TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Life Sciences

Studying *in vitro*-differentiated hepatocytes by proteomics

Johannes Krumm

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Prof. Angelika Schnieke, Ph.D

Prüfer der Dissertation: 1. Prof. Dr. Bernhard Küster
2. Prof. Dr. Maximilian Reichert

Die Dissertation wurde am 01.02.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 28.06.2022 angenommen.

Abstract

Mass spectrometry-based proteomcis is a powerful tool for acquiring comprehensive and unbiased expression information on thousands of proteins. However, high sample amounts are required in order to achieve an adequate depth. Hence, in this study the suitability of several proteomics approaches to start from quantity-limited samples was evaluated. The most promising methodologies were further fine-tuned and their usefulness for low-input samples was proven. Optimizing the in-StageTip method yielded around 10,000 peptides from only 2,000 cells, which is 5 times more than with the original protocol. Only the SP3 approach, which is based on the precipitation of proteins to magnetic beads, was more sensitive. In addition, the latter enables the removal of detergents and demonstrated great compatibility with the chemical Tandem Mass Tag labeling approach. In this regard, the identification and quantification quality of several other magnetic microparticles was assessed, including IMAC-beads for phosphopeptide enrichment.

In the second part of this work, hepatocytes generated from human pluripotent stem cells were studied. The *in vitro* differentiated hepatocyte-like cells are increasingly discussed as potential alternatives for primary hepatocytes, which are scarce, show high donor variability, and quickly change their properties upon culturing. Especially in the field of regenerative medicine and drug discovery, the pluripotent stem cell-derived hepatocytes could be of great value. However, the underlying differentiation process is incompletely characterized. Therefore, quantitative proteomics experiments at multiple time points during the differentiation were employed resulting in the detection of around 9,000 proteins, 12,000 phosphorylation sites, and 800 acetylation sites. The expression dynamics revealed a major protein rewiring between the hepatic endoderm and immature hepatocyte-like cells. While metabolic and cell surface proteins were upregulated, levels of cell cycle-related proteins, epigenetic modifiers, and transcription factors diminished. Furthermore, the high temporal resolution enabled to define novel and specific protein markers for each development stage.

In a complementary comprehensive (phospho)proteomic experiment, 2D and 3D pluripotent stem cell-derived hepatocytes were compared against fetal and adult liver samples to elaborate on how well they can recapitulate their *in vivo* counterpart. Indeed, the 3D model was superior to the monolayer approach in expressing several liver-specific proteins. However, both differentiation derivatives lacked the expression of sufficiently high levels of ADME/Tox proteins, which are crucial for metabolising therapeutic drugs in the liver. Moreover, the phosphorylation data indicated divergent kinase activity, suggesting that the stem cell-derived hepatocytes resemble a rather fetal state. Based on the obtained results a molecular roadmap of hepatocyte differentiation was constructed, which enhances the understanding of the underlying biology and serves as a proteomics resource. Additionally, several starting points for differentiation protocol improvements were inferred from the collective data.

In summary, this study demonstrated the versatile application of an optimized SP3 beads-based proteomic workflow, specifically for minute protein quantities. Furthermore, the analysis of protein expression and post-translational modifications was utilized to shed light on

developmental processess and to provide recommendations for protocol modifications of pluripotent stem cell-derived hepatocyte differentiation.

Zusammenfassung

Die Massenspektrometriebasierte Proteomik ist eine leistungsfähige Methode zur umfassenden und unvoreingenommenen Expressionsmessung von Tausenden Proteinen. Allerdings ist für das Erreichen einer solch hinreichenden Tiefe eine große Probenmenge nötig. Daher wurden in dieser Arbeit mehrere proteomische Ansätze hinsichtlich ihrer Eignung im Bezug auf Proben mit begrenzter Menge untersucht. Die vielversprechendsten Methoden wurden weiter verfeinert und ihre Nützlichkeit für Proben mit geringem Materialeinsatz wurde nachgewiesen. Durch die Optimierung der in-StageTip Methode wurden rund 10.000 Peptide aus nur 2.000 Zellen gemessen, was einer Verfünffachung gegenüber des originalen Protokolls entspricht. Lediglich der SP3 Ansatz, welcher auf der Präzipitation von Proteinen auf magnetische Kügelchen basiert, war noch sensitiver. Zusätzlich ermöglicht Letzterer die Beseitigung von Detergenzien und zeigte gute Kompatibilität mit der Tandem Mass Tag Markierung. In diesem Zusammenhang wurde die Qualität der Identifizierung und Quantifizierung verschiedener anderer magnetischer Mikropartikel getestet, einschließlich IMAC-Kügelchen die für die Anreicherung von Phosphopeptiden verwendet werden.

Im zweiten Teil dieser Arbeit wurden aus menschlichen pluripotenten Stammzellen hergestellte Hepatozyten untersucht. Diese *in vitro* differenzierten Hepatozytenähnlichen Zellen werden vermehrt als mögliche Alternativen für primäre Zellen diskutiert, da diese rar sind, eine hohe Variabilität zwischen Spendern aufweisen und schnell ihre Eigenschaften in Kultur verlieren. Insbesondere im Bereich der regenerativen Medizin und der Arzneimittelforschung könnten diese Hepatozyten aus pluripotenten Stammzellen von großen Wert sein. Allerdings ist deren zu Grunde liegender Differenzierungsprozess unzureichend charakterisiert. Daher wurde ein quantitatives proteomisches Experiment mit mehreren Zeitpunkten während der Differenzierung durchgeführt was zu 9.000 Proteinen, 12.000 Phosphorylierungsstellen und 800 Acetylierungsstellen führte. Die Dynamik der Expression zeigte starke Veränderungen zwischen dem hepatischen Endoderm und den unreifen Hepatozytenähnlichen Zellen. Während Stoffwechsel und Zelloberflächenproteine anstiegen, nahmen die Expressionslevel von Zellzyklusproteinen, epigentischen Modifikatoren und Transkriptionsfaktoren ab. Darüber hinaus konnten durch die hohe zeitliche Auflösung neue und spezifischen Proteinmarker für jedes Entwicklungsstadium definiert werden.

In einem ergänzenden umfangreichen (Phospho)proteom Experiment wurden aus pluripotenten Stammzellen gewonnene 2D- und 3D-Hepatozyten mit fötalen und adulten Leberproben verglichen, um zu untersuchen, wie gut diese ihre *in vivo* Gegenstücke darstellen können. Tatsächlich war das 3D-Model dem Monolayer-Ansatz bei der Expression mehrerer leberspezifischer Proteine überlegen. Jedoch zeigten beide Differenzierungsderivate keine ausreichend hohe Expression von ADME/Tox Proteinen, welche essentiell für die Verstoffwechselung von therapeutischen Medikamenten in der Leber sind. Zudem deuteten die Phosphorylierungsdaten auf abweichende Kinaseaktivität hin und suggerierten, dass die aus Stammzellen hergestellten Hepatozyten eher ein fötales Stadium darstellen. Auf der Grundlage dieser Ergebnisse wurde eine molekulare Roadmap für Hepatozytendifferenzierung erstellt, welche sowohl das Verständnis der zugrunde liegenden Biologie verbessert, als auch als

proteomische Resource dient. Außerdem wurden aus den gesammelten Daten einige Ansatzpunkte für Verbesserungen des Differenzierungsprotokolls abgeleitet.
Zusammenfassend lässt sich sagen, dass diese Arbeit die vielfältige Anwendung eines optimierten, auf SP3 Kügelchen-basierenden Arbeitsablauf für die Proteomik demonstriert, insbesondere für kleinste Proteinmengen. Darüber hinaus wurde die Analyse von Proteinexpression und post-translationaler Modifikationen genutzt, um Aufschluss über Entwicklungsprozesse zu geben und Empfehlungen für Protokolländerungen bei der Differenzierung von Hepatozyten aus pluripotenten Stammzellen zu geben.

Table of contents

Chapter I: General introduction

1 Stem cells and their value for research

1.1 Stem cell properties and classification

Stem cells are rapidly dividing cells which are capable of self-renewal over a long period of time and have the potential to differentiate into various different cell types. Depending on their differentiation potency, they can be classified into four groups (I-Figure 1; as reviewed in [1]).

Totipotent stem cells

Totipotent stem cells comprise the highest differentiation potential of all stem cells. They are capable of developing into any cell type and are therefore the only cells that can form an entire organism (reviewed in [2]). In the human body a totipotent cell, the zygote, is formed through the fusion of a spermatozoon with an oocyte. During further development, the zygote undergoes cell division which results in a decrease of potency. After approximately 4-5 days, the zygote specifies into the inner cell mass (ICM) and the extra-embryonic cell lineage. While the extra-embryonic cell lineage matures further to the placenta, the ICM gives rise to the embryonic lineage and possesses only pluripotent characteristics. The exact time and cell stage at which totipotency is lost is still controversial and challenging to study as the number of totipotent cells and their accessibility is limited (reviewed in [3]).

Pluripotent stem cells

The second most potent cells are pluripotent stem cells (PSCs). They are capable of differentiating into each of the three germ layers of the embryonic lineage, but are not able to form extra-embryonic structures (reviewed in [4]). Human embryonic stem cells (hESCs) are one example of pluripotent cells. They can be isolated from the ICM and cultured *in vitro* as first shown by Thompson *et al.* [5] as well as Reubinoff *et al.* [6]. However, the hESCs research is challenging as cells are scarce and their use is strictly regulated. Some of these challenges were overcome by the discovery of induced pluripotent stem cells (iPSCs) in 2006, which was a scientific breakthrough that was rewared with the Nobel Price in 2012. Yamanaka and colleagues were able to regain pluripotency from terminally differentiated somatic cells by the transduction of four transcription factors: Oct4, Sox2, Klf4, and c-Myc [7, 8]. While in the beginning iPSCs were mostly generated from fibroblasts, nowadays a variety of sources exist, such as blood cells or keratinocytes [9-11]. Another source are cells from the urinary system which are excreted in the urine. This is of special interest, because it is a noninvasive method with almost unlimited capacities [12]. Urine-derived iPSCs showed the potential to differentiate into cardiomyocytes [13] and neural progenitors [14] indicating a promising and valuable resource for iPSCs. In the original reprogramming protocol of Yamanaka and colleagues retroviral transduction was utilized to deliver the four crucial transcriptions factors [7]. As this method is prone for introducing genetic alterations, several alternatives were developed (reviewed in [15, 16]). For example, transduction with an adenovirus which enables iPSC generation of fibroblasts without transgenic integration [17]. One of the most commonly used systems for generating iPSCs at the moment is the Sendai virus vector [18], which is non-integrating and capable of infecting a broad range of host cells with a high efficiency [19]. Another alternative is the non-viral piggyback transposon system which is able to integrate the

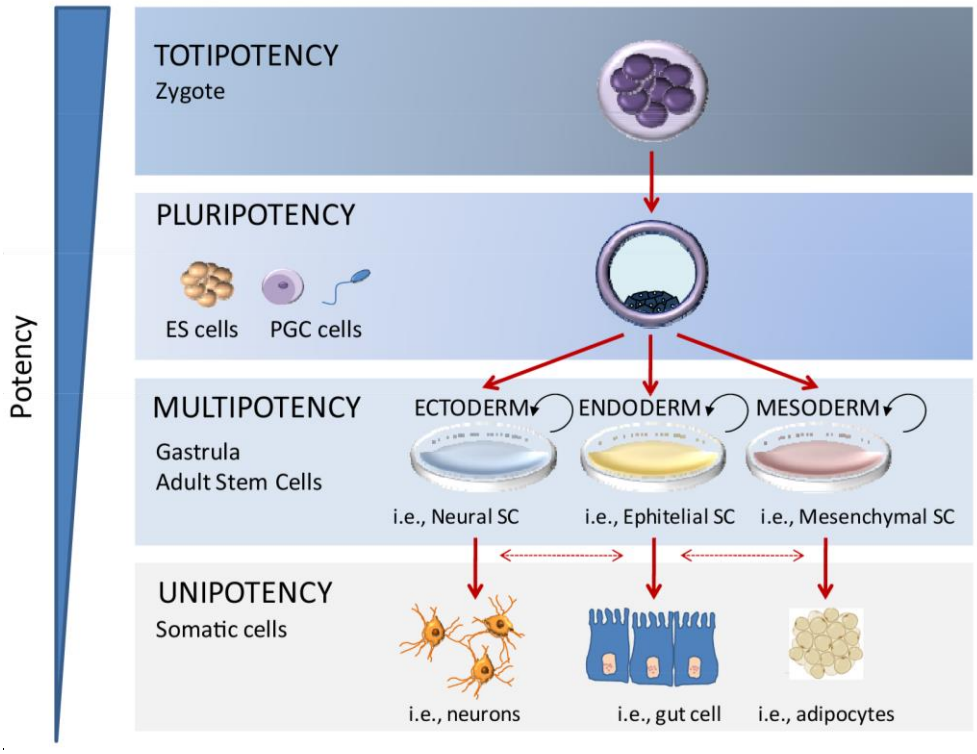
reprogramming factors into somatic cells [20] and appears to be a promising method, but has shown lower efficiency than virus-based methods so far [21]. In general, non-integrating and non-viral vectors with high reprogramming efficiency are desirable in order to make this method applicabale for medical administration. As iPSCs can be harvested and derived from patients, not only are the ethical concerns lower, but also the rejection of the immune system is decreased. Despite the great potential to renew or replace damaged patient tissues, no iPSC-based therapy has yet made its way into routine clinical applications, although several are in clinical trials [22]. For their routine application, the remaining challenges such as carcinogeneity [23] and genomic instability [24] have to be overcome.

Multipotent stem cells

Multipotent stem cells have less potency compared to PSCs, but still possess important stem cell characteristics such as high self-renewal capacity and the ability to specialize into several different cells. One example are mesenchymal stem cells, which are adult stem cells that can be obtained from various different origins, like the bone marrow or the umbilical cord [25, 26]. Mesenchymal stem cells can, for example, differentiate into cells from the mesodermal lineage such, adipocytes [27], osteocytes [28], or cartilage [29]. Furthermore, some studies postulated the differentiation of mesenchymal stem cells into the endodermal [30, 31] and ectodermal [32] lineage. However, this differentiation outside of the mesenchymal lineage is widely debated [33, 34]. Nevertheless, mesenchymal stem cells are considered as a promising source for regenerative medicine for example for cartilage repair or bone regeneration (as reviewed in [35]).

Unipotent stem cells

Apart from multipotent stem cells, unipotent stem cells comprise the second class of adult stem cells. They are characterized by the least differentiation potency and are only able to generate one specific cell type. For example, male germline stem cells are unipotent precursors which exclusively differentiate into sperm cells. However, studies have shown that these unipotent stem cells can be dedifferentiated upon culturing with growth factors to gain pluripotency [36, 37]. Compared to iPSCs and ESCs, pluripotent cells obtained from germline stem cells suffer from less ethical concerns, a decreased risk of tumorigenesis, and diminished immune rejection (reviewed in [38]).



I-Figure 1: Schematic of stem cell classification and their differentiation potency. Modified with permission from Berdasco *et al.* [39] and created with BioRender.com.

1.2 Directed stem cell differentiation

The potency of stem cells to specify into different cell lineages can be used and controlled in a process called directed differentiation. This process enables to mimic and study some of the fundamental developing steps of embryogenesis *in vitro*. Knowing the fundamental developmental mechanism, allows to spot errors in this process and explain the resulting malfunctions or diseases. Besides studying the underlying basic biology, directed cell differentiation offers a promising approach for several clinical applications like regenerative medicine or personalized medicine as well as for drug discovery.

Directed differentiation of pluripotent stem cells *in vitro*

In order to direct cell fate, the knowledge of essential developmental pathways, key transcription factors, or other crucial proteins is required. Based on this expertise, recombinant proteins and small molecules, which are both powerful tools, can be utilized to manipulate cell differentiation. Although the application of recombinant proteins has been successful in recent years, it is associated with high costs and batch-to-batch variations. Since chemical molecules are usually more stable, cheaper, and non-immunogenic, they might be the go-to tool in the future.

In order to accomplish directed differentiation from PSCs, the first step comprises germ layer induction using growth factors [40]. Once the desired germ layer is formed, additional supplementation of specific proteins and small molecules enable further cell lineage specification. This process usually needs extensive optimization, as the right concentration, duration, and timing

of the supplemented reagents is essential. Many protocols are developed for just one specific pluripotent cell line and the adaption to a different cell line can be tedious. Moreover, further modifications are usually required when the differentiation protocol was developed in a different organism. Although *in vitro* differentiated cells can resemble many functions of their *in vivo* counterparts, they very often represent rather immature characteristics. One reason for this is the varying differentiation capacity observed between cell lines as well as between clones from the same cell line, which leads to a heterogenous cell population rather than a very specific cell type [41]. Another reason is that the cell development *in vitro* usually only takes a fraction of the time compared to the developmental process in a living organism.

Organoid models and their applications

Traditionally, *in vitro* differentiation was performed in a petri dish and grown in two-dimensional (2D) monolayer. However, this experimental setup does not entirely recapitulate the *in vivo* situation, as for example cell-cell interactions are vastly diminished and thus, only a limited view of the developmental process is represented. To overcome some of these challenges, three-dimensional (3D) organoid models have been developed, which are self-organized structures composed of one or more cell types. They resemble the *in vivo* conditions more closely and have been superior to 2D models in terms of cell-cell communication, viability, differentiation, cell polarization, and drug metabolism, just to name a few (as reviewed in [42, 43]). Stem cells have been used to generate functional organoids from all three germ layers [44]. For example, the development of a brain organoid system, called cerebral organoids, enabled to elucidate on human brain development, which has been difficult to assess before. By using organoids derived from PSCs, not only discrete brain regions were differentiated but also features of microcephaly were mimicked and a potential cause of this disease was derived [45]. Antonica and colleagues were able to generate functional thyroid follicles from ESCs *in vitro*, which had the capacity to organify iodide [46]. Upon transplantation into mice, thyroid hormone deficiency was overcome showing the potential impact of this powerful tool for regenerative medicine in the future [46].

Personaliced medicine is another research field with great capacity for the organoid technology. For example, it has been used for drugs targeting cystic fibrosis [47]. Here, iPSCs were generated from CF patients and further differentiated into cholangiocytes which matured through organoid formation. These organoids not only mimicked the CF disease, but further reproduced drug effects *in vitro*. In a second study, rectal organoids from biopsies of CF patients were derived [48]. The drug responses of this *in vitro* model positively correlated with data from clinical trials demonstrating the ability of this technology for predicting drug effects of individual patients.

While personaliced medicine is applied to individual patients, drug discovery usually makes use of high-throughput screening of biologically active molecules. In pharma industry these screenings are often based on primary cells or animal models, but the organoid culture system might support or even substitute some of the current pipelines. To this regard, StemoniX® offers a platform of commercially available brain and heart organoids derived from iPSCs and developed for high-throughput applications. Further, a platform for colon organoids in a 384-well as well as 1536-well formats have been introduced for drug screening [49]. However, high-throughput applications have been challenging so far, because organoids are formed by self-organization which is a stochastic manner that is difficult to control. This decreases the reproducibility leading to batch-

to-batch variations and has to be overcome in order to make large-scale screening on organoid models reasonable. Since the 3D culture system is still in its infancy, further optimization is required to exploit its full potential and to make it widely applicable.

