 1. Targeted NGS data

Overview of identified variants in each sector. Provided by dr. Volker Endris. Genes in which mutation were found are colored in red. Gene highlighted in yellow indicate the only non homogenous mutation.

| | Allelic frequency (%) | | | |
|---|---|---|---|---|
| Aminoacid change | peritumor | Sector1 | Sector2 | Sector3 |
| ABL2:p.Lys909Arg(NM_001168236) | 47.91 | 43.01 | 50.87 | 42.69 |
| ADGRA2:p.Glu111Lys(NM_032777) | | | | 12.99 |
| ADGRA2:p.Pro76fs*16(NM_032777) | 78.00 | 79.60 | 79.80 | 79.20 |
| ADGRB3:p.Asn503Ser(NM_001704) | 100.00 | 100.00 | 100.00 | 98.07 |
| AFF3:p.Asn519Ser(NM_001025108) | 51.87 | 48.48 | 53.71 | 48.02 |
| AFF3:p.Ser383Asn(NM_001025108) | 100.00 | 100.00 | 99.38 | 100.00 |
| AKAP9:p.Asn2792Ser(NM_005751) | 56.00 | 66.36 | 30.00 | 54.00 |
| ALK:p.Asp1529Glu(NM_004304) | 97.60 | 98.43 | 98.36 | 98.65 |
| ALK:p.Ile1461Val(NM_004304) | 100.00 | 100.00 | 100.00 | 100.00 |
| ALK:p.Lys1491Arg(NM_004304) | 98.68 | 98.86 | 100.00 | 98.40 |
| ARID1A:p.Gln321Glu(NM_006015) | 35.00 | 51.88 | 45.00 | 38.00 |
| ASXL1:p.Leu815Pro(NM_015338) | 98.95 | 99.16 | 100.00 | 100.00 |
| ATM:p.Asn1983Ser(NM_000051) | 98.96 | 98.18 | 96.30 | 94.87 |
| ATR:p.Met211Thr(NM_001184) | 92.99 | 98.52 | 100.00 | 98.48 |
| AURKA:p.Ile57Val(NM_003600) | 100.00 | 100.00 | 100.00 | 100.00 |
| AURKA:p.Phe31Ile(NM_003600) | 47.33 | 55.29 | 52.61 | 51.29 |
| AURKB:p.Met257Thr(NM_001256834) | 100.00 | 100.00 | 100.00 | 100.00 |
| AXL:p.Asn266Asp(NM_001699) | 98.75 | 99.06 | 98.86 | 100.00 |
| BCL2L2:p.Gln133Arg(NM_001199839) | 100.00 | 100.00 | 100.00 | 100.00 |
| BCL9:p.Pro671Ser(NM_004326) | 74.11 | 72.47 | 84.80 | 84.76 |
| BCR:p.Asn796Ser(NM_004327) | 100.00 | 100.00 | 100.00 | 100.00 |
| BIRC5:p.Glu129Lys(NM_001168) | 100.00 | 98.86 | 99.05 | 100.00 |
| BUB1B:p.Arg349Gln(NM_001211) | 42.56 | 44.92 | 46.39 | 44.19 |
| CASC5:p.Ala460Ser(NM_144508) | 94.73 | 98.80 | 98.50 | 98.23 |
| CASC5:p.Arg43Thr(NM_144508) | 88.13 | 100.00 | 98.75 | 95.41 |
| CASC5:p.Arg910Gly(NM_144508) | 95.43 | 100.00 | 100.00 | 98.16 |
| CASC5:p.Lys1259Glu(NM_144508) | 46.45 | 47.20 | 47.62 | 43.74 |
| CASC5:p.Met572Thr(NM_144508) | 39.67 | 50.00 | 57.00 | 56.00 |
| CDH11:p.Ser373Ala(NM_001797) | 50.64 | 50.51 | 52.75 | 51.34 |
| CDH11:p.Thr255Met(NM_001797) | 42.56 | 51.60 | 46.97 | 52.14 |
| CDH2:p.Ala118Thr(NM_001792) | 48.16 | 51.33 | 52.14 | 50.21 |
| CDH5:p.Ile503Thr(NM_001795) | 47.25 | 47.39 | 44.03 | 49.14 |
| **CDK12:p.His369Arg(NM_015083)** | | **24.89** | **6.79** | |
| CHEK1:p.Ile437Val(NM_001244846) | 98.95 | 100.00 | 100.00 | 100.00 |
| CMPK1:p.Gly8Arg(NM_001136140) | 60.00 | 57.00 | 64.00 | 64.00 |
| COL1A1:p.Thr1075Ala(NM_000088) | 100.00 | 100.00 | 100.00 | 100.00 |
| CREBBP:p.Met2183fs*120(NM_001079846) | 14.41 | 9.55 | 10.04 | 12.54 |
| CSMD3:p.Ile219Met(NM_052900) | 48.06 | 57.59 | 48.36 | 48.04 |
| CTNNB1:p.Asn287Ser(NM_001098209) | 50.59 | 44.00 | 50.00 | 48.41 |
| CYP2D6:p.Cys245Arg(NM_001025161) | 100.00 | 100.00 | 98.66 | 100.00 |
| CYP2D6:p.Pro34Ser(NM_000106) | 10.71 | 2.00 | 2.00 | 6.63 |

| | | | | |
|---|---|---|---|---|
| CYP2D6:p.Thr435Ser(NM_001025161) | 98.79 | 100.00 | 100.00 | 100.00 |
| DCC:p.Arg201Gly(NM_005215) | 52.30 | 53.65 | 52.35 | 49.70 |
| DCC:p.Gly1031Arg(NM_005215) | 7.09 | 7.79 | 11.11 | 3.19 |
| DCC:p.Phe23Leu(NM_005215) | 97.99 | 100.00 | 98.76 | 100.00 |
| DICER1:p.1420_1421del: (NM_001195573) | 59.25 | 59.03 | 53.80 | 52.66 |
| **DNMT3A:p.Tyr247Phe(NM_153759)** | | **29.53** | **31.45** | **26.83** |
| DPYD:p.Arg29Cys(NM_000110) | 99.06 | 100.00 | 100.00 | 98.53 |
| DST:p.Gln1308Arg(NM_015548) | 52.64 | 50.28 | 52.32 | 48.82 |
| DST:p.Lys4066Arg(NM_015548) | 3.00 | 3.93 | 5.15 | 3.46 |
| DST:p.Met2813Ile(NM_015548) | 100.00 | 100.00 | 100.00 | 100.00 |
| DST:p.Thr2726Ala(NM_015548) | 97.64 | 99.00 | 99.32 | 100.00 |
| EGFR:p.Arg521Lys(NM_005228) | 50.59 | 49.61 | 50.42 | 45.54 |
| EML4:p.Ile324Val(NM_001145076) | 99.17 | 100.00 | 100.00 | 100.00 |
| EML4:p.Lys225Glu(NM_001145076) | 98.85 | 98.33 | 100.00 | 100.00 |
| ERBB2:p.Ile655Val(NM_001289937) | 98.43 | 98.57 | 100.00 | 100.00 |
| ERBB2:p.Pro1170Ala(NM_004448) | 52.53 | 48.20 | 57.69 | 50.44 |
| ERBB3:p.Ser1119Cys(NM_001982) | 47.62 | 50.56 | 51.30 | 51.89 |
| ERBB3:p.Thr1024Asn(NM_001982) | 6.75 | 11.38 | 12.86 | 3.49 |
| ERCC2:p.Lys751Gln(NM_000400) | 44.79 | 52.10 | 55.41 | 49.57 |
| ERCC5:p.Gly1053Arg(NM_000123) | 100.00 | 100.00 | 100.00 | 100.00 |
| ERCC5:p.Gly1080Arg(NM_000123) | 100.00 | 94.69 | 100.00 | 100.00 |
| ESR1:p.Gln238fs*12(NM_000125) | 93.86 | 94.63 | 93.67 | 94.84 |
| ESR1:p.Gln238Leu(NM_000125) | 93.86 | 94.63 | 93.67 | 94.84 |
| FANCA:p.Gly501Ser(NM_000135) | 43.49 | 51.44 | 53.67 | 48.47 |
| FANCA:p.Gly809Asp(NM_000135) | 48.98 | 51.55 | 52.22 | 52.31 |
| FANCA:p.Leu1143Val(NM_000135) | 43.99 | 45.13 | 47.51 | 45.59 |
| FANCA:p.Ser1088Phe(NM_000135) | 7.24 | 0.10 | 1.00 | 2.00 |
| FANCA:p.Thr266Ala(NM_000135) | 55.75 | 48.40 | 48.34 | 54.75 |
| FGFR3:p.Pro473Thr(NM_022965) | 4.48 | 4.01 | 8.44 | 14.42 |
| FGFR3:p.Val570Ile(NM_022965) | 93.86 | 90.57 | 90.46 | 88.12 |
| FGFR4:p.Gly388Arg(NM_002011) | 95.95 | 97.54 | 99.83 | 100.00 |
| FGFR4:p.Pro136Leu(NM_022963) | 97.63 | 100.00 | 100.00 | 98.80 |
| FLT3:p.Thr227Met(NM_004119) | 78.00 | 100.00 | 90.00 | 100.00 |
| FLT4:p.His890Gln(NM_002020) | 96.62 | 100.00 | 100.00 | 99.34 |
| FN1:p.Gln15Leu(NM_002026) | 55.65 | 45.39 | 54.02 | 48.57 |
| FN1:p.Thr817Pro(NM_002026) | 100.00 | 100.00 | 100.00 | 100.00 |
| FN1:p.Val1960Ile(NM_212474) | 100.00 | 100.00 | 100.00 | 100.00 |
| GATA2:p.His323Tyr(NM_001145662) | 65.38 | 62.83 | 39.06 | 34.62 |
| GNAS:p.Gly35fs*654(NM_080425) | 100.00 | 100.00 | 100.00 | 100.00 |
| GNAS:p.Ser565fs*63(NM_001077490) | 18.10 | 18.60 | 18.50 | 18.70 |
| HNF1A:p.Ser574Gly(NM_000545) | 100.00 | 100.00 | 100.00 | 98.68 |
| HSP90AA1:p.Met71Leu(NM_001017963) | 100.00 | 100.00 | 100.00 | 100.00 |
| IGF1R:p.Gly596Val(NM_000875) | 5.14 | 5.14 | 4.69 | 3.20 |
| IGF2R:p.Arg1619Gly(NM_000876) | 100.00 | 100.00 | 100.00 | 100.00 |
| IL6ST:p.Gly148Arg(NM_001190981) | 44.84 | 48.18 | 51.32 | 48.37 |
| IL6ST:p.Leu397Val(NM_001190981) | 46.69 | 49.78 | 68.18 | 43.48 |
| IL7R:p.Ile356Val(NM_002185) | 93.64 | 98.87 | 100.00 | 98.49 |
| IL7R:p.Ile66Thr(NM_002185) | 100.00 | 100.00 | 100.00 | 100.00 |
| IL7R:p.Val138Ile(NM_002185) | 95.50 | 100.00 | 95.04 | 100.00 |
| ITGA9:p.Gly507Glu(NM_002207) | 92.99 | 100.00 | 100.00 | 97.68 |