Transdifferentiation of somatic cells

Besides the dedifferentiation of somatic cells to iPSC and their subsequent directed differentiation, transdifferentiation of somatic cells offers an alternative for generating specific cell types (reviewed in [50]). Here, somatic cells are directly reprogrammed into a different cell type using small molecules, transcription factors, or a combination of both. This was first shown in 1987, when Davis and colleagues converted fibroblast into muscle cells by the transfection of MyoD [51]. Then, in 2008 the first *in vivo* application was published where the combination of Ngn3, Pdx1, and Mafa resulted in the reprogramming of pancreatic exocrine cells into functional β-cells [52]. The directed transdifferentiation with small molecules offers a fast method without the requirement of viral transduction to generate specific cells types, while the risk of tumorigenesis is low. However, this method is limited by low differentiation efficiencies as well as the small number of cases where transdifferentiation was successful so far. Noteworthy, besides the artificially induced transdifferentiation, reprogramming and lineage transition has been observed naturally in zebrafish, where atrial cardiomyocytes converted into functional ventricular cardiomyocytes [53].

Post-transclational modifications and their function in regulating pluripotency and differentiation

As shown by Yamanaka and colleagues the introduction of transcriptions factors enables the reprogramming of somatic cells into iPSCs, which possess the capacity to specify into several different cell types [7, 8]. This shows how crucial gene expression for the pluripotent state is. However, studies have shown varying repogramming and differentiation efficiencies among PSCs, which indicates that transcript levels are not solely decisive for cell fate decisions [41]. This seems obvious as many signaling pathways related to cell differentiation are regulated by post-translational modifications (PTMs) (reviewed in [54]). For example, several studies have elucidated on the connection between pluripotency and PTMs, such as ubiquitination [55], sumoylation [56], glycosylation [57], and methylation [58]. Various different types of PTMs exist, however the majority rely on enzymatic modifications of amino acid side chains or the protein's termini. By manipulating the biophysical properties of proteins, PTMs are capable of influencing gene transcription, enzymatic activity, protein translocation, protein degradation, or protein-protein interactions, just to name a few [59].

Phosphorylation is not only one of the most commonly observed PTMs, but it is also involved in controlling various signaling pathways [60]. Phosphorylation is enzymatically controlled by kinases and phosphatases, which add or remove phosphate groups from hydroxyl groups of serine, threonine, and tyrosine amino acid residues. Transfer of this phosphate group changes the net charge of amino acids, which often results in a conformational change and altered protein characteristics. Many signaling pathways known for regulating cell differentiation and pluripotency are controlled through phosphorylation, e.g. via receptor tyrosine kinases such as FGFR. A large scale study performed by Swaney and colleagues detected more than 10,000

phosphorylation sites (P-sites) in ESC and among them several on the two critical transcription factors for pluripotency SOX2 and OCT4 [61]. In a different study more than 50% of the detected P-sites were observed to be modified already after 1 hour of differentiation showing how rapid and vast this modification is [62]. To this regard, Yang and colleagues performed a multi-omics survey of ESCs transitioning from the naïve state to the primed state and found that changes on the phosphoproteome level precede changes on epigenome, transcriptome, and proteome [63]. This shows that phosphorylation possessess a crucial role in cell development, but further research is necessary in order to grasp the full picture.

The acetylation of the lysine side chain ε-amino group is another reversible PTM with significant effects on protein characteristics. While lysine/histone acetyltransferases catalyze the transfer of acetyl groups from acetyl-CoA to lysine residues, lysine/histone deacetylases (KDAC/HDAC) are capable of reverting this process [64]. Although acetylation was first discovered only on histone proteins [65], acetylation on many non-histone proteins has been reported (as reviewed in [64]). The addition of an acetyl group removes the positive charge of the lysine residue which can have a significant impact on protein characteristics. For example, elevated histone acetylation is associated with increased chromatin accessibility and therefore higher gene transcription rates [66]. On the other hand, low levels of histone acetylation result in decreased chromatin accessibility and transcriptional hypoactivity [54]. Considering that in general pluripotency is associated with a high rate of transcriptional activity while differentiation is accompanied with decreased activity [67], this PTM is highly associated with cell development. To this regard, several studies have shown positive effects on the reprogramming efficiency of somatic cells using HDAC inhibitors which target histone proteins [21, 68, 69]. On the contrary, HDAC inhibitors have also been shown to support endoderm differentiation [70], improve the bone formation from mesenchymal stem cell [71], induce neural cell differentiation [72, 73], and support the expansion of hematopoietic stem cells [74]. This process gets even more complex when considering the timing component of this PTM. For example, high levels of acetylation support the acinus fate of pancreatic progenitors, whereas low levels promote ductal specification [75, 76]. Although the understanding so far is rather rudimentary, it could already be shown how closely pluripotency and acetylation levels are interwoven and we are just beginning to understand their connection.

In summary, although the addition and modulation of transcription factors have shown to induce pluripotency and regulate cell differentiation, several studies have shown a considerable impact of PTMs in these processes. However, their influence is incompletely understood and a more detailed perception may be helpful for appreciating the differentiation better and improve future applications.

2 Stem cell differentiation towards hepatocytes

2.1 Liver: Physiology and embryonic development

The liver is located in the right upper part of the abdomen and is the largest internal organ of the human body with approximately 1.6 kg. It consists of two unequally sized lobes and it is tightly connected to the blood system via the hepatic artery as well as the portal vein.

Biological functions of the liver

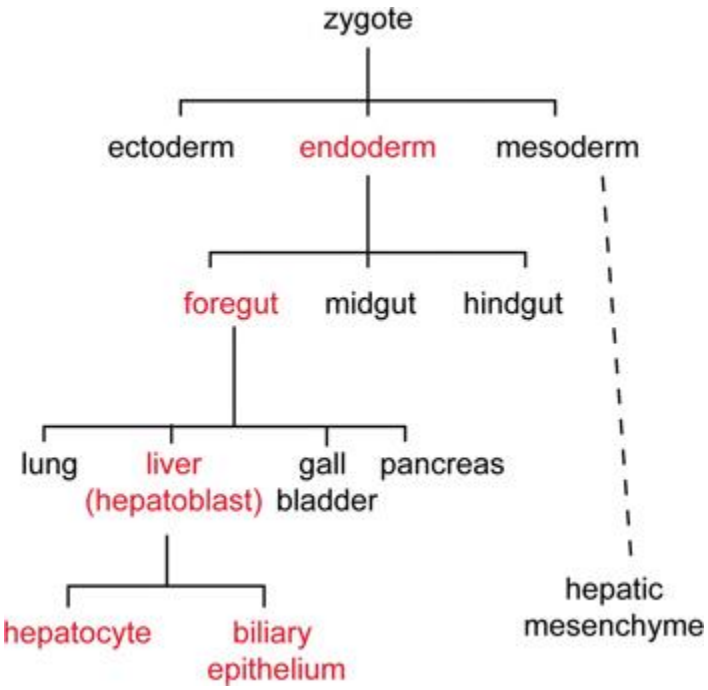
Around 70-80% of the liver mass can be accounted to hepatocytes, which is the main parenchymal cell type fulfilling most of the vital organ functions. One key feature of the liver is the regulation of the blood glucose level (as reviewed in [77]). In order to do this, hepatocytes surround the portal vein, which passes through the liver, and absorb nutrients coming from the gastrointestinal tract. Under carbohydrate-rich conditions, glucose is converted into glyocogen or fatty acids and stored. As the glucose level in the blood decreases, glycogen is enzymatically broken down to maintain the blood glucose level constant and to nourish other organs, such as the brain [77]. Once the glycogen storage is exhausted, the liver is able to convert precursors, such as lactate, glycerol, alanine, or glutamine, into glucose in a process called gluconeogenesis [78]. Apart from the carbohydrate metabolism, the majority of plasma proteins, such as serum albumin, are synthesized here [79]. The liver is also an important organ for lipid metabolism, as it produces cholesterol and degrades fatty acids. Excessive fatty acids which are not internally used for energy supply can be converted into ketone bodies by liver cells and distributed to other tissues for energy supply. The liver also produces bile and secretes it via the biliary tract into the gallbladder for storage. Another important feature of hepatocytes is the detoxification of molecules such as ammonia, a toxic product of amino acid catabolism, which can be converted into urea through the urea cycle and subsequently excreted. Furthermore, hepatocytes are capable of absorbing and metabolizing bioactive compounds from the blood. This is of special interest for drug discovery, as this process often leads to toxic intermediates which is one of the main causes for drugs failing in preclinical and clinical trials [80, 81]. Additionally, the absorption significantly affects a drugs half life and thus its efficacy.

Apart from hepatocytes, the liver is composed of non-parenchymal cells including kupffer, stellate, and sinusoidal endothelial cells. The kupffer cells are macrophagic cells, which are capable of taking up particles or pathogens from their environment by phagocytosis. They can stimulate cytokine production, which makes them a crucial part of the innate immune system, but also a key player in disease states like inflammation or steatosis [82]. The hepatic stellate cells reside in a quiescent state and their functions are not fully understood. However, it is known that under chronical stress conditions, such as viral infections or alcohol abuse, stellate cells are activated and differentiate into myofibroblast, which are the main driver for liver fibrosis [83, 84]. A key regulator of stellate cell activation are the third type of non-parenchymal cells, the sinusoidal endothelial cells. An *in vitro* study showed that they are capable of keeping stellate cells in a quiescent state making them an important gatekeeper for liver fibrosis [85]. Moreover, sinusoidal endothelial cells comprise immunological functions, such as the elimination of pathogens, similar to kupffer cells [86]. Furthermore, the non-parenchymal liver cells maintain and regulate liver growth and participate in the control of hepatocyte proliferation [87].

Embryonic development of hepatocytes

Embryonic development is initialized by the fertilization of a male and a female gamete resulting in a zygote. Through several rounds of mitotic cell division, the single fertilized egg forms a group of cells, the blastomere, during the first days. After 5-7 days, a blastocyst is formed which consists of an outer layer, the trophoectoderm which will later develop into the placenta, and the ICM, which will give rise to the embryo. This blastocyst implants into the uterus and starts the

gastrulation. During this process, cells migrate through the process of epithelial-mesenchymal transition (EMT) from the epiblast into the primitive streak and form the ectoderm as well as the mesendoderm, which is the precursor of the definitive endoderm (DE) and the mesoderm (I-Figure 2). While organs like skin and brain are formed from the ectoderm, the mesoderm developes, for instance, into muscle cells and kidneys [88]. On the other hand, the endoderm forms a primitive gut consisting of anterior foregut, posterior foregut, midgut, and hindgut regions. From the posterior foregut hepatic progenitors, the hepatoblasts, are derived. These hepatoblasts form the liver bud, which is at this developmental stage the main producer of fetal blood cells [89]. Later the hepatoblasts can differentiate to biliary endothelial cells, which are progenitors for the bile ducts that are used to transport bile salts generated from the liver into the gallbladder. The majority of hepatoblasts, however, develop into immature hepatocytes which undergo a maturation process that even continues after birth. While hepatocytes are obtained from the endodermal lineage, the non-parenchymal cells are of mesodermal origin (reviewed in [89]).



I-Figure 2: Hepatocyte cell lineage specification. Through gastrulation the fertilized zygote forms the three germ layers ectoderm, endoderm, and mesoderm. Then the foregut formation is enhanced in order to differentiate hepatoblasts, which are progenitors with the capacity to develop hepatocytes and biliary epithelium. The stepwise hepatic development is highlighted in red (adapted from [89]).

2.2 Hepatocyte differentiation and applications

The liver possesses the capacity to fully re-grow even after 70% was removed, which is unique among organs in the human body [90]. Despite this regenerative potential, organ donation is the only cure for severe liver failure. As organ donors are scarce, the artificial production of liver cells is recognized as an alternative source. Hence, several different methods for hepatocyte generation have been developed over recent years. This chapter will highlight some of these

methods with a special emphasis on stem cell approaches and will draft a picture of the versatile applications of differentiated hepatocytes not only in medicine but also for basic research as well as drug discovery.

Hepatocytes derived from pluripotent stem cells

PSCs are a widely used source for generating hepatocyte-like cells *in vitro*. In the beginning, embryonic bodies, which are self-aggregated, heterogeneous, 3D structures of ESCs, were employed to generate hepatocyte-like cells [91]. However, this technique suffers from low differentiation efficiency and reproducibility. Gaining more knowledge about crucial differentiation conditions, including signaling pathways, transcription factors, and the extracellular matrix, paved the way to a directed differentiation of PSCs towards hepatocytes. Several protocols have been developed in recent years and are commonly divided into three steps: DE differentiation, hepatoblast specification, and hepatocyte maturation (reviewed in [92]).

As DE is the common origin of several endodermal tissues, such as pancreas, thyroid, and liver, the faithful formation of DE is crucial for further differentiation. *In vivo* TGFβ signaling is activated during gastrulation in order to suppress ectodermal specification and induce the mesendodermal lineage formation. Mouse experiments showed that the mesendodermal lineage formation is highly regulated by the TGFβ superfamily member Nodal together with downstream Smad2 signaling [93]. While high levels of Nodal lead to DE generation, low levels support the specification towards the mesodermal lineage [94]. These Nodal signals can be mimicked *in vitro* with activin A, which binds to the same receptor protein and activates TGFβ signaling leading to an increased expression of DE markers like SOX17 and FOXA2 [95]. Other groups have shown that DE differentiation can also be supported by the canonical Wnt pathway [96, 97]. Therefore, several protocols use a combination of activin A and WNT3a to trigger both pathways [98-100]. However, the duration of supplementation has to be controlled as prolonged Wnt signaling during primitive gut tube formation favours hindgut over foregut specification [101]. Hence, during this time Wnt inhibitors are expressed to maintain foregut identity [102]. Adding to the elaborated timely interplay, Toivonen and colleagues demonstrated that extended WNT stimulation for more than 1 day resulted in decreased competence of stem cells to form pancreas progenitors, while hepatocyte-like cells could still be obtained after 5-7 days [103]. In a different approach recombinantly expressed proteins are substituted by small molecules, which are more stable and chemically defined. Siller and colleagues used the GSK-3 inhibitor CHIR99021 to activate the canonical Wnt pathway and promote DE induction, which led to comparable differentiation efficiencies [104].

During the second step, hepatoblasts, the precursors of hepatocytes, are specified from the DE. *In vivo* this process is promoted by the secretion of cytokines from the mesodermal tissues in close proximity to the endoderm. Jung and colleagues showed that expression of fibroblast growth factors (FGF) 1, 2, and 8, in the surrounding mesoderm triggered liver development from foregut endoderm in mice [105]. In addition, bone morphogenetic protein (BMP) was shown to be essential for hepatic specification from the endoderm [106]. Multiple different protocols have been developed to mimic this developmental process *in vitro*. Although they vary in the exact administration of the recombinant protein and its concentration as well as the duration, they commonly activate the FGF and BMP signaling pathways [99, 107, 108]. Other protocols

circumvent the application of recombinant proteins by using small molecules, such as sodium butyrate or DMSO to enhance hepatoblast formation [100, 104, 109].

Once hepatoblasts are derived from stem cells, they are solely capable of differentiating into hepatocytes or cholangiocytes. Cholangiocyte commitment can be supported with the supplementation of activin A, FGF-10, and retionoic acid, which results in the expression of early biliary markers, such as KRT19 and SOX9 [110]. On the contrary, hepatocyte maturation is promoted by the combination of hepatocyte growth factor (HGF), oncostatin M (OSM), fibroblast growth factor (FGF), and dexamethasone. OSM is an interleukin-6 family member which is highly expressed in hematopoietic cells in the fetal liver [111]. While its supplementation promotes hepatocyte maturation of hepatic progenitors *in vitro*, the differentiation towards hematopoietic lineage is suppressed [112-114]. Like OSM, HGF attenuates cholangiocyte differentiation and promotes upregulation of hepatic proteins, such as albumin [112]. As an alternative, the HGF mimetic N-hexanoic-tyrosine-isoleucine-(6) aminohexanoic amide can be used, which showed high affinity towards HGF and a comparable induction of c-Met phosphorylation [104, 115]. Lastly, dexamethasone is a glucocorticoid with antiphlogistic characteristics. However in the context of liver development, it enhanced the expression of HNF4 and C/EBPα, which are key transcription factors triggering hepatocyte maturation *in vitro* [116]. As reviewed by Palakkan and colleagues various protocols for hepatocyte maturation exist [92], which all use different combinations and conditions of the afore mentioned recombinant proteins or small molecules.

Besides the utilization of recombinant proteins and small molecules, cell-cell and cell-matrix interactions are essential drivers for cell fate regulation and proliferation [117]. Especially for hepatocytes as they are epithelial cells with pronounced adhesion properties and strong interactions with the surrounding extracellular matrix (ECM). Proteins of the ECM interact with cellular receptors, such as integrins, which enables them to manipulate the cell via downstream signal transduction [118]. There are several artificial ECM alternatives that mimic *in vivo* properties and allow for more physiological differentiation [119]. Among them, matrigel and laminin have been shown to be useful [120, 121].

Applications of plutipotent stem cell-derived hepatocytes

Hepatocytes derived from directed differentiation are capable of mimicking various functions of their *in vivo* counterparts. To this end, Takebe and colleagues were the first to demonstrate that *in vitro* generated liver organoids have the capacity to fulfill organ functions *in vivo* [122]. The liver organoids were formed by self-aggregation of iPSC-derived hepatic endoderm (HE), endothelial cells, and mesenchymal cells, and further matured upon transplantation into mice. Moreover, the liver organoids were vascularized and exhibited various hepatic characteristics which allowed to extend the life of liver damaged mice [122]. Additionally, Takebe and colleagues showed that liver organoids can entirely be generated from PSCs in a clinical relevant scale which highlights their great potential for regenerative medicine [123].

In addition to clinical applications, stem cell-derived hepatocytes are also an attractive model system in the field of drug development. Firstly, they are considered as a potential alternative for assessing the toxicity of bioactive drugs [124, 125]. So far, primary human hepatocytes (PHH) are the ‘gold standard’ in such assays, but they come with several limitations, such as they are scarce, show low proliferation rates *in vitro*, and quickly lose hepatic features in cell culture. To overcome

some of these limitations, 3D culture system generated from human ESCs [126] as well as human iPSCs [127] have been developed and applied to test toxicity of xenobiotics. The second promising application is drug screening (as reviewed in [128]). Jing and colleagues used the CRISPR/Cas9 technology to generate iPSCs that mimic the mitochondrial dysfunction of mtDNA depletion syndrome 3 [129]. The mutated iPSCs were plated in 96-well plates and differentiated into hepatocytes that were further used to assess the effects of 2,400 drugs on the metabolism. Based on this large-scale screen, 15 interesting hits were found of which one was confirmed to increase mitochondrial activity *in vivo*. In a similar study, Choi and colleagues employed patient-derived iPSCs with an alpha-1 antitrypsin deficiency [130]. They differentiated these cells into hepatocytes in a 96-well format and measured the effects of more than 3,100 drugs using an immunofluorescent readout.