| | | | | |
|---|---|---|---|---|
| ITGB2:p.Gln354His(NM_000211) | 100.00 | 100.00 | 100.00 | 100.00 |
| KAT6B:p.Glu787_790del(NM_001256469) | 9.18 | 9.52 | 23.00 | 9.24 |
| KDM6A:p.Thr647Lys(NM_001291418) | 95.50 | 100.00 | 100.00 | 89.63 |
| KDR:p.Gln472His(NM_002253) | 51.95 | 62.96 | 50.00 | 57.06 |
| KMT2C:p.Asp348Asn(NM_170606) | 15.50 | 14.99 | 15.77 | 15.72 |
| KMT2C:p.Leu291Phe(NM_170606) | 39.00 | 51.00 | 38.00 | 40.00 |
| KMT2C:p.Tyr816_Ile817delins*(NM_170606) | 39.79 | 45.04 | 40.02 | 40.08 |
| LIFR:p.Val785Ile(NM_001127671) | 41.82 | 47.76 | 45.00 | 42.39 |
| LRP1B:p.Arg3783Gln(NM_018557) | 50.00 | 45.81 | 46.15 | 51.85 |
| LRP1B:p.Gln48Arg(NM_018557) | 45.83 | 45.16 | 48.73 | 45.41 |
| LRP1B:p.Pro1896Ser(NM_018557) | 50.18 | 43.81 | 45.45 | 45.83 |
| LTF:p.Ala29Thr(NM_002343) | 13.42 | 11.88 | 9.68 | 9.08 |
| LTF:p.Arg23delinsArgArg(NM_002343) | 100.00 | 100.00 | 100.00 | 100.00 |
| LTF:p.Lys3Arg(NM_001199149) | 50.24 | 42.39 | 50.85 | 52.94 |
| LTK:p.Pro62Leu(NM_001135685) | 48.22 | 46.51 | 48.07 | 42.33 |
| MALT1:p.Arg217Gly(NM_006785) | 47.30 | 52.20 | 51.23 | 51.83 |
| MEN1:p.Thr546Ala(NM_000244) | 100.00 | 100.00 | 100.00 | 100.00 |
| MLH1:(NM_001167619) | 49.90 | 53.45 | 56.25 | 52.89 |
| MLH1:p.Ile219Val(NM_000249) | 63.25 | 62.83 | 59.51 | 64.63 |
| MTOR:p.Pro201fs*5(NM_004958) | 14.40 | 12.00 | 12.70 | 12.60 |
| MTR:p.Arg52Gln(NM_000254) | 52.50 | 53.36 | 54.55 | 48.85 |
| MTRR:p.Ile22Met(NM_002454) | 46.20 | 49.41 | 49.66 | 49.50 |
| MTRR:p.Ser175Leu(NM_002454) | 44.06 | 35.76 | 54.70 | 48.25 |
| MUC1:p.Ser227Thr(NM_001204285) | 47.39 | 40.08 | 46.94 | 39.30 |
| MUTYH:p.Gln324His(NM_001048171) | 50.00 | 54.00 | 51.00 | 42.00 |
| MUTYH:p.Ser501Phe(NM_001048171) | 55.00 | 52.00 | 54.00 | 50.00 |
| MYCL:p.Thr362Ser(NM_001033081) | 50.86 | 55.78 | 68.27 | 55.39 |
| **MYH11:p.Leu1563Pro(NM_002474)** | **3.00** | **32.81** | **31.58** | **27.23** |
| NCOA4:p.Phe8Val(NM_001145260) | 43.50 | 49.80 | 45.58 | 37.65 |
| NIN:p.Gln1125Pro(NM_020921) | 98.20 | 100.00 | 92.53 | 100.00 |
| NIN:p.Gly1320Glu(NM_020921) | 96.80 | 98.22 | 98.60 | 96.73 |
| NIN:p.Pro1111Ala(NM_020921) | 46.55 | 51.98 | 49.91 | 51.31 |
| NLRP1:p.Met1154Val(NM_033006) | 10.23 | 2.00 | 5.00 | |
| NOTCH4:(NM_004557) | 41.92 | 49.95 | 40.06 | 49.11 |
| NOTCH4:p.Gly348fs*49(NM_004557) | 14.80 | 16.20 | 20.40 | 22.80 |
| NOTCH4:p.Lys117Gln(NM_004557) | 47.04 | 53.52 | 55.85 | 50.60 |
| NOTCH4:p.Thr320Ala(NM_004557) | 37.64 | 44.33 | 47.58 | 41.31 |
| NUP214:p.Pro574Ser(NM_005085) | 48.22 | 50.04 | 44.70 | 51.21 |
| NUP98:p.Asn297Asp(NM_005387) | 3.99 | 4.33 | 4.03 | 3.74 |
| NUP98:p.Gln1142Glu(NM_016320) | 43.00 | 50.00 | 51.00 | 44.77 |
| PARP1:p.Gln353fs(NM_001618) | 62.19 | 61.05 | 62.80 | 61.85 |
| PARP1:p.Val762Ala(NM_001618) | 38.36 | 51.14 | 40.28 | 37.18 |
| PAX5:p.Thr264Ile(NM_001280550) | 96.28 | 98.61 | 100.00 | 97.09 |
| PBX1:p.Gly21Ser(NM_001204961) | 40.00 | 43.11 | 39.22 | 40.31 |
| PDE4DIP:p.Ala1066Thr(NM_001198834) | 50.47 | 45.54 | 50.65 | 49.63 |
| PDE4DIP:p.Ala106Val(NM_001002811) | 49.86 | 49.90 | 49.62 | 52.24 |
| PDE4DIP:p.Ala127Thr(NM_001002811) | 53.77 | 52.59 | 53.13 | 54.48 |
| PDE4DIP:p.Ala1742Ser(NM_001198834) | 11.13 | 10.64 | 5.51 | 7.69 |
| PDE4DIP:p.Arg1867Cys(NM_001198834) | 22.85 | 27.10 | 18.11 | 29.03 |
| PDE4DIP:p.Arg25Leu(NM_001002810) | 47.55 | 46.68 | 49.35 | 46.35 |

| | | | | |
|---|---|---|---|---|
| PDE4DIP:p.Arg458His(NM_001002811) | 19.90 | 23.55 | 25.52 | 25.89 |
| PDE4DIP:p.Arg844His(NM_001002811) | 46.30 | 49.53 | 43.36 | 45.08 |
| PDE4DIP:p.Asp1910Glu(NM_001198834) | 45.23 | 48.00 | 41.81 | 45.17 |
| PDE4DIP:p.Glu573Val(NM_001002811) | 49.05 | 48.88 | 47.38 | 48.42 |
| PDE4DIP:p.His1598Arg(NM_001198834) | 24.93 | 25.32 | 23.35 | 29.71 |
| PDE4DIP:p.Leu1727Pro(NM_001198834) | 13.63 | 14.55 | 11.91 | 11.46 |
| PDE4DIP:p.Leu71Ile(NM_001198832) | 12.71 | 13.06 | 11.85 | 12.09 |
| PDE4DIP:p.Lys1359Glu(NM_001198834) | 100.00 | 98.08 | 100.00 | 98.48 |
| PDE4DIP:p.Lys257Glu(NM_001002811) | 47.30 | 51.45 | 52.65 | 55.35 |
| PDE4DIP:p.Phe1013Ile(NM_001198834) | 45.22 | 42.67 | 42.66 | 48.16 |
| PDE4DIP:p.Ser1097Gly(NM_001198834) | 24.29 | 23.49 | 26.70 | 27.77 |
| PDE4DIP:p.Ser438Leu(NM_001002811) | 49.15 | 56.95 | 43.65 | 45.56 |
| PDE4DIP:p.Ser699Thr(NM_001002811) | 49.37 | 41.63 | 38.84 | 50.45 |
| PDE4DIP:p.Thr2297Ala(NM_001198834) | 25.56 | 22.79 | 25.23 | 25.17 |
| PDE4DIP:p.Thr6Ala(NM_001198832) | 9.44 | 9.71 | 7.49 | 9.19 |
| PDE4DIP:p.Trp1396Arg(NM_001198834) | 81.63 | 77.59 | 75.47 | 76.96 |
| PDE4DIP:p.Trp60*(NM_022359) | 57.06 | 58.27 | 56.22 | 55.50 |
| PDE4DIP:p.Trp723*(NM_001002811) | 25.40 | 20.71 | 16.06 | 26.29 |
| PDE4DIP:p.Val1736Glu(NM_001198834) | 8.35 | 8.72 | 4.85 | 6.27 |
| PER1:p.Ala962Pro(NM_002616) | 100.00 | 99.20 | 100.00 | 99.16 |
| PIK3CA:p.Ile391Met(NM_006218) | 64.15 | 48.57 | 60.00 | 69.00 |
| PIK3R1:p.Met56Ile(NM_181504) | 41.67 | 43.02 | 43.23 | 37.73 |
| PIK3R2:p.Ser313Pro(NM_005027) | 100.00 | 100.00 | 100.00 | 100.00 |
| PKHD1:p.Ala1262Val(NM_138694) | 52.23 | 49.64 | 54.99 | 51.34 |
| PKHD1:p.Arg760Cys(NM_138694) | 58.80 | 48.49 | 50.77 | 52.27 |
| PKHD1:p.Asn830Ser(NM_138694) | 53.15 | 47.54 | 39.68 | 50.31 |
| PKHD1:p.Gln3899Arg(NM_138694) | 40.08 | 54.37 | | 44.12 |
| PKHD1:p.Leu1870Val(NM_138694) | 98.88 | 98.72 | 98.46 | 98.16 |
| PML:p.Gly732Val(NM_033250) | 44.64 | 49.12 | 48.68 | 43.68 |
| PML:p.Phe645Leu(NM_033238) | 44.20 | 45.45 | 56.21 | 55.46 |
| PML:p.Ser724Gly(NM_033250) | 41.73 | 47.57 | 48.00 | 41.39 |
| PMS2:p.Lys541Glu(NM_000535) | 96.31 | 96.49 | 95.45 | 95.55 |
| PMS2:p.Pro470Ser(NM_000535) | 52.84 | 48.97 | 49.52 | 50.74 |
| RECQL4:p.Ser92Pro(NM_004260) | 100.00 | 100.00 | 100.00 | 100.00 |
| RET:p.Gly691Ser(NM_020630) | 7.49 | 1.00 | | 2.00 |
| RNASEL:p.Arg462Gln(NM_021133) | 57.66 | 51.74 | 55.19 | 51.98 |
| RNASEL:p.Asp541Glu(NM_021133) | 53.88 | 58.33 | 48.92 | 55.87 |
| RNF213:p.Asn1045Asp(NM_001256071) | 100.00 | 100.00 | 100.00 | 100.00 |
| RNF213:p.Asp1331Gly(NM_001256071) | 52.26 | 50.00 | 49.29 | 56.71 |
| RNF213:p.Gln1133Lys(NM_001256071) | 49.00 | 44.96 | 50.77 | 45.59 |
| RNF213:p.Glu1272Gln(NM_001256071) | 50.15 | 44.30 | 54.73 | 46.09 |
| RNF213:p.Gly1828Ala(NM_001256071) | 9.06 | 11.34 | 4.39 | 5.41 |
| RNF213:p.Lys1034Met(NM_020954) | 8.24 | | | 2.00 |
| RNF213:p.Met270Thr(NM_001256071) | 29.74 | 32.61 | 33.74 | 35.33 |
| RNF213:p.Ser2334Asn(NM_001256071) | 51.91 | 52.33 | 54.28 | 48.66 |
| RNF213:p.Val1195Met(NM_001256071) | 52.04 | 49.68 | 51.51 | 49.91 |
| RPS6KA2:p.Glu32Gly(NM_001006932) | 100.00 | 100.00 | 100.00 | 100.00 |
| RPS6KA2:p.Thr34Ala(NM_001006932) | 99.95 | 99.95 | 100.00 | 100.00 |
| SEPT9:p.Met464Val(NM_001113495) | 99.95 | 99.95 | 100.00 | 100.00 |
| SEPT9:p.Pro138Leu(NM_001113493) | 45.17 | 46.15 | 52.41 | 44.79 |

| | | | | |
|---|---|---|---|---|
| SETD2:p.Pro1962Leu(NM_014159) | 53.12 | 41.01 | 50.11 | 51.42 |
| SMARCA4:p.Pro169Ser(NM_001128845) | 41.89 | 49.18 | 45.52 | 36.90 |
| SMO:p.Pro698Arg(NM_005631) | 49.18 | 53.21 | 50.41 | 48.00 |
| SYNE1:p.Ala2802Val(NM_033071) | 52.98 | 50.99 | 47.16 | 50.87 |
| SYNE1:p.Phe7231Val(NM_033071) | 92.46 | 95.74 | 94.78 | 95.37 |
| TBX22:p.Glu67Lys(NM_001303475) | 92.77 | 97.49 | 95.19 | 97.52 |
| TET1:p.Asp162Gly(NM_030625) | 49.91 | 43.41 | 53.99 | 51.29 |
| TET1:p.Ile1123Met(NM_030625) | 100.00 | 98.22 | | 83.00 |
| TET2:p.Gln729Glu(NM_001127208) | 6.90 | 3.64 | 11.59 | 11.24 |
| TET2:p.His1778Arg(NM_001127208) | 49.52 | 56.95 | 54.32 | 56.36 |
| TET2:p.Ile1762Val(NM_001127208) | 49.23 | 42.29 | 44.89 | 42.85 |
| TET2:p.Leu34Phe(NM_001127208) | 45.75 | 47.19 | 38.27 | 45.56 |
| TET2:p.Val218Met(NM_001127208) | 43.85 | 40.87 | 40.73 | 38.40 |
| TGM7:p.Arg552Gln(NM_052955) | 60.49 | 44.44 | 52.00 | 53.93 |
| TGM7:p.Pro564Leu(NM_052955) | 48.26 | 41.69 | 54.94 | 55.35 |
| THBS1:p.Gln882His(NM_003246) | 22.57 | 22.12 | 3.91 | 8.30 |
| THBS1:p.Thr523Ala(NM_003246) | 39.95 | 47.35 | 48.33 | 45.03 |
| TLR4:p.Asp99Gly(NM_138557) | 51.34 | 50.66 | 46.93 | 44.35 |
| TLR4:p.Thr199Ile(NM_138557) | 42.35 | 49.06 | 48.40 | 42.57 |
| TNK2:p.Pro725Leu(NM_005781) | 48.58 | 58.06 | 61.56 | 58.24 |
| TP53:p.Pro33Arg(NM_001126118) | 95.45 | 84.81 | 85.89 | 97.18 |
| TPR:p.Ser960Asn(NM_003292) | 30.95 | 45.65 | 53.68 | 35.11 |
| TRIM33:p.Ile840Thr(NM_015906) | 96.60 | 100.00 | 100.00 | 96.84 |
| TRIP11:p.Gly1827Ser(NM_004239) | 11.79 | 11.24 | 11.37 | 9.95 |
| TSHR:p.Arg248Ser(NM_001018036) | 100.00 | 100.00 | 98.84 | 100.00 |
| TSHR:p.Glu727Asp(NM_000369) | 100.00 | 100.00 | 100.00 | 100.00 |
| WRN:p.Leu1074Phe(NM_000553) | 98.30 | 98.27 | 100.00 | 97.76 |
| XPC:p.Gln939Lys(NM_004628) | 48.21 | 51.00 | 53.13 | 45.44 |

## Supplementary Table 2. Patients characteristics

Provided by dr. Stephan Singer

| Tumor | Etiology | Grading | T-stage | Sex | Age |
|---|---|---|---|---|---|
| T0 (first analysis) | Steatohepatitis and chemicals | G2 | pT2 | male | 53 |
| T1 | HCV | G3 | ypT2 | male | 52 |
| T2 | HBV | G2 | pT2 | male | 51 |
| T3 | Steatohepatitis | G2 | pT1 | male | 76 |
| T4 | Steatohepatitis | G1 | pT3 | male | 81 |
| T5 | Unknown | G1 | pT2 | female | 40 |

y = pre-operative treatment

## 5.3. List of figures

## 5.4. List of tables



## 5.5. List of abbreviations

| | |
|---|---|
| ACN | acetonitrile |
| AML | acute myeloid leykemia |
| BPC | base peak chromatogram |
| CID | collision induced dissociation |
| CRM | charge reside model |
| CSC | cancer stem cell |
| ctDNA | circulating DNA |
| DC | direct current |
| DDA | data dependent acquisition |
| DIA | data independent acquisition |
| DMEM | Dulbecco's modified eagle medium |
| ER | endolasmic reticulum |
| ERAD | ER associated degradation |
| ESI | electrospray ionization |
| ETD | electron transfer dissociation |
| FA | formic acid |
| FASP | filter-aided sample preparation |
| FBS | fetal bovine serum |
| FFPE | formalin-fixed and paraffin-embedded |
| FG | phenylalanine and glycine repeats |
| GO | gene ontology |
| GPI | glycophosphatidylinositol |
| H&E | hematoxylin and eosin |
| HBV | hepatitis B virus |
| HCC | hepatocellular carcinoma |
| HCD | higher-energy C-trap dissociation |
| HCV | hepatitis C virus |
| HIAR | heat-induced antigen retrieval |
| IEM | ion evaporation model |

| | |
|---|---|
| IHC | immunohistochemistry |
| INM | inner nuclear membrane |
| ITH | intratumoral heterogeneity |
| KO | knockout |
| LC | liquid chromatography |
| LCM | laser capture microdissection |
| LFQ | label free quantification |
| m/z | mass-to-charge ratio |
| MCM | minichromosome maintenance |
| MS | mass spectrometry |
| mtDNA | mitochondrial DNA |
| NE | nuclear envelope |
| NGS | next generation sequencing |
| NPC | nuclear pore comolex |
| NTR | nuclear transport receptor |
| NTS | nuclear transport system |
| PBS | phosphate buffer saline |
| PRM | parallel reaction monitoring |
| PSM | peptide spectrum match |
| PTM | posttranslational modification |
| RF | radio frequency |
| RT | room temperature |
| SCCA | squamous cell carcinoma antigen |
| SCID | severe combined immunodeficiency |
| SILAC | stable isotope labeling by aminoacids in cell culture |
| SP3 | single-pot solid phase-enhanced sample preparation |
| SRM | selected reaction monitoring |
| TACE | trans arterial chemoembolization |
| TFA | trifluoroacetic acid |
| TMT | tandem mass tag |
| UPR | unfolded protein response |
| WT | wild type |

## 5.6 Publications

**Buczak, K**., Ori, A., Kirkpatrick, J.M., Holzer, K., Dauch, D., Roessler, S., Endris, V., Lasitschka, F.,Parca, L., Schmidt, A., Zender, L., Schirmacher, P., Krijgsveld, J., Singer, S., Beck, M., 2018. Spatial Tissue Proteomics Quantifies Inter- and Intratumor Heterogeneity in Hepatocellular Carcinoma (HCC). Mol. Cell. Proteomics 17, 810–825. doi:10.1074/mcp.RA117.000189

Sadian Y, Tafur L, Kosinski J, Jakobi AJ, Wetzel R, **Buczak K,** Hagen WJ, Beck M, Sachse C, Müller CW. Structural insights into transcription initiation by yeast RNA polymerase I. EMBO J. 2017 Sep 15;36(18):2698-2709. doi: 10.15252/embj.201796958. Epub 2017 Jul 24.