Alternative strategies for hepatocyte differentiation

Oval cells are hepatic progenitors which supposedly origin from the canal of Hering, which is the connection between bile canaliculus and the bile ducts in the liver [131]. They are assumed to be not terminally differentiated hepatoblasts, which possess the capability to mature into hepatocytes or bile endothelial cells. Although they have some differentiation potency, they are most likely only progeny of stem cells as they are not able to self-renew [131, 132]. Since they are scarce and their isolation is challenging, the application for research is limited. Awan and colleagues showed that oval cells can be derived from adult bone marrow stem cells, which were able to take over some of the functions of co-cultured injured hepatocytes [133]. Furthermore, upon transplantation the oval-like cells were capable of rescuing some functions of fibrotic livers in mice, showing their potential for future applications [133].

Multipotent stem cells are an additional type of cell with the capacity to differentiate into hepatocytes. They can, for example, be non-invasively obtained from umbilical cord blood and showed the potential to differentiate into hepatocyte-like cells [134]. Lin and colleagues proved that cells from the umbilical cord can be transplanted into rats with liver damage and improve liver recovery [135]. They further demonstrated the production of crucial hepatic proteins from these cells, such as albumin and hepatocyte growth factor.

A third alternative technique for hepatocyte generation is transdifferentiation. During this process terminally differentiated somatic cells are converted into a different cell type without going through an intermediate pluripotent state. For example, Shen and colleagues showed that the treatment of pancreatic cells with dexamethasone led to the conversion of exocrine cells into hepatocytes [136]. The transcription factor C/EBPβ proved to be a key player in this transformation and is suspected to regulate the differentiation of the two endoderm-derived cell types. Hepatocytes can also be obtained from the non-parenchymal stellate liver cells. As mentioned in the previous subchapter (2.1 Liver: Physiology and embryonic development), the exact function of stellate cells is inconclusive, but during liver fibrosis they get activated and transform into myofibroblasts. Song and colleagues illustrated that through the overexpression of the transcription factors FOXA3, GATA4, HNF1A, and HNF4A, myofibroblasts could be differentiated into hepatocyte-like cells *in vitro* [137]. They further evaluated their findings in an *in vivo* mouse model by converting the fibrogenesis-inducing myofibroblasts into hepatocyte-like cells resulting in alleviated liver fibrosis. In 2011, two groups showed independently that mouse

fibroblasts can be directly reprogrammed into hepatocyte-like cells, which were able to restore liver functions of diseased mice. While Huang and colleagues overexpressed Gata4, Hnf1α, and Foxa3 [138], Sekiya and Suzuki used combinations of Hnf4α, Foxa1, Foxa2, and Foxa3 [139]. However, transdifferentiation was incomplete as immature markers, like Afp and CK19, were upregulated and some cytochrome P450 proteins were not expressed [138]. Furthermore, this resulted in only a partial rescue making additional optimizations inevitable.

In summary, different strategies to obtain hepatocytes *in vitro* exist and several studies demonstrated exciting results for applications in the clinics as well as for drug discovery. If limitations, such as immature hepatic characteristics, heterogeneous cell populations, or varying reproducibility between batches are improved, *in vitro* generated hepatocytes can be a promising alternative to primary hepatocytes.

3 Proteomics for studying developmental processes

When studying developmental processes, very often transcriptomics or low-throughput protein-based approaches are employed. For transcriptomics analysis, mRNA is extracted from cells and further amplified resulting in great sensitivity and sequencing depth. However, since mRNA is just the precursor of proteins and the correlation of mRNA to protein levels can diverge [140], such studies do not provide the full picture of biological processes in the cell. Hence, studying proteins, the executing molecules of the cell, can provide more insight by adding another level of information. For this type of analysis, very often methods like fluorescent activated cell sorting (FACS), immunostainings, or western plots are utilized. However, these approaches heavily rely on antibodies, which constrain the analysis in a way that it is low-troughput, requires specific antibodies, and prior hypotheses to select the proteins of interest. An alternative, which overcomes some of the limitations, is mass spectrometry-based proteomics. This method allows the indentification and quantification of global protein levels in a hypothesis-free way with great depth and sensitivity. Experiments of this kind can either be performed as 'bottom-up' or 'top-down' approaches. The latter injects intact proteins into the mass spectrometer, which leads to a high sequence coverage and a potential full protein characterization [141]. In contrast, for the ´bottom-up´ technology, proteins are chemically or enzymatically digested and mass-to-charge (m/z) ratios of the resulting peptides are acquired. While this methodology is more sensitive and provides a higher proteome coverage compared to 'top-down' measurements, the protein identification possessess some challenges. As the identification is inferred from peptides, and many peptides or peptide sequences are shared among proteins, the assignment of a peptide to one distinct protein can be ambiguous. Nevertheless, the advantages of 'bottom-up' proteomics prevail and technological developments during the last years yielded more than 17,000 identified proteins in the first draft of the human proteome [142, 143], which shows how powerful this method is and what impact it might have for future scientific research questions. Besides analysing the proteome, the high accuracy and sensitivity of mass spectrometers allows the measurement of PTMs of which many are connected to cell differentiation and development [144]. The following paragraphs provide details on classical 'bottom-up' workflows that have been applied throughout this work.

3.1 Bottom-up sample preparation

Before the sample can be introduced into the mass spectrometer, it undergoes several sample preparation steps, such as protein extraction, protein digestion, and the removal of interfering substances (I-Figure 3). Depending on the sample type and aim of the experiment, different preparation steps are performed and some of them are described in more detail with a special focus on the methods used in this work.

Protein extraction and digestion

In general, proteomic sample preparation starts with protein extraction, which can vary depending on the type of sample (e.g. tissues, body fluids, cell lines) and the downstream analysis that is performed. Besides methods that use mechanical force to break cells (e.g. bead mills), non-

mechanical methods can be applied (reviewed in [145]). An example for a cheap and straightforward approach, is chemical lysis. Detergents (e.g. sodium dodecyl sulphate (SDS)) are very commonly used chemicals, because they contain hydrophilic and hydrophobic moieties, allowing the detergent to integrate into the phospholipid bilayer of the cell, which causes the membrane to break. Furthermore, detergents disrupt non-covalent interactions resulting in protein denaturation [146]. On the downside, detergents impair enzymatic digestion, hamper chromatographic separation, and suppress the ionization [147, 148], which requires their removal during the sample preparation for mass spectrometry-based proteomics. Chaotropic reagents, such as urea or guanidine, are alternatives for chemical lysis. They interact with water molecules surrounding the cell membrane, which leads to a decrease in hydrophobicity and subsequently in its disruption. Similar to detergents, urea breaks non-covalent interactions causing protein denaturation [149].

After their successful extraction, proteins are digested for bottom-up proteomics (reviewed in [150]). For this, disulfide bridges are reduced and the free cysteins are alkylated. Besides chemical digestion using acids or other chemicals like cyanogen bromide, peptides can be generated enzymatically. Trypsin is a widely used protease, because of its high efficiency and sequence specificity cleaving peptide bonds c-terminally at lysine and arginine residues. As lysine and arginine are highly abundant in the human proteome, the resulting peptides consist on average of around 14 amino acids [151]. Additionally, the positive charges of the two basic amino acids make tryptic peptides ideal for mass spectrometry (MS) as it is based on the manipulation of charged ions. Although trypsin digestion can be considered the 'gold standard' for proteomics, other sequence specific proteases, such as Glu-C, Lys-C, Asp-N, or chymotrypsin, are alternatively used and have been proven advantageous in some cases [152, 153]. Besides the large selection of proteases, the setting in which proteins are digested can be varied according to the sample type. Traditionally, enzymatic digestion for bottom-up proteomics was performed in-solution or in-gel. While in-gel digestion is robust and capable of removing any interferring contaminants, it is very time-consuming and needs many manual handling steps [154, 155]. On the contrary, the in-solution protocol can easily be automated and is scalable, but interferring substance from the lysis buffer cannot easily be removed making further peptide cleanup necessary. This is especially of essence for workflows based on detergents, as their removal is tedious and usually leads to sample loss. While detergents often comprise multiple advantages, such as an increased yield of membrane proteins, they hamper digestion efficiency and impair the MS measurements. Filter-based methods like S-Trap [156] or the filter-aided sample preparation [157] facilitate the removal of SDS and provide a technology for protein digestion in a single spin column. Similar to this approach is the in-StageTip (iST) method [158], which combines all processing steps from cell lysis to the elution of purified peptides in a single pipette tip filled with C18 material. While this methodology enables to process minute amounts of samples [159], detergents cannot be removed. An additional alternative is the utilization of magnetic beads like the single-pot, solid-phase-enhanced, sample preparation (SP3, [160]) method or the protein aggregation capture approach [161]. Here, proteins are precipitated onto magnetic particles facilitating the removal of unwanted contaminants, including detergents, by washing steps before proteins are digested. This technology proved to be advantageous for low-input samples [159, 160] and is additionally easily automatable, which makes it interesting for high-throughput experiments [162]. As

outlined, several options for protein extraction and digestion exist, but they all have their advantages and disadvantages. A robust, scalable, inexpensive, automatable, straightforward, and universal method with minimal sample loss is yet to be found.

Peptide fractionation

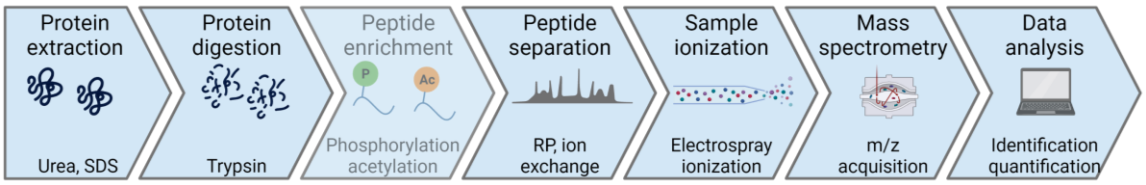
Digestion of the proteome results in a peptide mixture of tremendous complexity which makes data acquisition challenging. In order to decrease this complexity, additional sample fractionation strategies can be performed, which separate peptides based on their physicochemical properties. For this, usually high-performance liquid chromatography (HPLC) [163] or Stop-and-go-extraction tips (StageTips) [164] are used. In both methods the separation is based on the interaction of peptides between the mobile and the stationary phase (as reviewed in [165]). Reversed-phase chromatography is a fractionation method, which allows analyte separation based on hydrophobicity [166]. Here, the analyte is dissolved in the polar mobile phase and is separated via hydrophobic interactions with the non-polar stationary phase. This is a widely used chromatography approach in the field of proteomics, as it is highly orthogonal to the low pH reversed-phase chromatography commonly used for on-line fractionation preceeding the MS measurement. However, several other chromatographic methods, such as ion-exchange chromatography [167], hydrophilic interaction liquid chromatography [168], or a mixed mode [169, 170] can be applied as alternatives. Although thorough fractionation increases the number of protein and peptide identifications, this comes with the cost of increased measurement time.

Phosphopeptide enrichment

As many of the developmental signalling pathways are regulated by phosphorylation, the investigation of phosphorylation patterns and the inference of kinase activities are of great interest to elucidate biological function. While protein depth can be substantially increased by peptide fractionation, an additional enrichment step is required in order to achieve comprehensive phosphoproteome coverage, which is due to the substoichiometric abundance of phosphorylated peptides in nature (as reviewerd in [171]). Immobilized metal affinity chromatography (IMAC) is one widely utilized enrichment strategy [172]. Here, metal ions, such as Fe^{3+}, are chelated either with iminodiacetic acid or nitrilotriacetic acid and form a coordinative bond with the negatively charged phospho groups. As this binding step is performed at low pH, the unspecific binding of negatively charged carboxyl side chains is reduced. The phosphorylated peptides are further eluted by the addition of competing alkaline solvents or phosphate buffers [173]. The Fe^{3+}-IMAC enrichment can either be performed with beads [174] or in column format [175] and has recently been optimized for high-throughput and low input experiments on the AssayMap Bravo pipetting robot [176, 177]. In the last years a slightly modified setting using magnetic beads has been developed and is commercially available from ReSyn Biosciences (Pty) ltd (Edenvale, Gauteng, South Africa). Here, phosphopeptides are bound to metal(IV) ions, like Ti^{4+} or Zr^{4+}, which are chelated to phosphonate groups and connected via a linker to the magnetic microparticles. This setup has been proven to be especially attractive for small protein amounts and is in addition easily automatable enabling high-throughput enrichments [161, 178].

Acetylation enrichment

Besides phosphorylation, acetylation of lysine residues is another PTM capable of altering protein features like the regulation of enzymatic activity [179, 180]. In addition, histone acetylation affects chromatin accessibility and therefore gene transcription, which has been shown to comprise an important role for developmental processes [181-184]. However, studying acetylation is challenging due to its substoichiometric nature, which makes an enrichment step inevitable. In general, bead-conjugated antibodies targeting a specific acetyl-lysine motif are used for this purpose. Coupling this enrichment strategy with MS enables a global survey of acetylation levels demonstrating the widespread prevalence on a variety of different proteins [185, 186]. Despite the need of large input amounts, Svinkina and colleagues demonstrated that with an optimized proteomic sample preparation workflow more than 10,000 acetylated peptides can be identified [187].



I-Figure 3: Classical bottom-up proteomics workflow. Proteins are extracted from cells using lysis buffer and are further enzymatically digested to peptides. Optionally, peptides are enriched for phosphorylation or acetylation groups, for example. The peptides are separated on- or off-line using one- or multi-dimensional separation strategies. The resulting peptide mixture is ionized and injected into the mass spectrometer where m/z ratios are measured. For identification and quantification of peptides/proteins, the acquired mass spectra are matched against a protein database. Modified with permission from Steen and Mann [188] and created with BioRender.com.

3.2 Basics of bottom-up proteomics

Once sample preparation is finished, a clean peptide solution is injected into the mass spectrometer where m/z ratios are acquired. These m/z ratios can further be matched against an in-silico database, which leads to peptide identification and enables to infer protein information. Technological advances have generated a multitude of different online liquid chromatography (LC) systems, ionization methods, and mass analyzers. The next paragraphs will focus on the Orbitrap Fusion Lumos (Thermo Fisher) as this machine was solely used for this study (I-Figure 4).

Online fractionation

Before the sample is introduced into the mass spectrometer, peptides are separated using an on-line LC system. Reversed-phase HPLC is the most widely used separation technique, as it ensures high peak capacity and solvent compatibility [189, 190]. This approach is based on the direct interaction between hydrophobic peptide residues with the non-polar stationary phase, such as octadecyl alkane chains on silica beads. The analyte is dissolved in the acidic mobile phase, which puts a net positive charge on most tryptic peptides, a prerequisite for the generally applied positive MS mode. Moreover, it results in protonated hydroxyl groups of the silica beads, which

increases the peptide separation. One disadvantage of this approach is that polar peptide moieties are not retained by the stationary phase. To overcome this problem, ion-pairing reagents, such as formic acid (FA), are added to the mobile phase increasing the separation by indirect ion pairing effects [166, 191]. Peptides are eluted from the column by an increasing concentration of organic solvent in the mobile phase and further transferred to the mass spectrometer. Due to the high sensitivity and ionization efficiency, nanoflow LC has been widely used in the last years [192]. However, as the newly developed MS machines are more sensitive, micro-flow LC is gaining more popularity. At the expense of more sample material, this approach showed very promising results concerning throughput and robustness [193, 194].

Electrospray ionization

Peptides eluting from the reversed-phase column are passed through a thin capillary, the emitter, on their way into the mass spectrometer. Since the mass spectrometer acquires m/z ratios of ion, peptides have to be transferred into the gas phase. One mild and commonly used ionization method for this is the electrospray ionization, which efficiently produces very stable ions (as reviewed in [195]). While the high voltage applied between the emitter and the mass spectrometer pulls the charged molecules from the liquid to the counter electrode, the surface tension pulls the liquid back to the emitter in order to decrease the surface area. At a specific voltage the Taylor Cone is formed, which is a pointed cone where droplets are released towards the counter electrode. During the flight, solvents are evaporated until the repulsive forces of the remaining charges exceed the surface tension (Rayleigh limit) of the droplet. This results in the Coulomb fission causing the burst of the droplet and the formation of multiple nanodroplets containing charged peptides [196]. Although the principle of this effect was already described in 1964 [197], the exact mechanism of the final peptide ionization has not been resolved yet and can only be approximated by two models. The ion evaporation model postulates a decrease of droplet size due to solvent evaporation until the charged analyte can be expelled from the droplet [198, 199]. On the contrary, the charged residue model assumes that each nanodroplet contains only one analyte and upon complete solvent evaporated, the remaining charge is transferred to the peptide [200]. Ions enter the high-vacuum of the MS through a heated capillary, which further supports the complete solvent evaporation.