Kastritis PL, O'Reilly FJ, Bock T, Li Y, Rogon MZ, **Buczak K**, Romanov N, Betts MJ, Bui KH, Hagen WJ, Hennrich ML, Mackmull MT, Rappsilber J, Russell RB, Bork P, Beck M, Gavin AC. Capturing protein communities by structural proteomics in a thermophilic eukaryote. Mol Syst Biol. 2017 Jul 25;13(7):936. doi: 10.15252/msb.20167412

Ori, A., Iscar, M., **Buczak, K.,** Kastritis, P., Parca, L., Andrés-Pons, A., Singer, S.,Bork, P., and Beck, M. Spatiotemporal Variation of Mammalian Protein Complex Stoichiometries. *Genome Biology* 17 (2016): 47. doi:10.1186/s13059-016-0912-5.

Von Appen, A., Kosinski, J., Sparks, L., Ori, A., DiGuilio, A.L., Vollmer, B., Mackmull, M.-T., Banterle, N., Parca, L., Kastritis, P., **Buczak, K**., Mosalaganti, S., Hagen, W., Andres-Pons, A., Lemke, E.A., Bork, P., Antonin, W., Glavy, J.S., Khanh, H.B., and Beck, M. In Situ Structural Analysis of the Human Nuclear Pore Complex. *Nature* 526, no. 7571 (October 1, 2015): 140–43. doi:10.1038/nature15381.

## 6. Acknowledgements

Most importantly, I would like to express my sincere gratitude to my supervisor Dr. Martin Beck for giving me the opportunity to work in his laboratory. Thank you for all motivation, patience and support throughout the time of my PhD. I am glad that I could be a member of your group.

Next, I would like to thank Alessandro. During my time at EMBL I have learnt a lot for you. Thanks for your time and patience in explaining things to me, even after you have left EMBL. I can't image having a better teacher than you. It was a great experience to work with you.

Thanks Stephan for providing me with samples for the FFPE project and for trying to explain all the histological features of them during strange tours to the pathology department.

I want to acknowledge the former and current Proteomic Core Facility. Joanna, Mandy, Dominic and Per, thanks for keeping the machines running and performing well all the time!

Then I would like to thank all the member of the lab. You were the ones who really made me enjoying nearly every day of my work and often my time after lab. Marie, it was great to discuss some not necessarily serious stuff with you and to make fun of this "other part" of the lab! Shyamal, you really did great in "taking care" of my liver in last years. Thanks for every single beer we had! Thanks Bernhard for taking me up the crane and for not letting me starve during the writing time. Christian, thanks for the lesson I got from you, now I know that "it's not a wish concert". Thanks Andre for all of your ridiculous stories of loosing stuff, you made me laugh a lot. Thanks Beata for protecting me from strangers who appear close to döner places. Thanks Matteo for showing me how to use a single page to write entire book on it. Thanks Erica for taking care of my pipetting efficiency by fixing my thumb. Finally, I would like to thank Janina and Henrik for reading and correcting this thesis.

I also want to say thanks to former lab members: Amparo (for keeping everytging under control), Alex, Lenore (for your particular sense of humor), Amanda (for too many things to be listed here), Borja (for the great time in the guesthouse) and Panos.

I also want to thank my friends from both here and home for all fun we had together. But also for the support when things were not going well. Special thanks goes to Monika and Paweł.

Finally I would like to thank my family for always being supportive regardless of my decisions. Without you, I would not manage to achieve so much in life. Thank you.

# 7. References


Akhtar, A., Gasser, S.M., 2007. The nuclear envelope and transcriptional control. Nat. Rev. Genet. 8, 507–517. doi:10.1038/nrg2122

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., Sali, A., Rout, M.P., 2007. The molecular architecture of the nuclear pore complex. Nature 450, 695–701. doi:10.1038/nature06405

Alizadeh, A.A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., Caldas, C., Califano, A., Doherty, M., Elsner, M., Esteller, M., Fitzgerald, R., Korbel, J.O., Lichter, P., Mason, C.E., Navin, N., Pe'er, D., Polyak, K., Roberts, C.W.M., Siu, L., Snyder, A., Stower, H., Swanton, C., Verhaak, R.G.W., Zenklusen, J.C., Zuber, J., Zucman-Rossi, J., 2015. Toward understanding and exploiting tumor heterogeneity. Nat. Med. 21, 846–53. doi:10.1038/nm.3915

Alvero, A.B., Chen, R., Fu, H.-H., Montagna, M., Schwartz, P.E., Rutherford, T., Silasi, D.-A., Steffensen, K.D., Waldstrom, M., Visintin, I., Mor, G., 2009. Molecular phenotyping of human ovarian cancer stem cells unravels the mechanisms for repair and chemoresistance. Cell Cycle 8, 158–166. doi:10.4161/cc.8.1.7533

Antonin, W., Franz, C., Haselmann, U., Antony, C., Mattaj, I.W., 2005. The Integral Membrane Nucleoporin pom121 Functionally Links Nuclear Pore Complex Assembly and Nuclear Envelope Formation. Mol. Cell 17, 83–92. doi:10.1016/j.molcel.2004.12.010

Awad, M.M., Katayama, R., McTigue, M., Liu, W., Deng, Y.-L., Brooun, A., Friboulet, L., Huang, D., Falk, M.D., Timofeevski, S., Wilner, K.D., Lockerman, E.L., Khan, T.M., Mahmood, S., Gainor, J.F., Digumarthy, S.R., Stone, J.R., Mino-Kenudson, M., Christensen, J.G., Iafrate, A.J., Engelman, J.A., Shaw, A.T., 2013. Acquired Resistance to Crizotinib from a Mutation in *CD74 – ROS1*. N. Engl. J. Med. 368, 2395–2401. doi:10.1056/NEJMoa1215530

Bader, G.D., Hogue, C.W. V, 2003. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4, 2.

Beck, M., Schirmacher, P., Singer, S., 2017. Alterations of the nuclear transport system in hepatocellular carcinoma - New basis for therapeutic strategies. J. Hepatol. 67, 1051–1061. doi:10.1016/j.jhep.2017.06.021

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., Galon, J., 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25, 1091–3. doi:10.1093/bioinformatics/btp101

Bond, S.R., Naus, C.C., 2012. RF-Cloning.org: an online tool for the design of restriction-free cloning projects. Nucleic Acids Res. 40, W209–13. doi:10.1093/nar/gks396

Boyault, S., Rickman, D.S., de Reyniès, A., Balabaud, C., Rebouissou, S., Jeannot, E., Hérault, A., Saric, J., Belghiti, J., Franco, D., Bioulac-Sage, P., Laurent-Puig, P., Zucman-Rossi, J., 2007. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. Hepatology 45, 42–52. doi:10.1002/hep.21467

Brinkman, E.K., Kousholt, A.N., Harmsen, T., Leemans, C., Chen, T., Jonkers, J., van Steensel, B., 2018. Easy quantification of template-directed CRISPR/Cas9 editing. Nucleic Acids Res. 46, e58–e58. doi:10.1093/nar/gky164

Brosch, M., Yu, L., Hubbard, T., Choudhary, J., 2009. Accurate and sensitive peptide identification with Mascot Percolator. J. Proteome Res. 8, 3176–81. doi:10.1021/pr800982s

Brown, C.R., Silver, P.A., 2007. Transcriptional regulation at the nuclear pore complex. Curr. Opin. Genet. Dev. 17, 100–106. doi:10.1016/j.gde.2007.02.005

Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinović, S.M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., Reiter, L., 2015. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Mol. Cell. Proteomics 14, 1400–10. doi:10.1074/mcp.M114.044305

Buczak, K., Ori, A., Kirkpatrick, J.M., Holzer, K., Dauch, D., Roessler, S., Endris, V., Lasitschka, F., Parca, L., Schmidt, A., Zender, L., Schirmacher, P., Krijgsveld, J., Singer, S., Beck, M., 2018. Spatial Tissue Proteomics Quantifies Inter- and Intratumor Heterogeneity in Hepatocellular Carcinoma (HCC). Mol. Cell. Proteomics 17, 810–825. doi:10.1074/mcp.RA117.000189

Bui, K.H., von Appen, A., DiGuilio, A.L., Ori, A., Sparks, L., Mackmull, M.-T., Bock, T., Hagen, W., Andrés-Pons, A., Glavy, J.S., Beck, M., 2013. Integrated Structural Analysis of the Human Nuclear Pore Complex Scaffold. Cell 155, 1233–1243. doi:10.1016/j.cell.2013.10.055

Butler, S.L., Dong, H., Cardona, D., Jia, M., Zheng, R., Zhu, H., Crawford, J.M., Liu, C., 2008. The antigen for Hep Par 1 antibody is the urea cycle enzyme carbamoyl phosphate synthetase 1. Lab. Invest. 88, 78–88. doi:10.1038/labinvest.3700699

Chapman-Smith, A., Cronan Jr, J.E., 1999. Molecular Biology of Biotin Attachment to Proteins. J. Nutr. 129, 477S–484S. doi:10.1093/jn/129.2.477S

Cohen, M., Feinstein, N., Wilson, K.L., Gruenbaum, Y., 2003. Nuclear pore protein gp210 is essential for viability in HeLa cells and Caenorhabditis elegans. Mol. Biol. Cell 14, 4230–7. doi:10.1091/mbc.e03-04-0260

Cooks, R.G., 1995. Special feature: Historical. Collision-induced dissociation: Readings and commentary. J. Mass Spectrom. 30, 1215–1221. doi:10.1002/jms.1190300902

Cox, J., Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–72. doi:10.1038/nbt.1511

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J. V, Mann, M., 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. 10, 1794–805. doi:10.1021/pr101065j

Culjkovic-Kraljacic, B., Baguet, A., Volpon, L., Amri, A., Borden, K.L.B., 2012. The oncogene eIF4E reprograms the nuclear pore complex to promote mRNA export and oncogenic transformation. Cell Rep. 2, 207–15. doi:10.1016/j.celrep.2012.07.007

D'Angelo, M.A., Gomez-Cavazos, J.S., Mei, A., Lackner, D.H., Hetzer, M.W., 2012. A change in nuclear pore complex composition regulates cell differentiation. Dev. Cell 22, 446–58. doi:10.1016/j.devcel.2011.11.021

Dagogo-Jack, I., Shaw, A.T., 2017. Tumour heterogeneity and resistance to cancer therapies. Nat. Rev. Clin. Oncol. 15, 81–94. doi:10.1038/nrclinonc.2017.166

Das, M., Singh, S., Pradhan, S., Narayan, G., 2014. MCM Paradox: Abundance of Eukaryotic Replicative Helicases and Genomic Integrity. Mol. Biol. Int. 2014, 574850. doi:10.1155/2014/574850

Dauch, D., Rudalska, R., Cossa, G., Nault, J.-C., Kang, T.-W., Wuestefeld, T., Hohmeyer, A., Imbeaud, S., Yevsa, T., Hoenicke, L., Pantsar, T., Bozko, P., Malek, N.P., Longerich, T., Laufer, S., Poso, A., Zucman-Rossi, J., Eilers, M., Zender, L., 2016. A MYC-aurora kinase A protein complex represents an actionable drug target in p53-altered liver cancer. Nat. Med. 22, 744–53. doi:10.1038/nm.4107

Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D.F., Burkhard, P.R., Sanchez, J.-C., 2008. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. Anal. Chem. 80, 2921–2931. doi:10.1021/ac702422x

Dick, J.E., 2008. Stem cell concepts renew cancer research. Blood 112, 4793–4807. doi:10.1182/blood-2008-08-077941

Doerr, A., 2015. DIA mass spectrometry. Nat. Methods 12, 35–35. doi:10.1038/nmeth.3234

Dole, M., Mack, L.L., Hines, R.L., Mobley, R.C., Ferguson, L.D., Alice, M.B., 1968. Molecular Beams of Macroions. J. Chem. Phys. 49, 2240–2249. doi:10.1063/1.1670391

Dorner, A.J., Bole, D.G., Kaufman, R.J., 1987. The relationship of N-linked glycosylation and heavy chain-binding protein association with the secretion of glycoproteins. J. Cell Biol. 105, 2665–74.

Drummond, S.P., Wilson, K.L., 2002. Interference with the cytoplasmic tail of gp210 disrupts "close apposition" of nuclear membranes and blocks nuclear pore dilation. J. Cell Biol. 158, 53–62. doi:10.1083/jcb.200108145

Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–91. doi:10.1038/nprot.2009.97

Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365, 1687–1717. doi:10.1016/S0140-6736(05)66544-0

Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z., 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48. doi:10.1186/1471-2105-10-48

El-Serag, H.B., Kanwal, F., 2014. Epidemiology of hepatocellular carcinoma in the United States: where are we? Where do we go? Hepatology 60, 1767–75. doi:10.1002/hep.27222

Eriksson, C., Rustum, C., Hallberg, E., 2004. Dynamic properties of nuclear pore complex proteins in gp210 deficient cells. FEBS Lett. 572, 261–265. doi:10.1016/j.febslet.2004.07.044

Escher, C., Reiter, L., Maclean, B., Ossola, R., Herzog, F., Chilton, J., Maccoss, M.J., Rinner, O., 2012. Using iRT , a normalized retention time for more targeted measurement of peptides 1111–1121. doi:10.1002/pmic.201100463

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M., 1989. Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64–71.

Fialkow, P.J., 1979. Clonal Origin of Human Tumors. Annu. Rev. Med. 30, 135–143. doi:10.1146/annurev.me.30.020179.001031

Fox, C.H., Johnson, F.B., Whiting, J., Roller, P.P., 1985. Formaldehyde fixation. J. Histochem. Cytochem. 33, 845–853. doi:10.1177/33.8.3894502

Fraering, P., Imhof, I., Meyer, U., Strub, J.-M., van Dorsselaer, A., Vionnet, C., Conzelmann, A., 2001. The GPI Transamidase Complex of *Saccharomyces cerevisiae* Contains Gaa1p, Gpi8p, and Gpi16p. Mol. Biol. Cell 12, 3295–3306. doi:10.1091/mbc.12.10.3295

Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., Domon, B., 2012. Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. Mol. Cell. Proteomics 11, 1709–1723. doi:10.1074/mcp.O112.019802

Gatto, L., Lilley, K.S., 2011. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics 28, 288–289. doi:10.1093/bioinformatics/btr645

Gay, L., Baker, A.-M., Graham, T.A., 2016. Tumour Cell Heterogeneity. F1000Research 5. doi:10.12688/f1000research.7210.1

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y.H., Zhang, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80. doi:10.1186/gb-2004-5-10-r80

Gerdes, J., Schwab, U., Lemke, H., Stein, H., 1983. Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. Int. J. cancer 31, 13–20.

Gerlinger, M., Horswell, S., Larkin, J., Rowan, A.J., Salm, M.P., Varela, I., Fisher, R., McGranahan, N., Matthews, N., Santos, C.R., Martinez, P., Phillimore, B., Begum, S., Rabinowitz, A., Spencer-Dene, B., Gulati, S., Bates, P.A., Stamp, G., Pickering, L., Gore, M., Nicol, D.L., Hazell, S., Futreal, P.A., Stewart, A., Swanton, C., 2014. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat. Genet. 46, 225–233. doi:10.1038/ng.2891

Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N.Q., Butler, A., Jones, D., Raine, K., Latimer, C., Santos, C.R., Nohadani, M., Eklund, A.C., Spencer-Dene, B., Clark, G., Pickering, L., Stamp, G., Gore, M., Szallasi, Z., Downward, J., Futreal, P.A., Swanton, C., 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N. Engl. J. Med. 366, 883–92. doi:10.1056/NEJMoa1113205

Gething, M.-J., 1999. Role and regulation of the ER chaperone BiP. Semin. Cell Dev. Biol. 10, 465–472. doi:10.1006/scdb.1999.0318

Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., Aebersold, R., 2012. 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 11, O111.016717. doi:10.1074/mcp.O111.016717

Gomez-Cavazos, J.S., Hetzer, M.W., 2015. The nucleoporin gp210/Nup210 controls muscle differentiation by regulating nuclear envelope/ER homeostasis. J. Cell Biol. 208.

Gould, V.E., Martinez, N., Orucevic, A., Schneider, J., Alonso, A., 2000. A novel, nuclear pore-associated, widely distributed molecule overexpressed in oncogenesis and development. Am. J. Pathol. 157, 1605–13. doi:10.1016/S0002-9440(10)64798-0

Greaves, M., Maley, C.C., 2012. Clonal evolution in cancer. Nature 481, 306–13. doi:10.1038/nature10762

Greber, U.F., Senior, A., Gerace, L., 1990. A major glycoprotein of the nuclear pore complex is a membrane-spanning polypeptide with a large lumenal domain and a small cytoplasmic tail. EMBO J. 9, 1495–502.

Green, N.M., 1975. Avidin. Adv. Protein Chem. 29, 85–133.

Guichard, C., Amaddeo, G., Imbeaud, S., Ladeiro, Y., Pelletier, L., Maad, I. Ben, Calderaro, J., Bioulac-Sage, P., Letexier, M., Degos, F., Clément, B., Balabaud, C., Chevet, E., Laurent, A., Couchy, G., Letouzé, E., Calvo, F., Zucman-Rossi, J., 2012. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat. Genet. 44, 694–8. doi:10.1038/ng.2256

Guna, A., Volkmar, N., Christianson, J.C., Hegde, R.S., 2018. The ER membrane protein complex is a transmembrane domain insertase. Science (80-. ). 359, 470–473. doi:10.1126/science.aao3099

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R., 1999. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17, 994–999. doi:10.1038/13690

Hammond, M.E.H., Hayes, D.F., Dowsett, M., Allred, D.C., Hagerty, K.L., Badve, S., Fitzgibbons, P.L., Francis, G., Goldstein, N.S., Hayes, M., Hicks, D.G., Lester, S., Love, R., Mangu, P.B., McShane, L., Miller, K., Osborne, C.K., Paik, S., Perlmutter, J., Rhodes, A., Sasano, H., Schwartz, J.N., Sweep, F.C.G., Taube, S., Torlakovic, E.E., Valenstein, P., Viale, G., Visscher, D., Wheeler, T., Williams, R.B., Wittliff, J.L., Wolff, A.C., 2010. American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. J. Clin. Oncol. 28, 2784–2795. doi:10.1200/JCO.2009.25.6529

Harris, L.N., Ismaila, N., McShane, L.M., Andre, F., Collyar, D.E., Gonzalez-Angulo, A.M., Hammond, E.H., Kuderer, N.M., Liu, M.C., Mennel, R.G., Van Poznak, C., Bast, R.C., Hayes, D.F., American Society of Clinical Oncology, 2016. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. J. Clin. Oncol. 34, 1134–1150. doi:10.1200/JCO.2015.65.2289

Hough, S.H., Ajetunmobi, A., Brody, L., Humphryes-Kirilov, N., Perello, E., 2016. Desktop Genetics. Per. Med. 13, 517–521. doi:10.2217/pme-2016-0068

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., Vingron, M., 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18 Suppl 1, S96–104.

Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., Steinmetz, L.M., Krijgsveld, J., 2014. Ultrasensitive proteome analysis using paramagnetic bead technology. Mol. Syst. Biol. 10, 757.