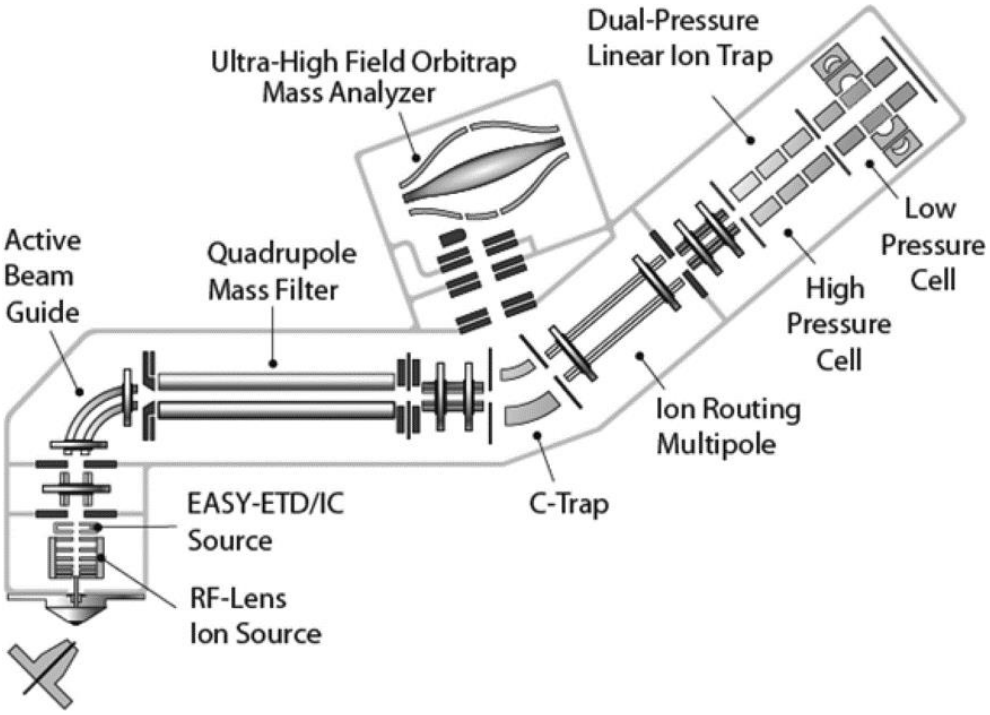
Precursor m/z acquisition

Once the ions are delivered into the MS, they are steered through electromagnetic fields on their path to the mass analyzers where their m/z values are determined. Several mass analyzers have been developed, each with different properties regarding mass resolution, mass accuracy, sensitivity, scan speed, dynamic range, and m/z range, as reviewed by Savaryn and colleagues [201]. Depending on the research question and sample type, specific mass analyzers are favoured and very often combined in one machine. The Fusion Lumos, which was used for this study, incorporates a quadrupole, a linear ion trap, and an orbitrap mass analyzer (I-Figure 4). The quadrupole consists of four parallel metal rods, where the two opposing rods are electrically connected. By setting different Alternating Current and Direct Current voltages, the trajectories of ions travelling through the quadrupole can be stabilized or destabilized [201]. For this reason the quadrupole is very often used as a prefilter, which allows the selection of ions with specific

m/z ratios. The linear ion trap follows a very similar principle, but is further capable of storing and fragmenting ions as well as determining m/z values [202]. It also consists of four rod-shaped electrodes of which two have exit slits for ions to leave the ion trap and enter the detector. The third mass analyzer of the Fusion Lumos is an orbitrap, which consists of two barrel-like outer electrodes and a coaxial inner spindle-like electrode [203]. Before ions enter the orbitrap, they are collected in the C-Trap and then injected as ion packages where they oscillate around the inner electrode. As the axial movement is proportional to the m/z value, the frequency of this oscillation can be recorded as an image current and further used to deduce m/z values by Fourier Transformation [204].

Tandem mass spectrometry

Although the determined m/z values from the MS1 scan are very accurate, this information alone is not enough to unambiguously identify peptides, because peptides with a different amino acid sequence but the same mass cannot be distinguished. In order to overcome this problem, two consecutive MS spectra are acquired in the so-called tandem MS. For mass spectrometers operated in data-dependent acquisition (DDA) an MS1 scan measures the m/z values and intensities of intact peptide percusors ions first. Then, the mass spectrometer decides automatically in real-time which precursors are selected for further fragmentation and MS2 spectrum acquisition. For a predefined number or time span, MS2 spectra are acquired based on their MS1 abundance, which limits the reproducibility due to the stochastic nature of precursor selection for fragmentation. In contrast, for data-independent acquisition (DIA) m/z windows are grouped and subsequently fragment ions are analyzed independent of the MS1 scan. During the fragmentation process, precursors collide with inert gas (e.g. helium, nitrogen) breaking the peptide bonds. The fragmented precursors are further used to record MS2 spectra which reveal the amino acid sequence. Combining the m/z values from the precursors with their amino acid sequence allows a very accurate peptide identification.



I-Figure 4: Schematic of the Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (reprinted with permission from [205]. Copyright (2013) American Chemical Society).

3.3 Mass spectrometry data analysis

Depending on the acquisition mode, the measurement time and the mass analyzer used for data generation, usually thousands of MS1 and MS2 spectra are recorded. In order to retrieve peptide information from these spectra, search engines with specific algorithms are used. Furthermore, this peptide information is used to identify and quantify proteins.

Protein identification

As discussed above, peptide identification solely based on MS1 spectra is ambiguous. Hence, acquiring MS2 spectra are inevitable to gain (partial) amino acid sequences as an additional layer of information for peptide identification. One method to deduce peptide information from MS2 spectra is *de novo* sequencing, which uses the delta mass between fragment ions to determine the discrete amino acid sequence [206]. While this approach is suitable for detecting previously unknown peptides, it requires the complete amino acid sequence, which is not achieved in most mass spectrometry-based experiments. In addition, this method is restricted to less complex samples, because a widely accepted approach for validating results from large-scale datasets is still lacking [207, 208]. However, the validation is an essential step for proteomics experiments, as usually thousands of spectra are recorded which have to be controlled with respect to their correct identification. Some of these challenges are overcome by an alternative method, which is based on database matching performed by softwares like Maxquant [209]. Here, the implemented search engine Andromeda [210] compares the obtained fragment spectra to an *in silico* digested protein database of theoretical spectra. For this, the noise of fragment masses is reduced, the charge state is deconvoluted, and de-isotoped. Furthermore, search parameters, such as the

applied protease or the mass tolerance, are specified to decrease the search space. Then, the search engines calculate a score describing the probability of a true identification and usually the highest scoring hit is reported. The correct annotation of the peptide spectrum match (PSM) is crucial for bottom-up proteomics, as protein identification is inferred from the experimentally acquired peptides. To control for correct identification, the experimental spectra are additionally searched against a database of reversed or scambled peptide sequences in a target-decoy approach [211]. Each hit to this decoy database is false by definition and enables the calculation of a false discovery rate (FDR) which is very often set to 1% to keep the number of wrong identifications low. As a peptide sequence cannot always be uniquely assigned to just one protein, but rather multiple, MaxQuant reports them as razor peptides. Due to this ambiguous identification, usually protein groups instead of proteins are reported, which demonstrates the protein-inference problem of bottom-up proteomics. Another drawback of this method for identification is that only sequenced organisms can be used, because a comprehensive database is required for sufficient peptide matching.

While protein identification is usually based on several peptides, in many cases one or only a few MS2 spectra are used to identify a phosphopeptide and to localize the P-site. As in theory each serine, threonine, or tyrosine can be modified, the search space is vastly elevated, which increases the possibility for false identification as well as localization. Hence, phosphorylation localization tools have been developed to this end [212, 213]. In addition, very often the data analysis is limited to so-called class I P-sites, which possess a localization probability >0.75 [212].

Protein quantification

In addition to protein identification, protein abundance information is usually of interest for research questions. Depending on the experiment and sample type, several different quantification methods exist and can be classified into label-free or label-based methods. In label-free approaches protein identification is based on MS2 spectra, while MS1 spectra are facilitated for quantification [214]. One quantification method is spectral counting, where the number of peptides for each protein is summed up and used to derive protein abundance. A more frequently used method uses the peak intensity of peptides, which correlates with their abundance [215]. However, a big drawback of label-free methods is that all samples from one experiment are only combined at the last step, the data processing. Hence, sample preparation and MS measurement has to be performed for each sample individually, which is not only time consuming but also error prone. In addition, due to the stochastic nature of DDA approaches missing values across many samples are unavoidable. In order to increase reproducibility and throughput, several techniques of multiplexing samples have been developed (reviewed in [216]). While some strategies are based on the incorporation of stable isotope labels with amino acids in cell culture, other approaches use chemical labeling, such as tandem mass tags (TMT). The latter technique uses succinimide chemistry to label the amino termini and lysine residues of peptides and thus enable the parallel quantification of up to 18 samples within a single experiment [217]. As TMT reagents have the same nominal mass, the MS1 spectra complexity is not increased. Upon precursor selection and subsequent fragmentation, the TMT reporter ions are cleaved off allowing for relative quantification as well as peptide identification based on MS2 scans. However, the precursor selection suffers from the coisolation of coeluting precursors and thereby leads to ratio

distortion of the TMT reporter ions [218, 219]. In order to overcome this problem, several ideas have been discussed in the past, such as decreased isolation window width [220], gas-phase purification [221], or utilization of complement reporter ions [222]. In this work, the quantification accuracy was increased by an additional MS3 scan, which can be performed by the Tribrid Fusion Lumos machine and has been shown to reduce ratio distortion of coisolating precursors [223]. Here, an MS2 spectrum was acquired for peptide identification and the most abundant fragments were further used for another round of fragmentation. These generated fragments were then recorded in an MS3 scan for quantification of reporter ions. Although this additional round of isolation leads to ratios closer to the reality, it suffers from decreased sensitivity. The MultiNotch MS3 measurements counteracts this pitfall with the synchronous precursor selection (SPS) of several of the most abundant MS2 fragments, which increases the MS3 intensity and thereby the sensitivity [224].

4 Objectives and Outline

Mass spectrometry-based proteomics is the most widely used technique for analyzing protein levels of complex samples. However, relatively large protein quantities are required to achieve comprehensive measurement depth, preventing its usage in some applications. While developments on the mass spectrometer-side in recent years have led to enhanced sensitivity of the devices itself, the sample preparation turned out to be one of the major bottlenecks for analyzing small protein quantities. Therefore, methods that can cope with minute protein quantities are warranted. The study of *in vitro* cell differentiation is an area of research that could benefit from such improved experimental conditions on quantity-limited samples, as differentiation experiments are cost and labor intensive. Moreover, few comprehensive and global proteomic characterizations of directed cell differentiations have been performed. Thus, the aim of this work was to exploit the insights that can be gained from the temporal dynamics of global protein levels and their phosphorylation and acetylation during hepatocyte differentiation. At the beginning, different sample preparation methods were bechmarked and their suitability for low input material was assessed. This evaluation revealed the SP3 bead approach as a versatile sample preparation method that can easily be implemented without the necessity of expensive equipment (Chapter II). In a second part, deep proteome profiling was utilized to gain insights in the differentiation process of pluripotent stem cell-derived hepatocytes (Chapter III). The high temporal resolution of this experiment enabled the identification of novel stage-specific marker proteins and elucidated the molecular changes occuring during the differentiation process. In addition, the study of phosphorylation and acetylation dynamics allowed the inference of kinases and deacetylase activity as key drivers of various signaling pathways. Finally, the same experimental setup was adopted to facilitate the comparison of *in vitro* generated hepatocytes to fetal and mature liver samples for the characterization of current differentiation protocols (Chapter IV).

Chapter II: Optimizing sample preparation workflows for proteomics

The following chapter includes data generated from Karl Kristian Krull during his research internship “Influence of surface modification in SP3 technology for functional proteome analysis” and his Master Thesis “Phosphoproteomics for low input amount using paramagnetic IMAC beads”, which he conducted under the author’s supervision at the Chair of Proteomics and Bioanalytics at the Technical University of Munich.

1 Summary

Albeit mass spectrometry-based proteomics has been successfully used to analyse a variety of different biological samples, an extraction and sample preparation method that can be universally applied is missing. In this regard, several different approaches for sample processing were benchmarked with a special emphasis on methods suitable for small quantities as well as phosphopeptide enrichments. Optimizing the in-StageTip workflow resulted in the identification of more than 10,000 peptides from 2,000 human cells, which is 5 times more than with the standard protocol. Only the use of magnetic carboxyl-coated SP3 showed better performance with about 40% higher number of peptide identifications. Additionally, the SP3 approach was compatible with TMT labeling, while decreasing processing time and handling steps compared to the standard in-solution workflow. Of note, alternative magnetic microparticles, such as amine and hydroxyl-modified beads, performed similar as SP3. In contrast, beads based on a HILIC-like protein binding decreased the number of identified peptides, had a bias towards more hydrophilic peptides, and showed lower reproducibility. In addition, protocols for phosphopeptide enrichments using magnetic Ti-IMAC and Zr-IMAC beads were optimized leading to the detection of 20% more phosphorylation sites than in the manufacturer's protocol. Based on these results a novel workflow, called OnePot, was developed which facilitates protein clean-up, digestion, and phosphopeptide enrichment in one reaction tube using magnetic IMAC beads. Compared to the Agilent AssayMAP® Bravo system the OnePot approach performed particularly well for enrichments of low input amounts.

2 Introduction

In bottom-up proteomics, protein abundances are indirectly inferred from the measurement of peptide solutions. Despite of multiple different methods for obtaining clean peptide solutions from protein extracts, their applicability depends on the experimental setup as well as the sample type. For example, different extraction buffers and additives can hamper enzymatic digestion, can interfere with fractionation columns, and can impair peptide labeling and enrichment. Hence, lysis reagents often need to be removed or diluted below a threshold to minimize their negative impact. In-solution digestion is a classical workflow for generating peptides, but is susceptible to contaminants and requires additional desalting steps which are time consuming and result in sample loss. Especially the latter is crucial for sample-limited experiments. In this regard, Kulak and colleagues designed an iST approach, which allows cell lysis, protein digestion, and desalting in a single reaction vessel [158]. While this approach decreases sample loss, it is not capable of removing detergents and is therefore not compatible with several lysis buffers. Another approach that has gained increasing interest in recent years is SP3 [160]. This method is based on the precipitation of proteins to magnetic carboxyl-coated beads using high concentrations of organic solvents. Once proteins are precipitated, detergents and other interfering substances can be removed. As the sample loss is small, this approach offers great potential for sample-limited experiments [159, 160, 225]. ReSyn Biosciences (Pty) ltd offers a commercially available alternative to SP3 beads, which feature a variety of different surface modifications. Compared to SP3 beads, they provide an increased surface area due to their bigger diameter as well as their hyper-porous characteristics which supposedly enhances their capacity of protein capturing.

Apart from the variety of workflows targeting the fullproteome, multiple different approaches for phosphopeptide enrichment exist. About 30% of all proteins are estimated to be phosphorylated [226], which alters their biological characteristics, including activity, subcellular localization, or protein-protein interactions. As this PTM suffers from low stoichiometry and abundance, an enrichment step is in general necessary. One commonly applied strategy is IMAC, where chelated metal ions form coordinative bonds with the phosphate groups of peptides. Traditionally, metal(III) cations such as Fe^{3+} or Ga^{3+} have been used. However, several studies have introduced metal(IV) cations as alternatives with potentially superior enrichment selectivity and enhanced phosphopeptide binding [227-229]. This led to the development of the noncovalent immobilization of Ti^{4+} and Zr^{4+} cations to magnetic beads via phosphonate linkers. Nevertheless, a big drawback of phosphoenrichment is the requirement of high sample amount. To this end, several platforms and approaches for low input quantities were designed over the last years. For example, the EasyPhos approach has enabled the identification of more than 20,000 phosphopeptides from only 200 µg of protein by consecutively performing protein digestion and phospho enrichment in a 96-well-plate format [230, 231]. Chen and colleagues have published the column-based Phospho-SISPROT method, which utilizes spintips for protein digestion, IMAC-enrichment, and subsequent desalting allowing to process samples in the low µg range [232, 233]. Furthermore, the liquid handling robot AssayMAP® Bravo Agilent offers a high-throughput system with great potential for minimizing sample quantities [177].

The following experiments were designed to optimize and evaluate the commonly applied in-solution, iST, and SP3 approaches for low amounts of starting material. Furthermore, magnetic

beads functionalized with several different chemical groups were benchmarked. Lastly, the enrichment of phosphorylated peptides using magnetic Ti^{4+} and Zr^{4+} IMAC-beads was tested. Utilizing the IMAC beads for protein precipitation, digestion, and phospho enrichment, led to the development of an integrated workflow performed in a single tube which revealed promising results for sample-limited experiments.

3 Material and methods

Protein extraction

Cells cultured in suspension were transferred into a falcon tube and centrifuged. The medium was discarded and the cell pellet was washed twice with phosphate-buffered saline (PBS) (w/o $CaCl_2$/$MgCl_2$). Then, the cell pellet was lysed with lysis buffer containing 8 M urea, 40 mM Tris/HCl (pH 7.6), 1x EDTA-free protease inhibitor (cOmplete™, Roche), and 1x phosphatase inhibitor mix (prepared in-house according to the Phosphatase Inhibitor 1, 2, and 3 from Sigma). The lysate was sonicated for 15 cycles of 30 sec ON/OFF using a Bioruptor® (diagenode) and the protein concentration was determined with the Pierce™ Coomassie assay (Thermo Scientific™). Alternatively to the urea lysis, washed cells were lysed with 2% SDS in 50 mM Tris/HCl (pH 8.5) and subsequently heated at 95°C for 10 min. In order to hydrolyse nucleic acids, a final concentration of 2% trifluoroacetic acid (TFA) was added to the boiling lysate for 1 min. The pH was quenched back using 3 M Tris until a neutral pH was reached and the lysate was additionally sonicated for 15 cycles of 30 sec ON/OFF using a Bioruptor® (diagenode). Finally, the protein concentration was determined using the Pierce™ BCA protein assay (Thermo Scientific™) and the lysate was stored at -80°C until further use.

In-solution workflow

Urea cell lysate was thawed and the respective protein amount was transferred to an Eppendorf tube. Next, proteins were reduced with 10 mM DTT for 45 min at 37°C and subsequently alkylated using CAA for 30 min at room temperature (RT). For an efficient digestion, the sample was diluted to a urea concentration below 1.6 M using 40 mM Tris/HCl (pH 7.6).Trypsin was added at a 1:50 enzyme-to-protein ratio and incubated at 37°C overnight. On the next day, the digestion was stopped by adding FA to a final concentration of 1%.

The digested peptides were further transferred onto self-packed StageTips for desalting as described previously [164]. Shortly, C18 material (Octadecyl Extraction Disks, 3M Empore™) was packed into a 200 µL pipette tip and equilibrated. The acidified peptides (pH 2-3) were loaded and contaminants were removed with two washes of 0.1% FA. Peptides were eluted with 50% acetonitrile (ACN) in 0.1% FA and dried down. For input amounts >100 µg protein, acidified peptides were loaded onto 50 mg SepPak columns (Water Corp.) and desalted. Contaminants were washed away with 0.1% FA, peptides were eluted with 0.1% FA in 50% ACN, and dried down. Depending on the experimental setup, peptides were further labeled with TMT (Thermo Scientific™). For labeling, the desalted peptides were reconstituted in 20 µL of 50 mM HEPES buffer (pH 8.5) as described previously [234]. Then, 5 µL 11.6 mM TMT reagent was added and incubated for 1 h at RT shaking at 400 rpm. The reaction was quenched using 2 µL of 5% hydroxylamine and subsequently all TMT channels were pooled. The labeled samples were further desalted using SepPak cartridges as described above and dried down.

iST workflow

The sample preparation workflow was based on the previously published study by Kulak and colleagues [158]. As described in the results section some modifications were made in order to optimize the workflow for low sample quantities. First, StageTips were constructed as described

in the previous section by packing 5 C18 disks in a pipette tip. While the original protocol recommended the use of 200 µL pipette tips, enhanced performance was observed utilizing 10 µL pipette tips in combination with lowering the digestion volumes. The StageTip was primed with 90 µL ACN followed by equilibration with 90 µL 100 mM Tris (pH 8.5). The column was heat-sealed and covered with 5 µL 100 mM Tris (pH 8.5) to avoid complete drying. Then, cells were washed twice with PBS (w/o $CaCl_2$/$MgCl_2$), counted in a Neubauer champer, and applied onto the StageTip (volume should be adjusted to 5 µL). For cell lysis, 10 µL urea buffer (8 M urea, 100 mM Tris, pH 8.5) was added and incubated for 10 min on ice. Proteins were reduced with a final concentration of 10 mM DTT for 45 min at 37°C and alkylated with 55 mM CAA for 30 min at RT. The urea concentration was decreased below 1.6 M using 25 mM Tris (pH 8.5) before trypsin was added with an enzyme-to-protein ratio of 1:50 and incubated at 37°C and 600 rpm overnight. On the next day, the sample was acidified with FA (1% final concentration) and the heat-sealed pipette tip was cut open. The digested peptides were loaded slowly to the C18 material by centrifugation, washed with 0.1% FA, and subsequently eluted with 0.1% FA in 60% ACN.