Hughes, C.S., McConechy, M.K., Cochrane, D.R., Nazeran, T., Karnezis, A.N., Huntsman, D.G., Morin, G.B., 2016. Quantitative Profiling of Single Formalin Fixed Tumour Sections: proteomics for translational research. Sci. Rep. 6, 34949. doi:10.1038/srep34949

Iribarne, J. V., Thomson, B.A., 1976. On the evaporation of small ions from charged droplets. J. Chem. Phys. 64, 2287. doi:10.1063/1.432536

Iwasa, Y., Nowak, M.A., Michor, F., 2005. Evolution of Resistance During Clonal Expansion. Genetics 172, 2557–2566. doi:10.1534/genetics.105.049791

Jacoby, M.A., Duncavage, E.J., Walter, M.J., 2015. Implications of Tumor Clonal Heterogeneity in the Era of Next-Generation Sequencing. Trends in cancer 1, 231–241. doi:10.1016/j.trecan.2015.10.006

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., von Mering, C., 2009. STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 37, D412–6. doi:10.1093/nar/gkn760

Joshi, S., Tolkunov, D., Aviv, H., Hakimi, A.A., Yao, M., Hsieh, J.J., Ganesan, S., Chan, C.S., White, E., 2015. The Genomic Landscape of Renal Oncocytoma Identifies a Metabolic Barrier to Tumorigenesis. Cell Rep. 13, 1895–908. doi:10.1016/j.celrep.2015.10.059

Jue-Liang Hsu, †, Sheng-Yu Huang, †, Nan-Haw Chow, ‡ and, Shu-Hui Chen*, †, 2003. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. doi:10.1021/AC0348625

Karlsson, C., Karlsson, M.G., 2011. Effects of long-term storage on the detection of proteins, DNA, and mRNA in tissue microarray slides. J. Histochem. Cytochem. 59, 1113–21. doi:10.1369/0022155411423779

Kennedy, D., Samali, A., Jäger, R., 2015. Methods for Studying ER Stress and UPR Markers in Human Cells. Humana Press, New York, NY, pp. 3–18. doi:10.1007/978-1-4939-2522-3_1

Kettenbach, A.N., Rush, J., Gerber, S.A., 2011. Absolute quantification of protein and post-translational modification abundance with stable isotope-labeled synthetic peptides. Nat. Protoc. 6, 175–86. doi:10.1038/nprot.2010.196

Kim, D.I., KC, B., Zhu, W., Motamedchaboki, K., Doye, V., Roux, K.J., 2014. Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc. Natl. Acad. Sci. 111, E2453–E2461. doi:10.1073/pnas.1406459111

Kinoshita, T., 2014. Enzymatic mechanism of GPI anchor attachment clarified. Cell Cycle 13, 1838–9. doi:10.4161/cc.29379

Konermann, L., Ahadi, E., Rodriguez, A.D., Vahidi, S., 2013. Unraveling the Mechanism of Electrospray Ionization. Anal. Chem. 85, 2–9. doi:10.1021/ac302789c

Kosinski, J., Mosalaganti, S., von Appen, A., Teimer, R., DiGuilio, A.L., Wan, W., Bui, K.H., Hagen, W.J.H., Briggs, J.A.G., Glavy, J.S., Hurt, E., Beck, M., 2016. Molecular architecture of the inner ring scaffold of the human nuclear pore complex. Science (80-. ). 352, 363–365. doi:10.1126/science.aaf0643

Kreso, A., Dick, J.E., 2014. Evolution of the Cancer Stem Cell Model. Cell Stem Cell 14, 275–291. doi:10.1016/J.STEM.2014.02.006

Kumar, L., E Futschik, M., 2007. Mfuzz: a software package for soft clustering of microarray data. Bioinformation 2, 5–7.

Kwon, K., Streaker, E.D., Ruparelia, S., Beckett, D., 2000. Multiple Disordered Loops Function in Corepressor-induced Dimerization of the Biotin Repressor. J. Mol. Biol. 304, 821–833. doi:10.1006/jmbi.2000.4249

Lange, V., Picotti, P., Domon, B., Aebersold, R., 2008. Selected reaction monitoring for quantitative proteomics: a tutorial. Mol. Syst. Biol. 4, 222. doi:10.1038/msb.2008.61

Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M.A., Dick, J.E., 1994. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature 367, 645–648. doi:10.1038/367645a0

Larsen, S., Nielsen, J., Hansen, C.N., Nielsen, L.B., Wibrand, F., Stride, N., Schroder, H.D., Boushel, R., Helge, J.W., Dela, F., Hey-Mogensen, M., 2012. Biomarkers of mitochondrial content in skeletal muscle of healthy young human subjects. J. Physiol. 590, 3349–60. doi:10.1113/jphysiol.2012.230185

Lemke, E.A., 2016. The Multiple Faces of Disordered Nucleoporins. J. Mol. Biol. 428, 2011–2024. doi:10.1016/J.JMB.2016.01.002

Ley, T.J., Mardis, E.R., Ding, L., Fulton, B., McLellan, M.D., Chen, K., Dooling, D., Dunford-Shore, B.H., McGrath, S., Hickenbotham, M., Cook, L., Abbott, R., Larson, D.E., Koboldt, D.C., Pohl, C., Smith, S., Hawkins, A., Abbott, S., Locke, D., Hillier, L.W., Miner, T., Fulton, L., Magrini, V., Wylie, T., Glasscock, J., Conyers, J., Sander, N., Shi, X., Osborne, J.R., Minx, P., Gordon, D., Chinwalla, A., Zhao, Y., Ries, R.E., Payton, J.E., Westervelt, P., Tomasson, M.H., Watson, M., Baty, J., Ivanovich, J., Heath, S., Shannon, W.D., Nagarajan, R., Walter, M.J., Link, D.C., Graubert, T.A., DiPersio, J.F., Wilson, R.K., 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 456, 66–72. doi:10.1038/nature07485

Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., Robertson, A.G., Hoadley, K., Triche, T.J., Laird, P.W., Baty, J.D., Fulton, L.L., Fulton, R., Heath, S.E., Kalicki-Veizer, J., Kandoth, C., Klco, J.M., Koboldt, D.C., Kanchi, K.-L., Kulkarni, S., Lamprecht, T.L., Larson, D.E., Lin, L., Lu, C., McLellan, M.D., McMichael, J.F., Payton, J., Schmidt, H., Spencer, D.H., Tomasson, M.H., Wallis, J.W., Wartman, L.D., Watson, M.A., Welch, J., Wendl, M.C., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y., Chiu, R., Chu, A., Chuah, E., Chun, H.-J., Corbett, R., Dhalla, N., Guin, R., He, A., Hirst, C., Hirst, M., Holt, R.A., Jones, S., Karsan, A., Lee, D., Li, H.I., Marra, M.A., Mayo, M., Moore, R.A., Mungall, K., Parker, J., Pleasance, E., Plettner, P., Schein, J., Stoll, D., Swanson, L., Tam, A., Thiessen, N., Varhol, R., Wye, N., Zhao, Y., Gabriel, S., Getz, G., Sougnez, C., Zou, L., Leiserson, M.D.M., Vandin, F., Wu, H.-T., Applebaum, F., Baylin, S.B., Akbani, R., Broom, B.M., Chen, K., Motter, T.C., Nguyen, K., Weinstein, J.N., Zhang, N., Ferguson, M.L., Adams, C., Black, A., Bowen, J., Gastier-Foster, J., Grossman, T., Lichtenberg, T., Wise, L., Davidsen, T., Demchok, J.A., Shaw, K.R.M., Sheth, M., Sofia, H.J., Yang, L., Downing, J.R., Eley, G., 2013. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. N. Engl. J. Med. 368, 2059–2074. doi:10.1056/NEJMoa1301689

Li, C., Heidt, D.G., Dalerba, P., Burant, C.F., Zhang, L., Adsay, V., Wicha, M., Clarke, M.F., Simeone, D.M., 2007. Identification of pancreatic cancer stem cells. Cancer Res. 67, 1030–7. doi:10.1158/0008-5472.CAN-06-2030

Li, W., Qi, Y., Cui, X., Sun, Y., Huo, Q., Yang, Y., Wen, X., Tan, M., Du, S., Zhang, H., Zhang, M., Liu, C., Kong, Q., 2017. Heteroplasmy and Copy Number Variations of Mitochondria in 88 Hepatocellular Carcinoma Individuals. J. Cancer 8, 4011–4017. doi:10.7150/jca.21218

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., Yates, J.R., 1999. Direct analysis of protein complexes using mass spectrometry 17.

Liu, F., Rijkers, D.T.S., Post, H., Heck, A.J.R., 2015. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. Nat. Methods 12, 1179–1184. doi:10.1038/nmeth.3603

Liu, K., Lee, J., Kim, J.Y., Wang, L., Tian, Y., Chan, S.T., Cho, C., Machida, K., Chen, D., Ou, J.-H.J., 2017. Mitophagy Controls the Activities of Tumor Suppressor p53 to Regulate Hepatic Cancer Stem Cells. Mol. Cell 68, 281–292.e5. doi:10.1016/j.molcel.2017.09.022

Liu, L., Xie, S., Zhang, C., Zhu, F., 2014. Aberrant regulation of alternative pre-mRNA splicing in hepatocellular carcinoma. Crit. Rev. Eukaryot. Gene Expr. 24, 133–49.

Louris, J.N., Cooks, R.G., Syka, J.E.P., Kelley, P.E., Stafford, G.C., Todd, J.F.J., 1987. Instrumentation, applications, and energy deposition in quadrupole ion-trap tandem mass spectrometry. Anal. Chem. 59, 1677–1685. doi:10.1021/ac00140a021

Lytton, J., Westlin, M., Hanley, M.R., 1991. Thapsigargin inhibits the sarcoplasmic or endoplasmic reticulum Ca-ATPase family of calcium pumps. J. Biol. Chem. 266, 17067–71.

Mackmull, M.-T., Klaus, B., Heinze, I., Chokkalingam, M., Beyer, A., Russell, R.B., Ori, A., Beck, M., 2017. Landscape of nuclear transport receptor cargo specificity. Mol. Syst. Biol. 13, 962.

Makarov, A., 2000. Electrostatic Axially Harmonic Orbital Trapping : A High-Performance Technique of Mass Analysis 72, 1156–1162.

Manza, L.L., Stamer, S.L., Ham, A.-J.L., Codreanu, S.G., Liebler, D.C., 2005. Sample preparation and digestion for proteomic analyses using spin filters. Proteomics 5, 1742–1745. doi:10.1002/pmic.200401063

Marcu, M.G., Doyle, M., Bertolotti, A., Ron, D., Hendershot, L., Neckers, L., 2002. Heat shock protein 90 modulates the unfolded protein response by stabilizing IRE1alpha. Mol. Cell. Biol. 22, 8506–13. doi:10.1128/MCB.22.24.8506-8513.2002

Martínez, N., Alonso, A., Moragues, M.D., Pontón, J., Schneider, J., 1999. The nuclear pore complex protein Nup88 is overexpressed in tumor cells. Cancer Res. 59, 5408–11.