SP3 workflow

The following workflow was modified from the original protocol of Hughes and colleagues [160, 235]. Magnetic hydrophilic (GE45152105050250) and hydrophobic (GE65152105050250) carboxylate-modified Sera-Mag™ SpeedBeads (50 mg/ml) were removed from the fridge and kept at RT for 10 min. The beads were vortexed and mixed with a 1:1 ratio. They were washed 3 times with water using a magnetic rack for immobilization. Before use, bead pellet was reconstituted with water to a final concentration of 100 mg/ml. Urea lysate (chapter 4.1) or SDS lysate (chapter 4.2), respectively, was thawed and the desired protein amount was aliquoted into reaction tubes. For up to 50 µg protein lysate 2µl of the prepared bead stock was transferred to the lysate, whereas 10 µl of the bead stock was applied for protein amounts above 50 µg. Next, ACN was added for protein aggregation to a final concentration of 75% and incubated for 20 min shaking at 800 rpm. The beads were separated using a magnetic rack and the supernatant was discarded. Beads were washed twice with 80% ethanol and once with 100% ACN to remove contaminants. Subsequently, digestion buffer (50 mM Tris/HCl, pH 8.5 or pH 7.6, 2 mM $CaCl_2$) and DTT (10 mM final concentration) were added and incubated for 45 min at 37°C. Afterwards cysteine residues were alkylated with 55 mM CAA for 30 min at RT in the dark. For the overnight digestion, trypsin was added in a 1:50 enzyme-to-protein ratio and incubated at 37°C shaking at 800 rpm. On the next day, beads were immobilized on the magnetic rack and the supernatant containing the digested peptides were transferred to a new vial. Beads were washed once with 1% TFA in water and the supernatant was pooled with the previous supernatant. Peptides were stored at -80°C until further use.

For TMT labeling the following modifications were introduced to the protocol. After reduction and alkylation, proteins were reaggregated using ACN (final concentration of 75%). The beads were washed again twice with 80% ethanol and once with 100% ACN to remove the acidic CAA. Next, beads were reconstituted in 100 mM HEPES buffer (pH 8.5) and trypsin was added with a 1:50 enzyme-to-protein ratio. On the next day, the supernatant was removed and dried down. The dried peptides were further reconstituted in 20 µl water (HEPES concentration should be above 40 mM) and the labeling was initiated by adding 5 µl of 11.6 mM TMT reagent as reported

previously [234]. After incubating the samples for 1 h at RT and 400 rpm, the reaction was terminated using 2 µl of 5% hydroxylamine. Subsequently, all TMT channels were pooled and further desalted using SepPak cartridges as described above.

Protein aggregation capture workflow

Samples were prepared according to the manufacturer's protocol (ReSyn Biosciences). In short, magnetic beads were removed from the fridge and kept at RT for 10 min before they were transferred to an Eppendorf tube. Beads were washed twice with 70% ACN and reconstituted with water to a final concentration of 20 µg/µl. SDS lysate was thawed and combined with the bead stock with a 5:1 bead-to-protein ratio. For protein aggregation, ACN was added to a final concentration of 75% and incubated for 20 min at RT shaking at 600 rpm. The subsequent washing steps were performed without removing the reaction tubes from the magnetic rack in order to keep the bead pellet intact. Beads were washed twice with 95% ACN followed by two washes with 70% ethanol. Next, beads were reconstituted in digestion buffer (50 mM Tris/HCl, pH 8.5, 2 mM $CaCl_2$), reduced with 10 mM DTT for 45 min at 37°C, and alkylated with 55 mM CAA for 30 min at RT. Trypsin was added with an enzyme-to-protein ratio of 1:50 and incubated overnight at 37°C and 800 rpm. On the following day, beads were separated and the supernatant was transferred to a new reaction vessel. Remaining peptides were eluted from the beads with 1% TFA and pooled with the previous supernatant. Peptides were stored at -80°C until further use.

HILIC workflow

Samples were prepared according to the manufacturer's protocol (ReSyn Biosciences). Briefly, beads were removed from the fridge and kept at RT for 10 min. Next, beads were transferred to a reaction tube, washed twice with equilibration buffer (15% ACN, 100 mM NH_4Ac, pH 4.5), and reconstituted in 2x binding buffer (30% ACN, 200 mM NH_4Ac, pH 4.5) to a final concentration of 20 µg/µl. The bead stock was subsequently mixed with the SDS lysate to obtain a 10:1 bead-to-protein ratio. An equivalent volume of binding buffer was added to the sample to reach a final volume of 50 µl. The solution was incubated at RT for 20 min shaking with 600 rpm. Then, the beads were washed twice with 95% ACN and reconstituted in digestion buffer (50 mM Tris/HCl, pH 8.5, 2 mM $CaCl_2$). Cysteine residues were reduced and alkylated with DTT and CAA, respectively, as described above and trypsin was added to a 1:50 enzyme-to-protein ratio. The digestion was performed overnight at 37°C. On the following day, beads were immobilized on a magnetic rack and peptides were stored at -80°C until further use.

StageTip fractionation

The high pH fractionation was conducted as described previously [236]. In short, peptides in solution were acidifed (pH 2-3) and loaded onto self-packed StageTips like mentioned above. Peptides were washed once with 0.1% FA before 50 µl of 25 mM ammonium formate (pH 10) was added. The flow-through was reapplied and then transferred to a new vial or plate as this was pooled with one of the fractions later. Next, peptides were step-wise eluted with 5%, 10%, 15%, 17.5%, and 50% ACN in 25 mM ammonium formate. The 5% and 50% fractions as well as the 17.5% fraction and the flow-through were combined. All 6 fractions were frozen and dried down. For dried samples that were not in solution, the protocol was similar. The only difference was that the

dried peptides were directly resuspended in 25 mM ammonium formate and loaded onto StageTips.

Phosphoenrichment using magnetic IMAC-beads

The enrichment was performed according to the manufacturer's protocol (ReSyn Biosciences) with some minor modifications. Briefly, Ti-IMAC and Zr-IMAC beads were removed from the fridge and kept at RT for 10 min. Depending on the experiment, equal amounts of beads were mixed or used separately, respectively. The magnetic beads were transferred to a reaction tube and equilibrated three times for 1 min with loading buffer (80% ACN, 5% TFA). The desalted peptides were reconstituted in 100 µl loading buffer, centrifuged for 5 min at 10,000 xg, and applied to the prepared bead solution. A bead-to-peptide ratio of 2:1 was used for quantities between 100 µg and 200 µg, while a 4:1 and 8:1 ratio was applied for 50 µg and 25 µg, respectively. This is half the bead amount recommended by the manufacturer, but this reduction revealed enhanced performance. For phosphopeptide binding, samples were incubated for 30 min at RT and 1,000 rpm. The beads were immobilized with a magnetic rack and the unbound fraction was discarded or further analyzed as the fullproteome fraction. Unbound and unspecifically bound sample was removed by washing once with loading buffer followed by two consecutive washes with wash buffer 1 (80% ACN, 1% TFA) and wash buffer 2 (10% ACN, 0.2% TFA). Subsequently, phosphopeptides were eluted by adding 100 µl of 1% NH_4OH and shaking for 10 min at 800 rpm. This step was repeated two more times to ensure complete elution. The three fractions were pooled and combined with 50 µl of neutralization buffer (50% FA).

OnePot workflow

Ti-IMAC and Zr-IMAC beads were mixed with equal amounts and washed three times with water to remove the storage solution. Cell lysate was combined with the beads and ACN was added to a final concentration of 75% for protein aggregation. The sample was incubated for 20 min at RT shaking with 800 rpm and subsequently separated on a magnetic rack. The unbound fraction was removed and the beads were washed twice with 80% ethanol and once with 100% ACN. Next, digestion buffer (25 mM TEAB, pH 8, 2 mM $CaCl_2$) was added, proteins were reduced with 10 mM DTT for 45 min at 37°C, and alkylated with 55 mM CAA for 30 min at RT. For protein digestion, trypsin was added to an enzyme-to-protein ratio of 50:1 and, depending on the experiment, incubated overnight or for 3 h at 37°C. For the subsequent phosphopeptide enrichment, 160 µl binding buffer (90% ACN, 6% TFA) was added to the 20 µl of digest and incubated for 45 min at RT and 800 rpm. Then, the beads were separated on the magnetic rack and the unbound fraction was discarded or analysed as the fullproteome. To remove unspecifically bound peptides, beads were washed with loading buffer (80% ACN, 5% TFA) followed by two washing steps with wash buffer 1 (80% ACN, 1% TFA) and wash buffer 2 (10% ACN, 0.2% TFA). Next, phosphopeptides were eluted by adding 100 µl of 1% NH_4OH for 10 min at RT and 800 rpm. This step was repeated two more times and the fractions were pooled together with 50 µl of neutralization buffer (50% FA). Samples were stored at -80°C until further use.

Phosphoenrichment using the Agilent AssayMAP® BRAVO platform

The standard programs from the AssayMAP® for cartridge charging (IMAC Cartridge Customization v1.0) and enrichment (Phosphopeptide Enrichment v2.0 App) were performed as recommended by the manufacturer. In short, AssayMAP® Fe(III)-NTA cartridges were stripped using 100 mM EDTA, washed with 0.1% TFA, and charged using 50 mM $FeCl_3$ with 100 mM acetic acid. The freshly charged cartridges were further primed with 80% ACN in 0.1% TFA and subsequently used. Desalted peptides were reconstituted in 200 µl loading buffer (30% ACN, 0.1% TFA) and loaded onto the cartridges. After washes with 0.1% TFA and 0.1% TFA in 99.9% ACN, phosphopeptides were eluted with 1% NH_4OH. Peptides were dried down and stored for further processing.

Data acquisition and processing

A nanoflow LC-MS setup was used for data acquisition by coupling a Dionex Ultimate 3000 UHPLC+ system to a Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific). Peptides were loaded onto an in-house packed trap column (75 µm x 2 cm, 5 µm C18 resin; Reprosil PUR AQ, Dr. Maisch) and washed for 10 min with 0.1% FA and 5% DMSO. Subsequently, peptides were separated using an analytical column (75 µm x 40 cm, packed in-house with 3 µm C18 resin; Reprosil PUR AQ) with a flow rate of 300 nl/min and an increasing ACN gradient. Measurements were performed in DDA and positive ionization mode.

The label-free fullproteome samples were acquired with a linear 50 min gradient from 4% to 32% LC buffer B (0.1% FA, 5% DMSO in ACN) in LC buffer A (0.1% FA, 5% DMSO in MS-grade water). The MS1 scan were recorded with 60,000 resolution in the orbitrap. The automatic gain control (AGC) target was set to 3e5 ions and the maximum injection time (maxIT) to 25 sec. The top 20 MS1 precursors were fragmented using higher-energy collisional dissociation (HCD) and MS2 spectra were subsequently acquired with 15,000 resolution in the orbitrap. Therefore, the AGC target was set to 1e5 charges and the maxIT to 25 sec.

The label-free phosphoproteome was recorded with a two-step 80 min gradient. While during the first 50 min a linear gradient from 4-18% LC buffer B was applied, for the last 30 min a linear gradient from 15-27% buffer B was used. The MS2 method was similar to the label-free fullproteome method, except that the resolution for the MS2 scan was 30,000 and the maxIT was set to 120 sec.

For the TMT6-plex labeled full proteome, a linear 50 min gradient from 8% to 34% buffer B (0.1% FA, 5% DMSO in ACN) in LC buffer A (0.1% FA, 5% DMSO in MS-grade water) was used. Full scan MS1 spectra were recorded at 60,000 resolution and a scan range from 360-1300 m/z in the orbitrap. The AGC target was set to 4e5 charges and a maxIT of 20 ms was used. Isolated precursor from the MS1 scan were fragmented via collision-induced dissociation (CID) and acquired with 15,000 resolution in the orbitrap. The AGC target was set to 5e4 charges and the maxIT to 22 ms. The MS3 spectra were obtained via SPS-MS3, which simultaneously selects 10 MS2 fragments that are further fragmented via HCD and read out in the orbitrap with 15,000 resolution. For this, the AGC target was set to 1e5 charges and the maxIT to 50 ms. A few minor changes were made for the TMT11-plex labeled FP samples. MS2 spectra were acquired in the ion trap (rapid mode) with an AGC target of 2e4 charges and a maxIT of 60 ms. SPS-MS3 spectra were recorded with 50,000 resolution, an AGC target of 1.2e5 charges and a maxIT of 120 ms.

The TMT-labeled phoshopeptides were acquired with a linear 80 min gradient from 4% to 32% LC buffer B. MS1 spectra were recorded with 60,000 resolution in the orbitrap, an AGC target of 4e5 charges, and maxIT of 20 ms (TMT6 samples) or 50 ms (TMT11 samples), respectively. MS1 precursors were isolated and fragmented via CID for subsequent MS2 spectra acquisition in the orbitrap. For this, 15,000 resolution, and AGC target of 5e4 charges and a maxIT of 22 ms was used. The following MS3 scan was recorded at 15,000 resolution in the orbitrap for the TMT6-labeled samples and with 50,000 resolution for the TMT11-labled samples.

Database searching

Raw files were searched with the Maxquant software with its search Engine Andromeda [210, 237] (version [1.6.2.3]) against the UniProtKB human reference list (downloaded 22.07.2013). Default setting were applied, unless stated otherwise. The enzyme trypsin was specified as protease allowing for up to two missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification, while oxidation of methionine, and N-terminal protein acetylation were defined as variable modifications. In addition, serine, threonine, and tyrosine phosphorylation was set as variable modification for the phosphoproteome. A target-decoy approach was used to adjust the data to 1% PSM and protein FDR. For the TMT-labeled samples, MS3-based reporter ion quantification was enabled and the corresponding TMT correction factors were added.

Data processing

Information about the peptide IDs and quantification were obtained from the summary.txt and evidence.txt, which are both output files from the Maxquant software (version [1.6.2.3]). Entries with reverse hits or potential contaminants were excluded from the analyses. Protein information was based on the proteingroups.txt from which 'only identified by site', reverse hits, and potential contaminants entries were removed.

Phosphosites from the phosphoproteome analyses were retrieved from the phosphosite.txt. Again, reverse hits and potential contaminants were excluded and entries with a localization probability <0.75 were removed. To calculate the phosphopeptide selectivity, reverse hits and potential contaminants were removed from the evidence.txt and the ratio of phosphorylated/unphosphorylated entries was calculated.

For data processing and visualization, mostly Microsoft Excel, Perseus software suite [238, 239] (v.1.6.2.3), GraphPad Prism 5, and RStudio (version [4.0.2]) were used.

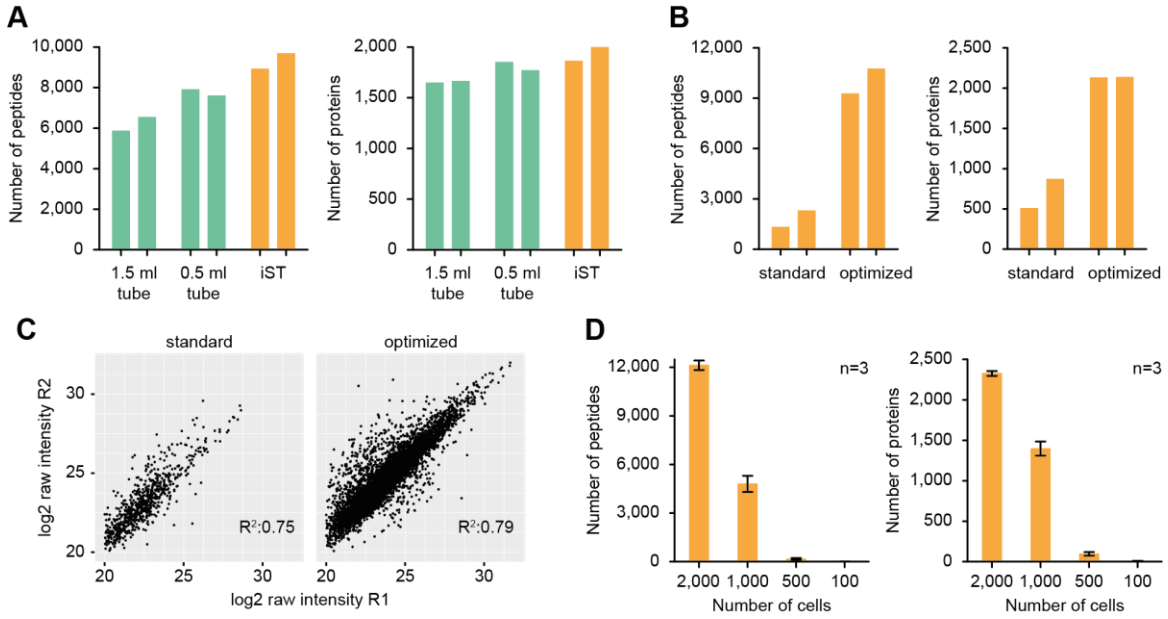
4 Results and discussion

A multitude of different sample preparation methods in the field of (phospho)proteomics have been developed over the last years. In this chapter several approaches were evaluated and optimized with a special emphasis on low input quantities.

4.1 Optimizing sample preparation for low input amount and TMT-labeling

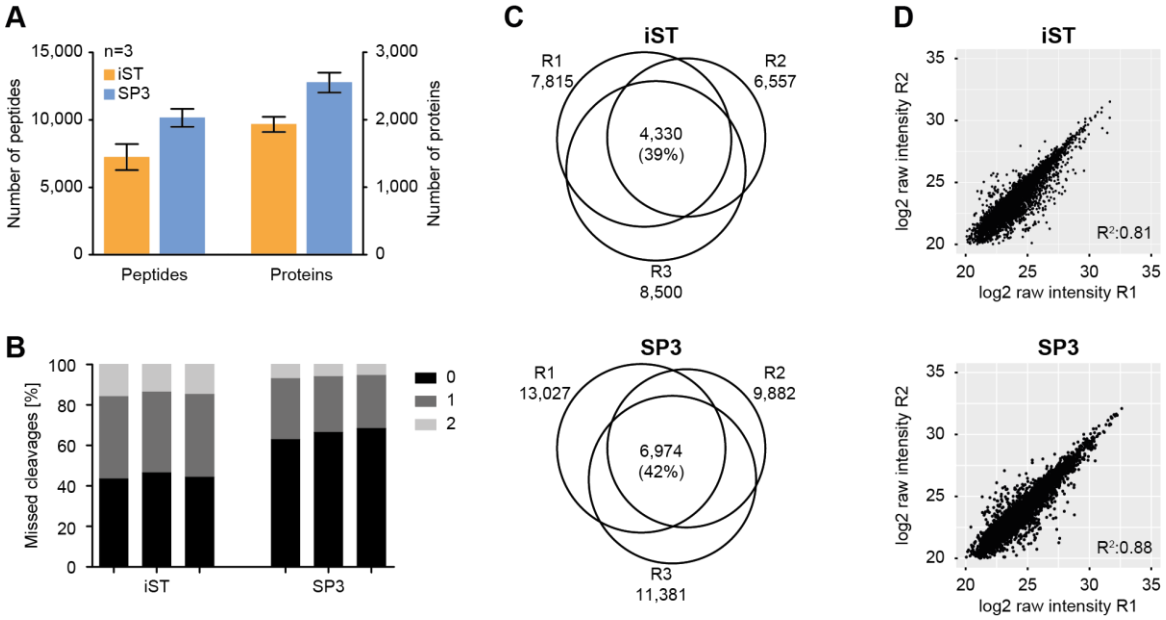
At first, the iST approach was benchmarked against an in-solution workflow, which is very common in bottom-up proteomics. For the standard in-solution workflow, cells were harvested and lysed in urea buffer, before the resulting protein was transferred into a new Eppendorf tube. Proteins were digested and subsequently desalted in separate StageTips. On the contrary, the iST approach was entirely performed in a single StageTip. For this, StageTips were prepared and cells were directly transferred onto the C18 disks. Then, lysis buffer was added and proteins were digested, desalted, and subsequently eluted from the same StageTip. To assess the performance on low starting amounts, 10,000 HL60 cells were processed in parallel with the in-solution workflow either using standard 1.5 ml or 0.5 ml low-binding tubes as well as with the iST protocol. Each method was performed in dublicate and proteins were subsequently acquired with MS. Utilizing the 0.5 ml low-binding tubes resulted in an increase of about 1,800 peptides and 150 proteins compared to the standard 1.5 ml tubes (II-Figure 1A). However, results from the iST approach still demonstrated superior performance, as more than 2,000 peptides and 200 proteins were identified here. Hence, these experiments demonstrated that minimizing the reaction tube surface area and the manual sample handling steps, were advantageous for low input amounts, as this decreased sample loss during the preparation process.

Based on these promising results, the next step was to optimize and minimize sample loss of the iST protocol. For this, the digestion volume was decreased from 250 µl to 55 µl and 10 µl pipette tips instead of the 200 µl were used for sample preparation. With this optimized protocol, approximately 5 times more peptides and twice as many proteins were obtained than in the standard iST protocol using only 2,000 cells as starting material (II-Figure 1B). With more than 9,000 peptides and 2,000 proteins, the optimized protocol yielded similar numbers as the first experiment (II-Figure 1A) despite applying only a fifth of the starting material. The experiment was performed in dublicate and demonstrated good quantitative reproducibility based on the peptide abundance (II-Figure 1C). To assess the detection limit, a downscaling of input amount was conducted next. For this, 2,000, 1,000, 500, and 100 cells were processed with the optimized workflow. Although the reduction from 2,000 to 1,000 cells led to a substantial decrease, still around 1,500 proteins were detected (II-Figure 1D). However, for less input material no considerable number of peptides/proteins were detected, which implies that the lower detection limit of this approach was reached.



II-Figure 1: Optimizing the iST workflow. (A) Barplots depicting the numbers of peptides and proteins identified with the standard in-solution workflow (green) in 1.5 ml or 0.5 ml tubes, respectively, as well as the iST approach (orange). The experiment was performed in technical dublicate and each bar represents one replicate. (B) Number of peptides and proteins identified with the standard and the optimized iST protocol. (C) Scatterplots showing the correlation between replicates of peptides identified with the standard (n=743 peptides, left panel) or the optimized iST protocol (n=6,149 peptides, right panel). (D) Barplots showing the number of peptides and proteins identified with decreasing input amounts using the optimized iST protocol. Bars represent the mean of three replicate and the error bars depict the standard deviation.

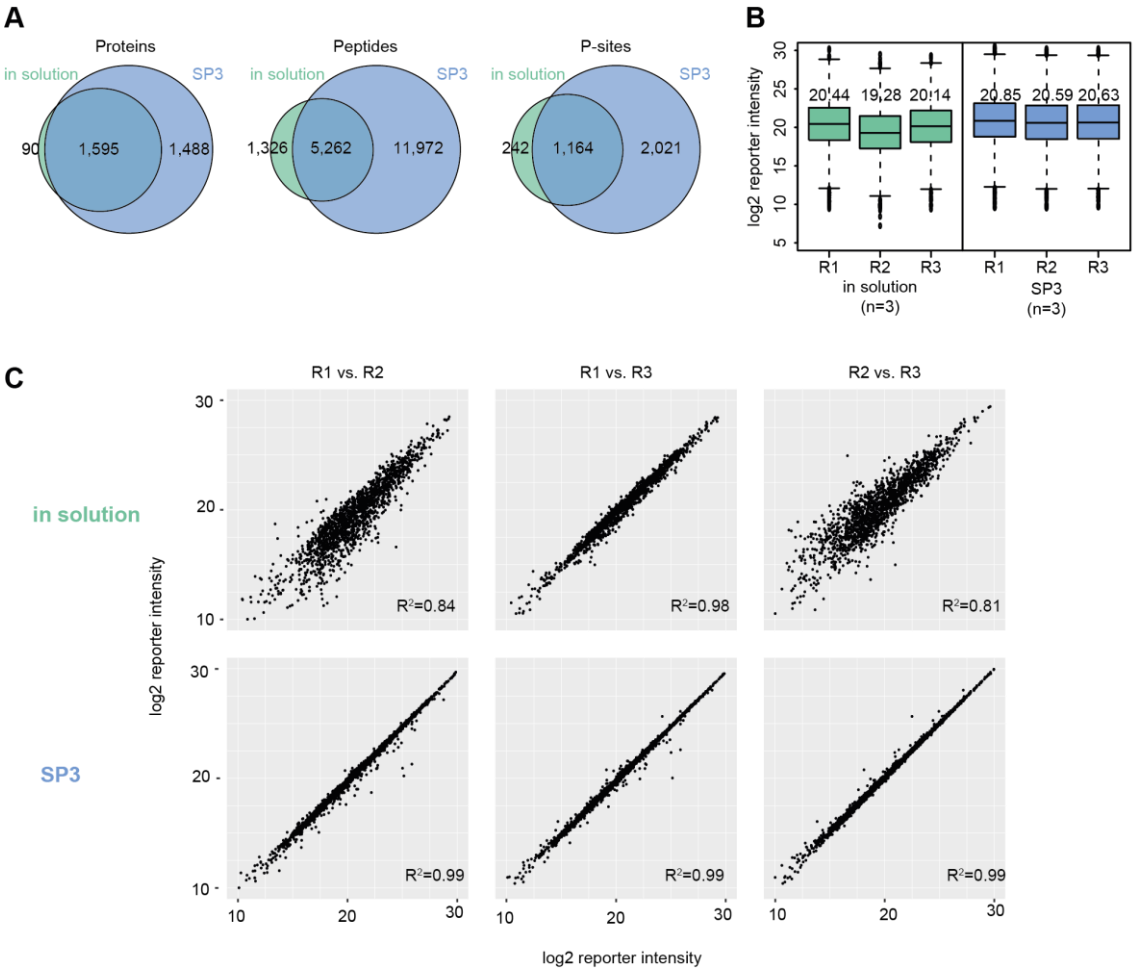
Next, the optimized iST protocol was compared with the previously published SP3 workflow [160], which utilizes paramagnetic beads for protein aggregation followed by washing steps and subsequent digestion. For this, 2,000 HL60 cells were processed separately in triplicate with both protocols and applied to a mass spectrometer. Approximately 40% more peptides and 30% more proteins were identified with the SP3 approach (II-Figure 2A). Furthermore, the number of peptides with one or two missed cleavage sites, which constitutes a relevant issue of the iST method, was greatly decreased (II-Figure 2B). One reason for the increased digestion efficiency in the SP3 method might be the improved removal of urea and other reagents potentially diminishing trypsin activity. Another reason for enhancement might be the lower digestion volume enabling a higher trypsin concentration and a higher digestion efficiency due to increased trypsin-protein interactions. While the number of identified peptides was higher for SP3 compared to iST, both approaches were similarly reproducible (II-Figure 2C). Furthermore, the peptide quantification showed comparable R^2 values when analysing raw intensities of the respective replicates (II-Figure 2D).



II-Figure 2: Comparison of iST and SP3 workflows. (A) Number of identified peptides and proteins from 2,000 HL60 cells with the optimized iST and the SP3 approach. Bars represent the mean and error bars the standard deviation of the three replicate. (B) Stacked barplot showing the percentage of 0, 1, and 2 missed cleavage sites of the three replicate derived with the iST or SP3 protocol, respectively. (C) Peptide overlap between replicates (R1-R3) of both approaches. (D) Scatterplot and coefficient of determination comparison between the peptides of two replicates acquired from the iST and SP3 workflow.

As the SP3 protocol was superior for label-free samples from minute quantities, its performance and reproducibility with TMT labelling was assessed in a next step. This method offers the advantage of increasing sample material by multiplexing up to 18 samples [217], which is desirable for comprehensive phosphopeptide enrichment as well as for deep-fractionation. To validate this workflow, triplicate of 160 µg protein lysate were processed with the SP3 or a standard in-solution protocol, respectively. For the in-solution protocol, protein lysate was digested overnight in Tris/HCl buffer, transferred onto SepPak cartridges, desalted, and dried down. The dried peptides were reconstituted in HEPES buffer and subsequently TMT-labeled. In contrast, the SP3 workflow reduced the number of manual handling steps as well as the processing time. First, proteins were precipitated on beads and washed. Then, contrary to the in-solution protocol, proteins were digested in HEPES buffer. The digested peptides were dried down and dissolved with water for subsequent TMT labelling. As interfering substances were washed away before the digestion, the additional SepPak desalting step from the in-solution protocol was superfluous for the SP3 approach. The two methods were not only benchmarked based on the fullproteome, but also on phosphoproteome level. Due to the sub-stoichiometric nature of phosphopeptides, their detection is very challenging and usually accompanied by special enrichment strategies. Here, the TMT-labeled peptides were enriched for phosphopeptides via a Fe^{3+}-IMAC column, while the flow-through, which contained the unphosphorylated fullproteome, was collected, and further fractionated using StageTips. The fullproteome as well as the phosphoproteome were acquired with a nanoflow system coupled to an Orbitrap Fusion Lumos, which was operated in MS3 mode with a 60 minute or 90 minute gradient, respectively. The SP3 approach outperformed the in-

solution protocol on the number of detected peptides, proteins, and P-sites (II-Figure 3A). However, the labeling efficiency (data not shown) for the in-solution experiment was only around 50%. On the contrary, the SP3 workflow led to more than 90% labeling, which is in the expected range based on previous TMT experiments. A reason for the unexpected low labelling efficiency might be the use of a malfunctioning TMT aliquot. Since the half-life of TMT reagent in buffer is only a few minutes, old aliquots can lead to decreased labeling efficiency if they were not stored adequately or for an extended period of time. This was presumably the case, as only the reporter intensities of the second replicate (R2) was considerably decreased (II-Figure 3B). This observation was further supported by comparing the reproducibility between replicates. While the reporter intensities of replicate 1 and 3 correlated nicely (R^2 = 0.96), the correlation to R2 was strikingly decreased (II-Figure 3C). In contrast, the SP3 approach yielded comparable levels of reporter ion intensities (II-Figure 3B) as well as high quantitative reproducibility between replicates (II-Figure 3C).



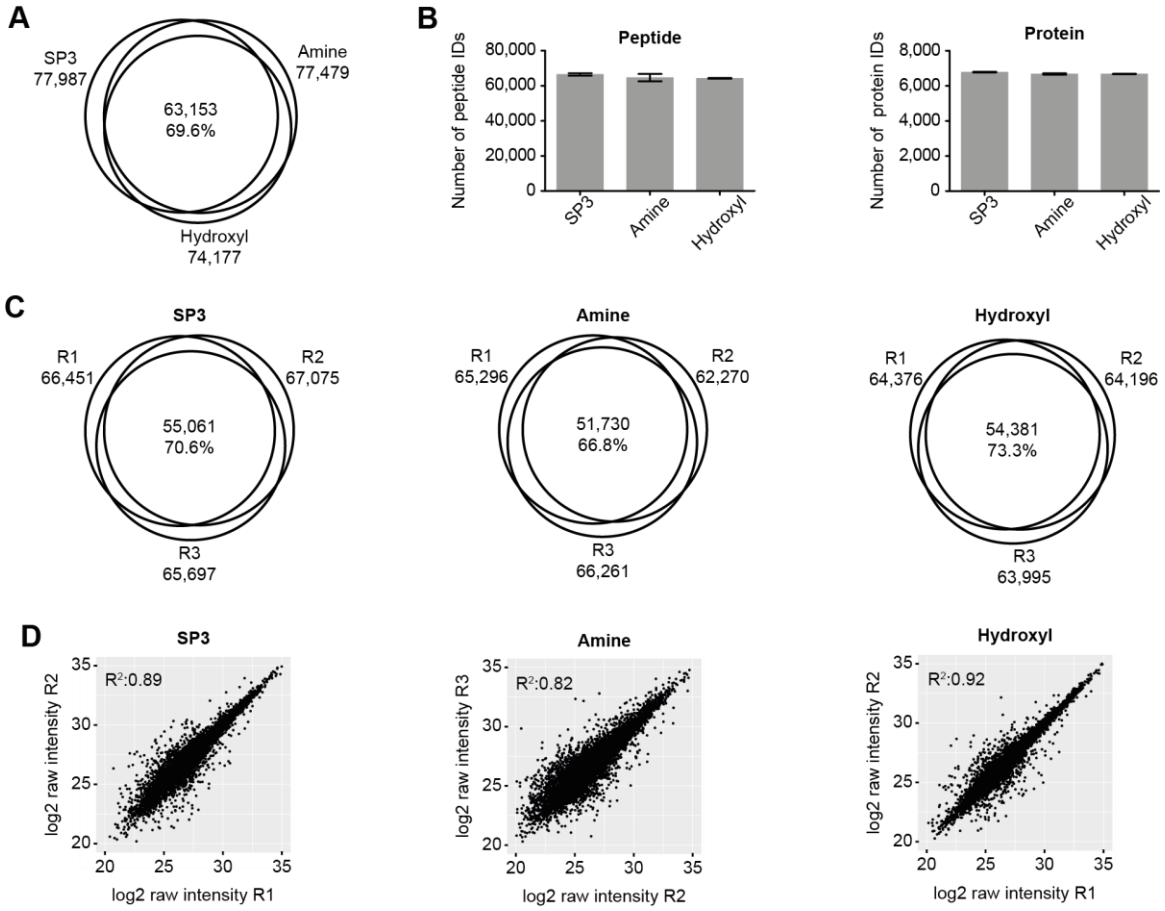
II-Figure 3: TMT labelling with the SP3 workflow. (A) Overlap of proteins, peptides, and P-sites identified with the in-solution and the SP3 approach. (B) Boxplots showing the log2 intensities of each TMT channel. Numbers depict the median. (C) Multi-scatter plot and coefficient of determination between replicate based on the reporter intensities of the identified proteins with both sample preparation workflow.

In summary, although optimizing the iST protocol increased the number of detected peptides 5 times compared to the standard iST protocol, the SP3 approach was superior for low input quantities in a label-free setting. In addition, this approach comprises advantages like the removal of detergents, scalability, and automation capability [162]. Furthermore, the coupling to subsequent TMT-labeling revealed high reproducibility indicating the high application area of this workflow from low input amount to very deep full- and phosphoproteome coverage.

4.2 Benchmarking different beads for fullproteome analysis

The following results are largely based on data obtain from Karl Kristian Krull during his internship *"Influence of surface modification in SP3 technology for functional proteome analysis"* and his master thesis *"Phosphoproteomics for low input amount using paramagnetic IMAC beads"* conducted under the author's supervision at the Chair of Proteomics and Bioanalytics at the Technical University of Munich.

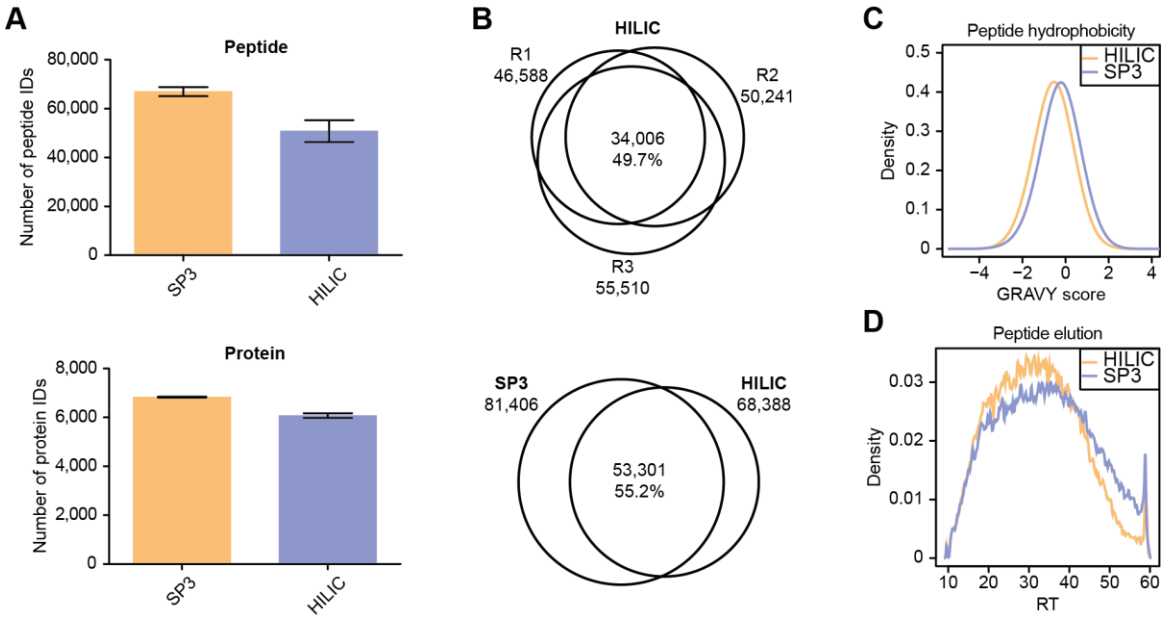
In light of the promising results obtained with the SP3 protocol, several other magnetic beads from ReSyn Biosciences (Pty) ltd were tested as alternative approaches. Compared to the carboxyl-coated SP3 beads, ReSyn Biosciences offers beads with various different surface modifications, such as amine, hydroxyl, or carboxyl groups. These magnetic microparticles are based on a patented hyper-porous matrix, which increases the surface area leading to enhanced protein-bead interactions according to the manufacturer. In order to assess the performance regarding proteomic sample preparation, amine and hydroxyl-modified ReSyn Biosciences beads were compared to the previously established SP3 protocol. Therefore, 20 µg of K652 cell lysate was processed in triplicate according to the respective protocols. In short, lysates were incubated with the beads and interfering substances were removed before the proteins were digested overnight. The resulting peptides were loaded onto a StageTip and separated into 6 fractions. Subsequently, each fraction was measured with a 1 h gradient on a Lumos Fusion mass spectrometer and the acquired data was further searched with Maxquant against a human database. In total around 78,000 peptides were identified with the SP3 and amine beads, whereas slightly less was detected with the hydroxyl beads (II-Figure 4A). The high overlap of 70% implies that the mechanism of protein capture is similar for all three approaches. This is in line with the observation that the identified peptides showed no significant differences in length, hydrophobicity, charge, or molecular mass between the beads types (data not shown). Comparable robustness was demonstrated as approximately 6,700 proteins and 65,000 peptide IDs were obtained in each condition (II-Figure 4B), which was additionally confirmed by the high reproducibility (II-Figure 4C). The high analogy of beads supported the finding of Batth and colleagues which have demonstrated that the protein-bead interaction is solely based on protein precipitation and is thus independent of the functional group immobilized on the microparticles [161]. Nevertheless, the peptide intensities acquired with SP3 and hydroxyl beads were slightly more robust with R^2 values around 0.9 compared to the amine beads (II-Figure 4D). Interestingly, the R^2 value of SP3 beads was comparable to the previous experiment where only a fraction of the input material had been used (II-Figure 2D). This again emphasizes the robustness of the SP3 bead method for sample-limited experiments.