Marusyk, A., Polyak, K., 2010. Tumor heterogeneity: causes and consequences. Biochim. Biophys. Acta 1805, 105–17. doi:10.1016/j.bbcan.2009.11.002

McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wühr, M., Huttlin, E.L., Erickson, B.K., Rad, R., Haas, W., Gygi, S.P., 2014. MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes. Anal. Chem. 86, 7150–7158. doi:10.1021/ac502040v

Mogensen, M., Bagger, M., Pedersen, P.K., Fernström, M., Sahlin, K., 2006. Cycling efficiency in humans is related to low UCP3 content and to type I fibres but not to mitochondrial efficiency. J. Physiol. 571, 669–681. doi:10.1113/jphysiol.2005.101691

Murtaza, M., Dawson, S.-J., Tsui, D.W.Y., Gale, D., Forshew, T., Piskorz, A.M., Parkinson, C., Chin, S.-F., Kingsbury, Z., Wong, A.S.C., Marass, F., Humphray, S., Hadfield, J., Bentley, D., Chin, T.M., Brenton, J.D., Caldas, C., Rosenfeld, N., 2013. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 497, 108–112. doi:10.1038/nature12065

Nassar, A., Radhakrishnan, A., Cabrero, I.A., Cotsonis, G.A., Cohen, C., 2010. Intratumoral Heterogeneity of Immunohistochemical Marker Expression in Breast Carcinoma. Appl. Immunohistochem. Mol. Morphol. 18, 1. doi:10.1097/PAI.0b013e3181dddb20

Nowell, P.C., 1976. The clonal evolution of tumor cell populations. Science 194, 23–8.

Olsen, J. V, Macek, B., Lange, O., Makarov, A., Horning, S., Mann, M., 2007. Higher-energy C-trap dissociation for peptide modification analysis. Nat. Methods 4, 709–712. doi:10.1038/nmeth1060

Olsson, M., Schéele, S., Ekblom, P., 2004. Limited expression of nuclear pore membrane glycoprotein 210 in cell lines and tissues suggests cell-type specific nuclear pores in metazoans. Exp. Cell Res. 292, 359–70.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., 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, 376–86. doi:10.1074/MCP.M200025-MCP200

Ori, A., Banterle, N., Iskar, M., Andrés-Pons, A., Escher, C., Khanh Bui, H., Sparks, L., Solis-Mezarino, V., Rinner, O., Bork, P., Lemke, E.A., Beck, M., 2013. Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines. Mol. Syst. Biol. 9, 648. doi:10.1038/msb.2013.4

Ostasiewicz, P., Zielinska, D.F., Mann, M., Wiśniewski, J.R., 2010. 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, 3688–700. doi:10.1021/pr100234w

Panté, N., Kann, M., 2002. Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. Mol. Biol. Cell 13, 425–34. doi:10.1091/mbc.01-06-0308

Pao, W., Miller, V.A., Politi, K.A., Riely, G.J., Somwar, R., Zakowski, M.F., Kris, M.G., Varmus, H., 2005. Acquired Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib Is Associated with a Second Mutation in the EGFR Kinase Domain. PLoS Med. 2, e73. doi:10.1371/journal.pmed.0020073

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B. V, Curry, W.T., Martuza, R.L., Louis, D.N., Rozenblatt-Rosen, O., Suvà, M.L., Regev, A., Bernstein, B.E., 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344, 1396–401. doi:10.1126/science.1254257

Picard, M., Taivassalo, T., Ritchie, D., Wright, K.J., Thomas, M.M., Romestaing, C., Hepple, R.T., 2011. Mitochondrial Structure and Function Are Disrupted by Standard Isolation Methods. PLoS One 6, e18317. doi:10.1371/journal.pone.0018317

Polyak, K., Haviv, I., Campbell, I.G., 2009. Co-evolution of tumor cells and their microenvironment. Trends Genet. 25, 30–38. doi:10.1016/j.tig.2008.10.012

Quintana, E., Shackleton, M., Sabel, M.S., Fullen, D.R., Johnson, T.M., Morrison, S.J., 2008. Efficient tumour formation by single human melanoma cells. Nature 456, 593–598. doi:10.1038/nature07567

Rabut, G., Doye, V., Ellenberg, J., 2004. Mapping the dynamic organization of the nuclear pore complex inside single living cells. Nat. Cell Biol. 6, 1114–21. doi:10.1038/ncb1184

Raices, M., Bukata, L., Sakuma, S., Borlido, J., Hernandez, L.S., Hart, D.O., D'Angelo, M.A., 2017. Nuclear Pores Regulate Muscle Development and Maintenance by Assembling a Localized Mef2C Complex. Dev. Cell 41, 540–554.e7. doi:10.1016/j.devcel.2017.05.007

Raza, A., Sood, G.K., 2014. Hepatocellular carcinoma review: current treatment, and evidence-based medicine. World J. Gastroenterol. 20, 4115–27. doi:10.3748/wjg.v20.i15.4115

Roessler, S., Jia, H.-L., Budhu, A., Forgues, M., Ye, Q.-H., Lee, J.-S., Thorgeirsson, S.S., Sun, Z., Tang, Z.-Y., Qin, L.-X., Wang, X.W., 2010. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer Res. 70, 10202–12. doi:10.1158/0008-5472.CAN-10-2607

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D.J., 2004. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 3, 1154–69. doi:10.1074/mcp.M400129-MCP200

Roux, K.J., Kim, D.I., Raida, M., Burke, B., 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. 196, 801–10. doi:10.1083/jcb.201112098

Saha, S., Anilkumar, A.A., Mayor, S., 2016. GPI-anchored protein organization and dynamics at the cell surface. J. Lipid Res. 57, 159–75. doi:10.1194/jlr.R062885

Sahu, A., Prabhash, K., Noronha, V., Joshi, A., Desai, S., 2013. Crizotinib: A comprehensive review. South Asian J. cancer 2, 91–7. doi:10.4103/2278-330X.110506

Sanyal, A.J., Yoon, S.K., Lencioni, R., 2010. The etiology of hepatocellular carcinoma and consequences for treatment. Oncologist 15 Suppl 4, 14–22. doi:10.1634/theoncologist.2010-S4-14

Sawayama, H., Ishimoto, T., Sugihara, H., Miyanari, N., Miyamoto, Y., Baba, Y., Yoshida, N., Baba, H., 2014. Clinical impact of the Warburg effect in gastrointestinal cancer (review). Int. J. Oncol. 45, 1345–54. doi:10.3892/ijo.2014.2563

Schatton, T., Murphy, G.F., Frank, N.Y., Yamaura, K., Waaga-Gasser, A.M., Gasser, M., Zhan, Q., Jordan, S., Duncan, L.M., Weishaupt, C., Fuhlbrigge, R.C., Kupper, T.S., Sayegh, M.H., Frank, M.H., 2008. Identification of cells initiating human melanomas. Nature 451, 345–349. doi:10.1038/nature06489

Scheuner, D., Song, B., McEwen, E., Liu, C., Laybutt, R., Gillespie, P., Saunders, T., Bonner-Weir, S., Kaufman, R.J., 2001. Translational control is required for the unfolded protein response and in vivo glucose homeostasis. Mol. Cell 7, 1165–76.

Schulze, K., Imbeaud, S., Letouzé, E., Alexandrov, L.B., Calderaro, J., Rebouissou, S., Couchy, G., Meiller, C., Shinde, J., Soysouvanh, F., Calatayud, A.-L., Pinyol, R., Pelletier, L., Balabaud, C., Laurent, A., Blanc, J.-F., Mazzaferro, V., Calvo, F., Villanueva, A., Nault, J.-C., Bioulac-Sage, P., Stratton, M.R., Llovet, J.M., Zucman-Rossi, J., 2015. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat. Genet. 47, 505–11. doi:10.1038/ng.3252

Schwartz, T.U., 2016. The Structure Inventory of the Nuclear Pore Complex. J. Mol. Biol. 428, 1986–2000. doi:10.1016/J.JMB.2016.03.015

Scigelova, M., Makarov, A., 2006. Orbitrap Mass Analyzer – Overview and Applications in Proteomics. Proteomics 6, 16–21. doi:10.1002/pmic.200600528

Sequist, L. V, Waltman, B.A., Dias-Santagata, D., Digumarthy, S., Turke, A.B., Fidias, P., Bergethon, K., Shaw, A.T., Gettinger, S., Cosper, A.K., Akhavanfard, S., Heist, R.S., Temel, J., Christensen, J.G., Wain, J.C., Lynch, T.J., Vernovsky, K., Mark, E.J., Lanuti, M., Iafrate, A.J., Mino-Kenudson, M., Engelman, J.A., 2011. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Sci. Transl. Med. 3, 75ra26. doi:10.1126/scitranslmed.3002003

Shaw, A.T., Friboulet, L., Leshchiner, I., Gainor, J.F., Bergqvist, S., Brooun, A., Burke, B.J., Deng, Y.-L., Liu, W., Dardaei, L., Frias, R.L., Schultz, K.R., Logan, J., James, L.P., Smeal, T., Timofeevski, S., Katayama, R., Iafrate, A.J., Le, L., McTigue, M., Getz, G., Johnson, T.W., Engelman, J.A., 2016. Resensitization to Crizotinib by the Lorlatinib *ALK* Resistance Mutation L1198F. N. Engl. J. Med. 374, 54–61. doi:10.1056/NEJMoa1508887

Shen, K., Sun, J., Cao, X., Zhou, D., Li, J., 2015. Comparison of Different Buffers for Protein Extraction from Formalin-Fixed and Paraffin-Embedded Tissue Specimens. PLoS One 10, e0142650. doi:10.1371/journal.pone.0142650

Shi, S.R., Key, M.E., Kalra, K.L., 1991. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. J. Histochem. Cytochem. 39, 741–8.

Shurtleff, M.J., Itzhak, D.N., Hussmann, J.A., Schirle Oakdale, N.T., Costa, E.A., Jonikas, M., Weibezahn, J., Popova, K.D., Jan, C.H., Sinitcyn, P., Vembar, S.S., Hernandez, H., Cox, J., Burlingame, A.L., Brodsky, J.L., Frost, A., Borner, G.H., Weissman, J.S., 2018. The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. Elife 7. doi:10.7554/eLife.37018

Sielaff, M., Kuharev, J., Bohn, T., Hahlbrock, J., Bopp, T., Tenzer, S., Distler, U., 2017. Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. J. Proteome Res. 16, 4060–4072. doi:10.1021/acs.jproteome.7b00433

Singer, S., Ehemann, V., Brauckhoff, A., Keith, M., Vreden, S., Schirmacher, P., Breuhahn, K., 2007. Protumorigenic overexpression of stathmin/Op18 by gain-of-function mutation in p53 in human hepatocarcinogenesis. Hepatology 46, 759–68. doi:10.1002/hep.21736

Sinn, H.-P., Kreipe, H., 2013. A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. Breast Care (Basel). 8, 149–54. doi:10.1159/000350774

Siravegna, G., Bardelli, A., 2016. Blood circulating tumor DNA for non-invasive genotyping of colon cancer patients. Mol. Oncol. 10, 475–480. doi:10.1016/J.MOLONC.2015.12.005

Smyth, G.K., 2005. Limma : Linear Models for Microarray Data.