II-Figure 4: Evaluating different magnetic beads as alternatives for SP3 beads. (A) Venn diagram showing the overlap of all peptides identified with SP3, amine, and hydroxyl beads. (B) Barplots showing the mean numbers of identified peptides (left panel) and proteins (right panel). Error bars depict the standard deviation of three replicates. (C) Same as (A) but showing the overlap between replicates. (D) Representative scatterplots showing the correlation of log2 raw intensity of peptides between two replicates.

In the next step, the commercially available HILIC beads from ReSyn Biosciences were compared with SP3 beads. According to the manufacturer's protocol, this bead type uses a HILIC-like binding mechanism where proteins are trapped between the bead surface and an aqueous layer. With an ACN concentration of 15% and a pH of 4.5 the binding conditions are clearly different to the SP3 protocol, where an ACN concentration of 70% at a neutral pH is recommended. Complementary to the previous experiment, 20 µg of cell lysates were processed in triplicate and digested according to the SP3 or HILIC protocol, respectively. The resulting peptides were further separated into 6 fractions via high pH StageTips fractionation and subsequently measured with a 1 h gradient. Alike the previous experiment (II-Figure 4B) around 65,000 peptides and 6,800 proteins were identified from each replicate with the SP3 approach (II-Figure 5A). In contrast, approximately 10-30% less protein and peptide numbers were obtained with the HILIC beads (II-Figure 5A). Additionally, the median peptide intensity was 2-4 fold decreased compared to the SP3 beads, indicating that HILIC binding was less efficient and potentially explaining the ID drop. Furthermore, only 50% of peptides were robustly identified in all three replicates, which was the

least among all tested magnetic bead types (II-Figure 5B, upper panel). The overlap of peptides identified with SP3 and HILIC beads was also low with only 55% (II-Figure 5B, lower panel), which suggests that the HILIC mode enhances the binding of different proteins. In order to elucidate on this in more detail, the hydrophobicity of peptides identified exclusively with either method were analysed. The GRAVY score distribution, which is a measure for hydrophobicity, suggested that peptides identified with the HILIC beads were slightly more hydrophilic (II-Figure 5C). This was further supported by the observation that peptides identified with the HILIC beads had the tendency to elute earlier in the gradient and thus at a lower ACN concentration compared to the SP3 approach (II-Figure 5D). This is contrary to Moggridge and colleagues [240], who reported a small bias towards hydrophobicity with HILIC beads.



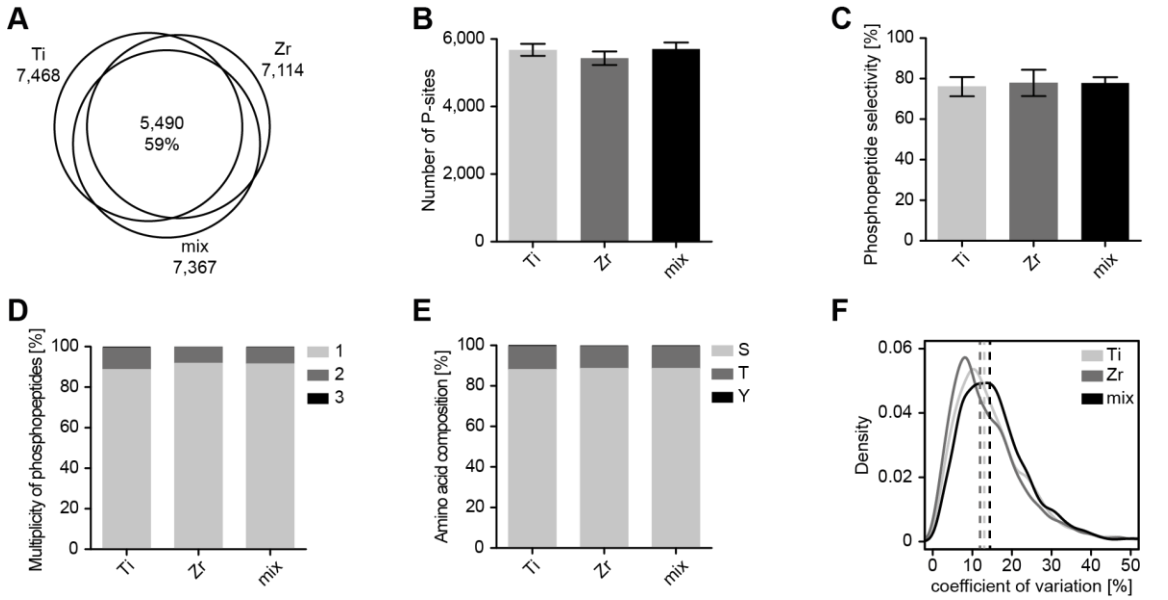
II-Figure 5: Comparison of SP3 beads and HILIC beads. (A) Number of identified peptides (upper panel) and proteins (lower panel). Barplot show the mean of triplicate and error bars depict the standard deviation. (B) Venn diagrams showing the overlap of identified peptides between replicates from the HILIC approach (upper panel) or between peptides identified with SP3 and HILIC beads (lower panel). (C) Density plot showing the hydrophobicity distribution of peptides exclusively identified with HILIC or SP3 beads, respectively. (D) Peptide elution profile along the 1 h MS gradient.

In summary, magnetic beads functionalized with carboxyl, amine, or hydroxyl groups performed similarly well regarding peptide identification, quantification, and reproducibility. This is in line with previous findings postulating that proteins are captured through precipitation rather than a specific binding to the functionalized group [161]. Only the HILIC beads performed slightly worse, which is probably due to the different binding mode and buffer conditions. Noteworthy, ReSyn Biosciences beads are 5-10 times bigger and settled quicker in the reaction tube compared to the SP3 beads, which slightly simplified manual handling steps. Since ReSyn Biosciences beads did not show any compelling benefits, the well-established SP3 beads were further used for digesting proteins.

4.3 Magnetic beads for phosphopeptide enrichment

Besides the application for fullproteome analysis, magnetic beads can also be employed to enrich phosphorylated peptides. Due to the very low abundance of phosphopeptides, an enrichment is inevitable for a comprehensive analysis. To this regard, ReSyn Biosciences developed magnetic microparticles with Ti^{4+} and Zr^{4+} cations coordinated to phosphonate groups which were evaluated in the following experiments. Besides the performance of each individual bead type, a combination of both beads was tested in order to increase the phosphoproteome coverage.

Triplicate of 100 µg desalted peptides were applied to Zr-IMAC, Ti-IMAC, or a 1:1 mixture of both beads, respectively. Samples were processed according to the manufacturer's protocol and further desalted using StageTips before half of the input amount was injected into a Fusion Lumos mass spectrometer. With a 1 h gradient around 7,400 P-sites were detected with the Ti-IMAC and the mixed approach, whereas 7,100 were acquired with the Zr-IMAC beads (II-Figure 6A). The high overlap of IDs indicates a high complementarity between Ti-IMAC and Zr-IMAC explaining why the mixed approach did not result in additional ID gain. All three methods demonstrated high robustness with the identification of 5,700 P-sites on average (II-Figure 6B). Moreover, the phosphopeptide selectivity was highly comparable between the different bead types (II-Figure 6C), albeit it was with 77% slightly lower than alternative enrichment methods, such as Fe-IMAC columns, Ti-IMAC tips, or TiO_2 beads, which commonly yield more than 90% selectivity [173, 175]. In addition, also the multiplicity of phosphopeptides and the amino acid composition was similar among the three methods (II-Figure 6D and 6E). In order to evaluate the quantitative reproducibility, the distribution of coefficient of variations (CV) within each triplicate was analyzed (II-Figure 6F). The median CV was between 12% and 14% for all approaches which is comparable to other enrichment strategies [173, 175].



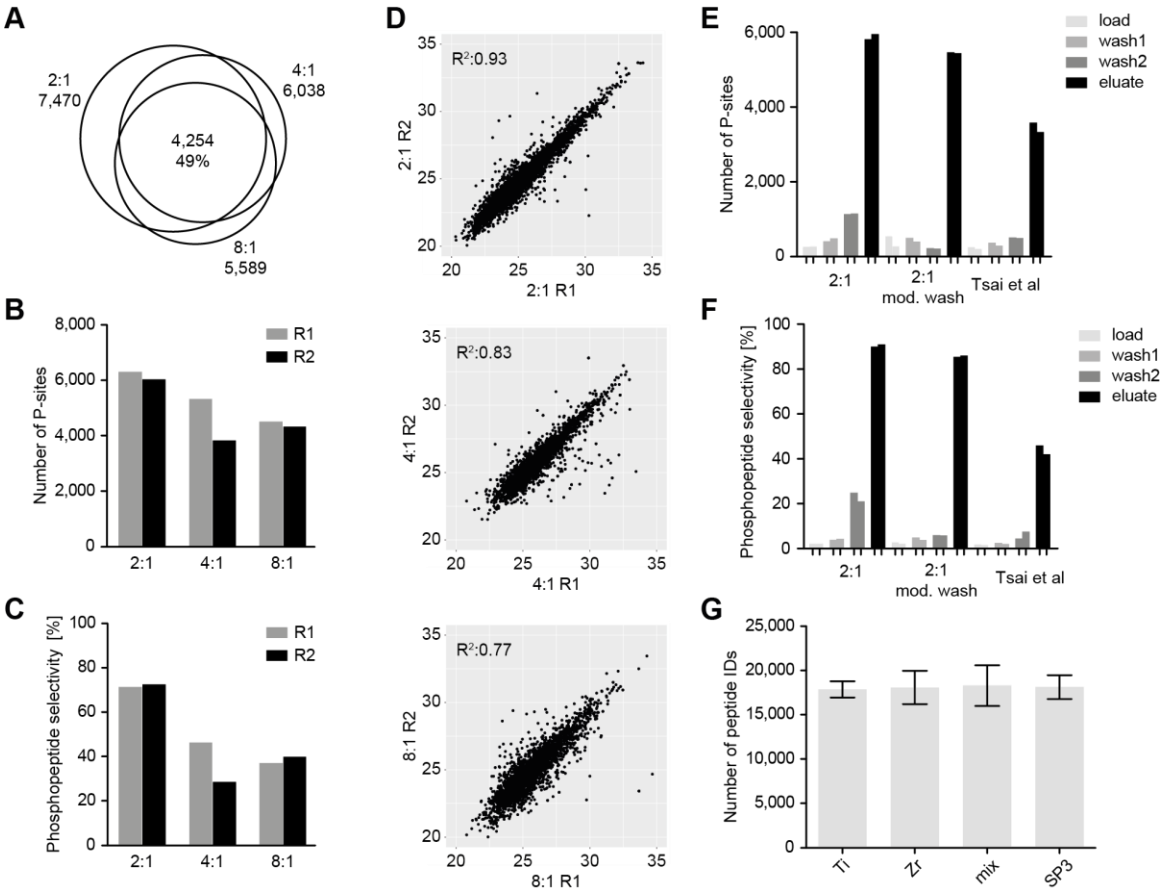
II-Figure 6: Benchmarking different magnetic IMAC-beads for phospho enrichment. (A) Overlap of phosphopeptides detected with Ti-IMAC, Zr-IMAC, or a 1:1 mix of both beads. (B) Number of P-sites identified with either of the three approaches. Barplot represents the average and error bars depict the standard deviation of the triplicates. (C) Same as (B) but for the phosphopeptide selectivity. (D) Stacked barplot showing the ratios of phosphopeptide multiplicity. (E) Same as (D) but for the amino acid

composition of the identified phosphopeptides. (F) Distribution of the coefficient of variation within each triplicate.

In order to optimize the enrichment protocol, different bead-to-peptide ratios were tested. For this, dublicate of 100 µg desalted peptides were processed using a 4:1 bead-to-peptide ratio, as recommended by the manufacturer, and additionally with 2:1 and 8:1 ratios. Again, half of the enriched phospopeptides were measured in a 1 h gradient yielding in total around 9,300 P-sites of which almost 50% were detected with all three approaches (II-Figure 7A). Interestingly, decreasing numbers of identified P-sites indicate that the enrichment is diminished with higher protein-to-bead ratios (II-Figure 7B and 7C). This is potentially caused by the increased binding capacity resulting from the higher bead amount, which in turn enhanced binding and enrichment of non-phosphorylated peptides and therefore decreased enrichment selectivity (II-Figure 7C). However, a high selectivity is crucial as non-phosphorylated peptides are usually more abundant and have higher stoichiometry compared to their modified counterparts. As DDA favours the detection of highly abundant peptides, the identification of phosphorylated peptides would therefore be decreased [241]. This effect is even more prevalent in complex lysates where the acquisition speed is not sufficient to sequence all present peptides [241]. Furthermore, studies have suggested that negatively charged phospo groups suppress the ionization efficiency, which would also favor the identification of their unmodified cognates [242]. In addition to the enhanced identification numbers, the quantitative reproducibility was considerably increased with lower bead-to-peptide ratios (II-Figure 7D). Hence, these results suggest the application of a lower bead-to-peptide ratio as recommended by the manufacturer. This is in line with a previous study which demonstrated that decreasing the amount of TiO_2 beads was beneficial [243].

Although lowering the bead-to-peptide ratio enhanced phosphopeptide enrichment, low selectivity remained a challenge. In order to investigate the enrichment in more detail, samples were analyzed after each processing step. For this, dublicate of 100 µg desalted peptides were enriched with a bead-to-peptide ratio of 2:1 and washed as recommended by the manufacturer with two different washing buffers (washing buffer 1: 80% ACN, 1% TFA; washing buffer 2: 10% ACN, 0.2% TFA). Samples from the unbound fraction after IMAC-bead incubation (load), after each of the two washing steps, and the eluate were analyzed. As expected, the vast majority of P-sites was detected in the eluate fraction (II-Figure 7E, left panel). However, more than 1,000 P-sites were lost during the second washing step. To retain these phosphopeptides, the washing procedure was modified such that washing buffer 1 was used twice and washing buffer 2 was omitted. This decreased the number of P-sites identified in the second washing step, but did not result in higher IDs in the eluate (II-Figure 7E, middle panel). A third enrichment approach was based on the publication of Tsai and colleagues [244] in which they postulated that full deprotonation of phosphopeptides and thus efficient binding to the Fe-IMAC column was only achieved at a pH above 3. As this comes at the cost of unspecific peptide binding, 6% acetic acid was added to compete and hamper the unspecific binding. However, in this setting the approach did not only identify the lowest number of P-sites but further led to decreased selectivity (II-Figure 7E and 7F, right panels). These results demonstrate that the washing steps vastly influence the enrichment and further optimization is required to reach the full potential of this workflow.

In summary, Ti-IMAC beads performed slightly better than Zr-IMAC beads and mixing both beads did not increase the phosphoproteome coverage. Samples prepared in parallel showed high reproducibility, however some variation was observed between samples prepared on separate days. While the phosphopeptide selectivity was around 70% (II-Figure 7C) in one experiment, 90% was reached in a conformable experiment (II-Figure 7F) utilizing the same lysates and protocol. Bead stability is likely a reason for the day-to-day variation. In general further optimizations are required in order to reach the robustness and IDs of the well established Fe-IMAC column [173].



II-Figure 7: Optimizing conditions for phospho enrichment with magnetic IMAC beads. (A) Overlap of phosphopeptides identified with bead-to-peptide ratios of 2:1, 4:1, and 8:1. (B) Barplot showing the number of detected P-sites with either approach. Each bar displays one of the two replicates (R1 and R2). (C) Same as (B) but for the phosphopeptide selectivity. (D) Scatterplots of the log2 transformed phosphopeptide abundance of the two replicates R1 and R2. (E) Barplots showing the number of P-sites obtained after the phosphopeptide binding (load), the washing steps (wash 1 and wash 2), and the elution (eluate). Each bar represents a technical replicate. (F) Same as (E) but for the phosphopeptide selectivity. (G) Bar charts displaying the number of peptides obtained with IMAC or SP3 beads, respectively. Bars represent the average of triplicate and error bars depict the standard deviation.