Song, J., Wang, Q., Luo, Y., Yuan, P., Tang, C., Hui, Y., Wang, Z., 2015. miR-34c-3p inhibits cell proliferation, migration and invasion of hepatocellular carcinoma by targeting MARCKS. Int. J. Clin. Exp. Pathol. 8, 12728–37.

Stavru, F., Nautrup-Pedersen, G., Cordes, V.C., Görlich, D., 2006. Nuclear pore complex assembly and maintenance in POM121- and gp210-deficient cells. J. Cell Biol. 173, 477–83. doi:10.1083/jcb.200601002

Stewart, B.W., Wild, C.P., n.d. World cancer report 2014.

Strimmer, K., 2008. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics 24, 1461–2. doi:10.1093/bioinformatics/btn209

Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 6, e21800. doi:10.1371/journal.pone.0021800

Takahashi, N., van Kilsdonk, J.W.J., Ostendorf, B., Smeets, R., Bruggeman, S.W.M., Alonso, A., van de Loo, F., Schneider, M., van den Berg, W.B., Swart, G.W.M., 2008. Tumor marker nucleoporin 88 kDa regulates nucleocytoplasmic transport of NF-kappaB. Biochem. Biophys. Res. Commun. 374, 424–30. doi:10.1016/j.bbrc.2008.06.128

Thavarajah, R., Mudimbaimannar, V.K., Elizabeth, J., Rao, U.K., Ranganathan, K., 2012. Chemical and physical basics of routine formaldehyde fixation. J. Oral Maxillofac. Pathol. 16, 400–5. doi:10.4103/0973-029X.102496

Thierry, A.R., El Messaoudi, S., Mollevi, C., Raoul, J.L., Guimbaud, R., Pezet, D., Artru, P., Assenat, E., Borg, C., Mathonnet, M., De La Fouchardière, C., Bouché, O., Gavoille, C., Fiess, C., Auzemery, B., Meddeb, R., Lopez-Crapez, E., Sanchez, C., Pastor, B., Ychou, M., 2017. Clinical utility of circulating DNA analysis for rapid detection of actionable mutations to select metastatic colorectal patients for anti-EGFR treatment. Ann. Oncol. 28, 2149–2159. doi:10.1093/annonc/mdx330

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K.A., Hamon, C., 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 75, 1895–904.

Titz, B., Elamin, A., Martin, F., Schneider, T., Dijon, S., Ivanov, N. V, Hoeng, J., Peitsch, M.C., 2014. Proteomics for systems toxicology. CSBJ 11, 73–90. doi:10.1016/j.csbj.2014.08.004

Toyama, B.H., Savas, J.N., Park, S.K., Harris, M.S., Ingolia, N.T., Yates, J.R., Hetzer, M.W., 2013. Identification of long-lived proteins reveals exceptional stability of essential cellular structures. Cell 154, 971–82. doi:10.1016/j.cell.2013.07.037

Troyanovsky, R.B., Sokolov, E., Troyanovsky, S.M., 2003. Adhesive and lateral E-cadherin dimers are mediated by the same interface. Mol. Cell. Biol. 23, 7965–72.

Tsukumo, Y., Tomida, A., Kitahara, O., Nakamura, Y., Asada, S., Mori, K., Tsuruo, T., 2007. Nucleobindin 1 Controls the Unfolded Protein Response by Inhibiting ATF6 Activation. J. Biol. Chem. 282, 29264–29272. doi:10.1074/jbc.M705038200

Turashvili, G., Brogi, E., 2017. Tumor Heterogeneity in Breast Cancer. Front. Med. 4, 227. doi:10.3389/fmed.2017.00227

Upla, P., Kim, S.J., Sampathkumar, P., Dutta, K., Cahill, S.M., Chemmama, I.E., Williams, R., Bonanno, J.B., Rice, W.J., Stokes, D.L., Cowburn, D., Almo, S.C., Sali, A., Rout, M.P., Fernandez-Martinez, J., 2017. Molecular Architecture of the Major Membrane Ring Component of the Nuclear Pore Complex. Structure 25, 434–445. doi:10.1016/j.str.2017.01.006

Vachet, R.W., McElvany, S.W., 1999. Application of external customized waveforms to a commercial quadrupole ion trap. J. Am. Soc. Mass Spectrom. 10, 355–359. doi:10.1016/S1044-0305(99)00007-0

Van Schadewijk, A., van't Wout, E.F.A., Stolk, J., Hiemstra, P.S., 2012. A quantitative method for detection of spliced X-box binding protein-1 (XBP1) mRNA as a measure of endoplasmic reticulum (ER) stress. Cell Stress Chaperones 17, 275–9. doi:10.1007/s12192-011-0306-2

Vogel, C., Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat. Publ. Gr. 13, 227–232. doi:10.1038/nrg3185

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Kinzler, K.W., Kinzler, K.W., 2013. Cancer genome landscapes. Science 339, 1546–58. doi:10.1126/science.1235122

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164. doi:10.1093/nar/gkq603

Wang, M., Kaufman, R.J., 2014. The impact of the endoplasmic reticulum protein-folding environment on cancer development. Nat. Rev. Cancer 14, 581–597. doi:10.1038/nrc3800

Warburg, O., 1956. On the origin of cancer cells. Science 123, 309–14.

Wiesner, J., Premsler, T., Sickmann, A., 2008. Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. Proteomics 8, 4466–4483. doi:10.1002/pmic.200800329

Winkler, J., Ori, A., Holzer, K., Sticht, C., Dauch, D., Eiteneuer, E.M., Pinna, F., Geffers, R., Ehemann, V., Andres-Pons, A., Breuhahn, K., Longerich, T., Bermejo, J.L., Gretz, N., Zender, L., Schirmacher, P., Beck, M., Singer, S., 2014. Prosurvival function of the cellular apoptosis susceptibility/importin-α1 transport cycle is repressed by p53 in liver cancer. Hepatology 60, 884–895. doi:10.1002/hep.27207

Wiśniewski, J.R., 2013. Proteomic sample preparation from formalin fixed and paraffin embedded tissue. J. Vis. Exp. doi:10.3791/50589

Wozniak, R.W., Bartnik, E., Blobel, G., 1989. Primary structure analysis of an integral membrane glycoprotein of the nuclear pore. J. Cell Biol. 108, 2083–92.

Wozniak, R.W., Blobel, G., Rout, M.P., 1994. POM152 is an integral protein of the pore membrane domain of the yeast nuclear envelope. J. Cell Biol. 125, 31–42.

Xu, S., Xu, Y., Chen, L., Fang, Q., Song, S., Chen, J., Teng, J., 2017. RCN1 suppresses ER stress-induced apoptosis via calcium homeostasis and PERK–CHOP signaling. Oncogenesis 6, e304. doi:10.1038/oncsis.2017.6

Yao, X., Freas, A., Ramirez, J., Demirev, P.A., Fenselau, C., 2001. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. Anal. Chem. 73, 2836–42.

Yates, L.R., Gerstung, M., Knappskog, S., Desmedt, C., Gundem, G., Van Loo, P., Aas, T., Alexandrov, L.B., Larsimont, D., Davies, H., Li, Y., Ju, Y.S., Ramakrishna, M., Haugland, H.K., Lilleng, P.K., Nik-Zainal, S., McLaren, S., Butler, A., Martin, S., Glodzik, D., Menzies, A., Raine, K., Hinton, J., Jones, D., Mudie, L.J., Jiang, B., Vincent, D., Greene-Colozzi, A., Adnet, P.-Y., Fatima, A., Maetens, M., Ignatiadis, M., Stratton, M.R., Sotiriou, C., Richardson, A.L., Lønning, P.E., Wedge, D.C., Campbell, P.J., 2015. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. Nat. Med. 21, 751–759. doi:10.1038/nm.3886

Ye, J., Rawson, R.B., Komuro, R., Chen, X., Davé, U.P., Prywes, R., Brown, M.S., Goldstein, J.L., 2000. ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. Mol. Cell 6, 1355–64.

Yoshida, H., Matsui, T., Yamamoto, A., Okada, T., Mori, K., 2001. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. Cell 107, 881–91.

Yu, H.A., Arcila, M.E., Hellmann, M.D., Kris, M.G., Ladanyi, M., Riely, G.J., 2014. Poor response to erlotinib in patients with tumors containing baseline EGFR T790M mutations found by routine clinical molecular testing. Ann. Oncol. 25, 423–428. doi:10.1093/annonc/mdt573

Yuan, R.-H., Jeng, Y.-M., Chen, H.-L., Lai, P.-L., Pan, H.-W., Hsieh, F.-J., Lin, C.-Y., Lee, P.-H., Hsu, H.-C., 2006. Stathmin overexpression cooperates with p53 mutation and osteopontin overexpression, and is associated with tumour progression, early recurrence, and poor prognosis in hepatocellular carcinoma. J. Pathol. 209, 549–58. doi:10.1002/path.2011

Zhang, S., Balch, C., Chan, M.W., Lai, H.-C., Matei, D., Schilder, J.M., Yan, P.S., Huang, T.H.-M., Nephew, K.P., 2008. Identification and Characterization of Ovarian Cancer-Initiating Cells from Primary Human Tumors. Cancer Res. 68, 4311–4320. doi:10.1158/0008-5472.CAN-08-0364

Zhao, Y.-J., Ju, Q., Li, G.-C., 2013. Tumor markers for hepatocellular carcinoma. Mol. Clin. Oncol. 1, 593–598. doi:10.3892/mco.2013.119

Zheng, T., Chen, M., Han, S., Zhang, L., Bai, Y., Fang, X., Ding, S.-Z., Yang, Y., 2014. Plasma minichromosome maintenance complex component 6 is a novel biomarker for hepatocellular carcinoma patients. Hepatol. Res. 44, 1347–1356. doi:10.1111/hepr.12303