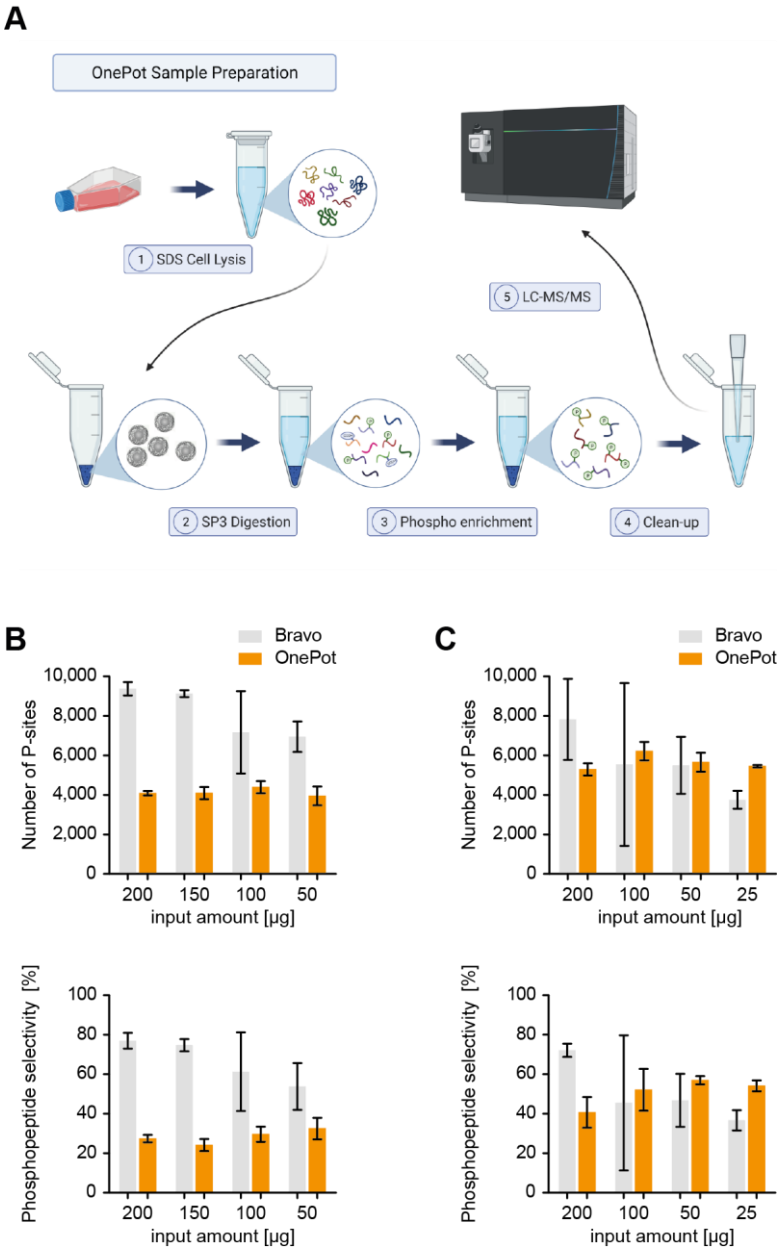
Although the magnetic IMAC beads revealed some challenges, this enrichment strategy was further tested for low input material because the previous experiment with magnetic SP3 beads performed particularly well for minute sample quantities (II-Figure 2). Extensive sample handling is one major reason for sample loss and thus unfavorable for sample-limited experiments. Hence,

the aim was to develop a workflow that can be performed in a single reaction tube with a minimal number of manual steps. This led to the development of the "OnePot" workflow, which allows both protein digestion and phosphopeptide enrichment with magnetic beads (II-Figure 8A). During the initial step proteins from cell lysate were precipitated on IMAC beads and digested according to the SP3 protocol. Changing the buffer conditions enforced the interaction of phosphopeptides with the metal(IV) cations from the IMAC beads, while unspecifically bound peptides were removed through several washing steps. Finally, phosphopeptides were eluted and transferred onto StageTips for desalting or fractionation before they were injected into the mass spectrometer.

The initial protein capture and digestion was successfully performed with IMAC beads and yielded similar peptide IDs to the SP3 approach (II-Figure 7G). This result was in agreement with previous experiments, which revealed that protein precipitation is independent of the bead surface (II-Figure 4). Nonetheless, this was a crucial processing step for the OnePot method to be successful. As the phospho enrichment of desalted peptides was also demonstrated (II-Figure 6 and II-Figure 7), only evidence for the efficient combination of both steps was lacking. Therefore, 50 µg to 200 µg of protein lysate were processed to evaluate the novel OnePot protocol. For comparison, the same starting material was digested with SP3 beads, desalted, and subsequently applied to the Agilent AssayMAP® Bravo robot. This liquid handling platform enables phosphopeptide enrichment with Fe(III)-NTA cartridges and has been successfully implemented on minute sample amounts [177]. Despite high reproducibility, the numbers of identified P-sites as well as the phosphopeptide selectivity were considerably lower in the OnePot approach than after the Bravo enrichment (II-Figure 8B). Interestingly, the IDs from the OnePot approach were not correlating with the input amount, which would be expected and was observed from the Bravo samples. As the IMAC beads were not designed for an overnight digestion at a basic pH, these conditions were potentially compromising the bead stability and thus the enrichment efficiency. In order to optimize the workflow, the digestion time was decreased to only 3 hours similar to the protocol reported by Leutert and colleagues [245]. This resulted in around 50% more P-sites and a selectivity of up to 60%, while the standard deviation remained low indicating high robustness (II-Figure 8C). Still, the Bravo workflow acquired more IDs for quantities above 100 µg, whereas the OnePot approach was superior for the lower input amounts (II-Figure 8C). The number of unique phosphopeptides obtained from 25 µg was almost twice as high as in a previous study that used the Bravo system [177] and still 15% more than Leutert and colleagues who developed the R2-P2 workflow [245], a robotic platform for phospho enrichment using magnetic beads. Noteworthy, the experimental setup as well as the mass spectrometers were different in these studies which makes a one-to-one comparison difficult. While the low standard deviation from the OnePot confirmed high robustness over the complete range of sample amount, the high variation of the Bravo system (II-Figure 8C) was unexpected as this method worked reproducibly in multiple different in-house experiments and studies before [176, 177].

To conclude, even though the OnePot revealed promising results from small sample quantities, a similar day-to-day variation was observed as described before (II-Figure 6 and II-Figure 7). This challenge has to be overcome in order to exploit the full potential of this workflow and to make it widely applicable. However, a benefit of this approach is the potential automatization, which would increase sample throughput as well as robustness. In principal automatization should be

easily implementable as the protein digestion [162] as well as the phosphopeptide enrichment [245, 246] have been performed with pipetting robots before.



II-Figure 8: Comparing phosphopeptide enrichment of the OnePot approach with the Agilent Bravo platform. (A) Schematic workflow of the OnePot. (B) Number of identified P-sites (upper panel) and phosphopeptide selectivity (lower panel). Bar represents the average of the triplicate and error bars display the standard deviation. (C) Same as (B) but with 3 h digestion instead of overnight.

Chapter III: High temporal resolution investigation of *in vitro* hepatocyte differentiation

The following chapter is largely based on the publication "High temporal resolution proteome and phosphoproteome profiling of stem cell-derived hepatocyte development" [247] published in Cell Reports on March 29, 2022.

Authors contributions for [247]:

Conceptualization, J.K., B.T., and B.K.; methodology, J.K., K.S., A.B., M.B., R.Y., A.K., H.T., B.T., J.G.C., and B.K.; software, J.K., P.S., and M.W.; validation, J.K., P.S., M.W., M.B., A.K., and B.K.; formal analysis, J.K., P.S., M.W., and B.K.; investigation, J.K., K.S., A.B., R.Y., and M.B.; resources, J.K., K.S., A.B., M.W., B.T., J.G.C., and B.K.; data curation, J.K., P.S., and M.W.; writing – original draft, J.K. and B.K.; visualization, J.K., P.S., and M.W.; supervision, B.T., J.G.C., and B.K.; project administration, J.K. and B.K.; funding acquisition, B.T. and B.K.

1 Summary

Pluripotent stem cell differentiation provides a novel resource of generating hepatocytes in a standardized and expandable way. To elucidate on the complex biological processes occuring during the *in vitro* differentiation, deep proteome and phosphoproteome datasets with high temporal resolution of two independent iPSC lines were acquired. This led to the quantification of more than 9,000 proteins, which offered sufficient depth to separate the different intermediates along the maturation with PCA analysis. A massive metabolic switch towards higher utilization of oxidative phosphorylation and fatty acid degradation was observed between the hepatic endoderm and immature hepatocyte-like stage. Simultaneously, multiple proteins associated to DNA replication showed a congruent drop in abundance. Interestingly, this drop on protein level was preceded on cell cycle checkpoint phosphorylation. Furthermore, statistical analysis revealed the differential expression of 78 stage-specific novel protein markers, which demonstrated the high regulation of multiple WNT-related activators and inhibitors. The high temporal resolution led to a detailed roadmap through hepatocyte differentiation, which does not only show alteration of biological processes, but is furthermore able to propose potential key regulators and elucidate on the activities of multiple kinases. These datasets enable to appreciate the timely sequence of biological processes and suggest starting points for future protocol improvements in order to increase the maturation of *in vitro* hepatocytes.

2 Introduction

Stem cells possess the ability to regenerate and to differentiate into various different cell types, which make them an attractive model system for regenerative medicine, developmental processes, and various diseases. Depending on their differentiation potential they can be classified into totipotent, pluripotent, multipotent, and unipotent cells [1]. One widely used model system for directed cell differentiation are ESCs, but they come with restricted accessibility and ethical concerns. In 2006 Yamanaka and colleagues discovered that somatic cells can regain pluripotency by transducing the transcription factors Oct4, c-Myc, Klf4, and Sox2 [7, 8]. This discovery led to the generation of iPSCs as an alternative to ESCs. Since iPSCs can be easily generated, expanded, and have less ethical contraints, the stem cell research was highly simplified, thereby, paving the way for a much broader application in medical research. Based on this knowledge, tissue-specific cell types from each germ layer can nowadays be derived from iPSC by the addition of various different supplements [248-251]. For hepatocyte differentiation, iPSCs are initialy differentiated towards the DE lineage. This is achieved by the supplementation of activin A from the TGF-ß family and by activation of the WNT pathway [103], either with the addition of WNT3A or the small molecule CHIR99021, which inhibitis glycogen synthase kinase 3 [252]. The DE is a progenitor stage for multiple endoderm-derived organs like the pancreas, the liver or the intestines. Next, hepatocyte differentiation is promoted by the addition of hepatocyte growth factor, dexamethasone, and oncostatin M. Although, stem cell research helped to understand basic concepts of directed differentiation towards specific cell types, the *in vitro* models mostly result in a heterogenous mix of cells. In order to improve the thruthful mimicry of fully mature hepatocytes, several studies have tried to characterize the differentiation of stem cell-derived hepatocytes in detail. However, such studies were based on a limited number of selected proteins [123, 253, 254] or solely on mRNA expression [255]. Camp and colleagues for example have used single-cell transcriptomics to shed light on the cell lineage progression and to derive information about the heterogeneity of cell populations during hepatocyte differentiation [100]. However, as the correlation between mRNA and protein expression is not fully understood [140], such studies have limitations in fully recapitulating the differentiation process. Since proteins and not mRNAs are the key players controlling most of the biological functions, a thorough investigation of the global proteome is inevitable to understand hepatocyte differentiation in more detail. Few studies focusing on protein expression have been performed so far, but insights have either been limited to early embryogenesis [62, 63, 256, 257] or have not provided a comprehensive proteome depth and data analysis [258, 259]. Besides protein expression, studies of PTMs, such as phosphorylation or acetylation, are of special interest as they are key regulators of many enzymes and control various signaling pathways. In recent years the depth and robustness of such experiments has been increased but their potential to understand and improve *in vitro* differentiations has not been exploited. To elucidate to this end, deep proteome, phosphoproteome, and acetylome data was acquired from multiple developmental stages along hepatocyte differentiation of two iPSC lines from different donors.

3 Material and methods

2D Hepatocyte differentiation

Hepatocyte-like cells were generated from two different human iPSC cell lines, TkDA3-4 [260] and Ff-I01 [123]. Both cell lines were tested for mycoplasma before the experiment was conducted. The TkDA3-4 cells were differentiated like stated previously [100]. Briefly, TkDA3-4 cells were cultivated in standard feeder-free conditions in mTeSR1 (StemCell Technologies) on laminin 511-E8 (iMatrix-511, Nippi)-coated dishes, dissociated using Accutase® (Sigma-Aldrich) and seeded in RPMI 1640 (Gibco™) medium with 1% B27™ (Gibco™), 50ng/mL WNT3a (R&D Systems), and 100 ng/mL activin A (Sigma-Aldrich) on laminin 511-E8 (iMatrix-511, Nippi)-coated dishes. For the first day after seeding, 10 µM ROCK Inhibitor Y-27632 (Fujifilm Wako Pure Chemical) was supplemented. Between day 6 and day 13, cells were cultivated in KnockOut™-DMEM (Gibco™) with 1% (vol/vol) DMSO (Sigma-Aldrich), 20% (vol/vol) KnockOut™ Serum Replacement, 1 mM GlutaMax™, 1% (vol/vol) non-essential amino acids, and 0.1 mM β-mercaptoethanol (all Gibco™). Lastly, cells were cultured in hepatocyte culture medium (Lonza) without EGF and supplemented with 20 ng/mL hepatocyte growth factor and 20 ng/mL oncostatinM (both R&D Systems) until day 21. The medium was exchanged daily during the differentiation process and the cell morphology was monitored by microscopy. For qPCR and proteomics analysis, samples at day 0 (iPSC), day 6 (definitive endoderm, DE), day 10 (hepatic endoderm, HE), day 13 (immature hepatocyte, IH), and day 21 (mature hepatocytes, MH) were harvested and further processed.

The Ff-I01 cells were differentiated like described earlier [253] and were kindly provided from our collaborators from the research group of Dr. Keisuke Sekine (Department of Regenrative Medicine, Yokohama City University, Japan; Laboratory of Cancer Cell System, Tokyo, Japan). In short, cells were cultivated in StemFit™ Basic03 (Ajinomoto) medium with 80 ng/ml bFGF (Fujifilm Wako Pure Chemical) and dissociated for differentiation using Accutase® (Sigma-Aldrich). Dishes were coated with laminin 511-E8 (iMatrix-511, Nippi) and cells were seeded in the presence of ROCK inhibitor Y-27632 (Fujifilm Wako Pure Chemical). For the first 6 days, cells were cultivated in RPMI 1640 (GIBCO™) supplemented with 20% StemFit™ For Differentiation and 100 ng/mL activin A (both Ajinomoto). Additionally, 2 µM CHIR99021 (Cayman Chemical) was added for the first 3 days and 0.5 mM sodium butyrate (Sigma-Aldrich) was supplemented from day 1 to day 4, which resulted in almost 100% CXCR4 positive cells [253]. From day 6 to day 13, cells were cultured in StemFit™ Basic03 (Ajinomoto) medium supplemented with 1% DMSO (Sigma-Aldrich), 0.1 mM β-mercaptoethanol, 0.5% L-glutamine, and 1% non-essential amino acids (all Gibco™). The medium was exchanged daily during the first 13 days of differentiation. For the final 8 days, cells were cultured in DMEM medium (GIBCO™) supplemented with 5% StemFIT™ For Differentiation (Ajinomoto) and 0.1 µM dexamethasone (Sigma-Aldrich). The medium was exchanged every second day. At day 0 (iPSC), day 6 (DE), day 10 (HE), day 13 (IH), and day 21 (MH) samples were taken for further qPCR and proteomics analysis.

RNA extraction and qPCR analysis

RNA from the harvested cells was isolated using the RNeasy Kit (QIAGEN) by following the manufacturer´s protocol. The concentration was determined via Nanodrop2000 and around 150 ng of RNA were further used for cDNA synthesis using the iScript™ cDNA Synthesis Kit (Bio-

Rad). For each reaction, 500 ng of cDNA were combined with the SensiMix SYBR Kit (Bioline) and data was acquired with the QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific). Raw data was normalized to the housekeeping gene GAPDH (TkDA3-4) or to the S18 RNA (Ff-I01) before the averaged $2^{(-\Delta CT)}$ was calculated. The RNA extraction and qPCR measurement for the Ff-I01cells was performed by our collaborators from the research group of Dr. Keisuke Sekine (Department of Regenrative Medicine, Yokohama City University, Japan; Laboratory of Cancer Cell System, Tokyo, Japan). For the heatmap visualizations, the calculated values were z-scored across all five time points. A list of all used primers can be found in **0-Table 1**.

Protein extraction, digestion, and labeling

For cell harvest, medium was removed and cells were washed with PBS before they were detached with Accutase. Then, cells were pelleted and washed twice with PBS (w/o $CaCl_2$/$MgCl_2$, Sigma-Aldrich) and resuspended with lysis buffer containing 8 M urea, 40 mM Tris/HCl (pH 7.6), 1x EDTA-free protease inhibitor (cOmplete™, Roche), and 1x phosphatase inhibitor mix (prepared in-house according to the Phosphatase Inhibitor 1, 2, and 3 from Sigma-Aldrich). Cell lysates were frozen at -80°C, thawed on ice, and centrifuged at 20.000 xg at 4°C for 20 min before the protein concentration was determined via the Pierce™ Coomassie Bradford solution (Thermo Scientific). From each time point, 100 µg (for TkDA3-4 cells) or 70 µg (for Ff-I01 cells) protein aliquots were reduced with 10 mM DTT for 45 min at 37°C, alkylated for 30 min at RT, and subsequently diluted below 1.6 M urea using 40 mM Tris/HCl (pH 7.6). For subsequent pre-digestion, trypsin was added at a 1:100 enzyme:substrate ratio and incubated at 37°C shaking with 700 rpm. After 3 h, the same amount of trypsin was additionally added to a final 1:50 ratio, and proteins were digested overnight. The reaction was stopped by adding FA to a final concentration of 1%.

Next, peptides were desalted using self-packed StageTips as described previously [164]. For this, C18 material (Octadecyl Extraction Disks, 3M Empore ™) was packed into a 200 µL pipette tip and the acidified peptides were loaded. After washing twice with 0.1% FA, peptides were eluted with 50% ACN in 0.1% FA, and dried down.

TMT labeling was performed as described previously [234]. In short, dried peptides were reconstituted in 20 µl 50 mM HEPES (pH 8.5) and 5 µl of 11.6 mM TMT reagent was added and incubated for 1 h at RT shaking at 400 rpm. The reaction was stopped by adding 2 µl of 5% hydroxylamine (Sigma-Aldrich) and all channels were pooled. The reaction vessels were rinsed with 10% FA in 10% ACN and combined to the pooled samples. The labeled peptides were dried down and further desalted using 50 mg SepPak columns (Water Corp.). After loading, peptides were washed with 0.07% TFA and eluted with 0.07% TFA in 50% ACN.

Phosphopeptide enrichment, immunoprecipitation, and off-line fractionation

Phosphopeptides were enriched using a Fe-IMAC column (Thermo Fisher Scientific) connected to an Aekta HPLC system (GE Healthcare Life Sciences) as described previously [173]. Samples were loaded in IMAC loading buffer (0.07% TFA in 30% ACN) onto the column, while the unbound full proteome flow-through was collected and dried for further analysis of the non-phosphorylated fullproteome. The bound phosphopeptides were eluted from the column with an increasing gradient of elution buffer (0.315% NH_4OH) and dried. Phosphopeptides were separated into 6 fractions using a micro-column format (StageTips with 5 discs of C18 material, 3M Empore™) and

an increasing ACN concentration (5%, 7.5%, 10%, 12.5%, 15%, 17.5%, and 50%) [236]. To obtain 6 fractions, the 5% fraction was combined with the 50% fraction and the 17.5% fraction with the flow-through. The desalted and fractionated phosphopeptides were dried and stored at -20°C until they were measured on the mass spectrometer.

The non-phosphorylated IMAC flow-through from the TkDA3-4 cells was further enriched for acetylated peptides via immunoprecipitation according to the manufacturer´s protocol with some modifications. PTMScan® Acetyl-Lysine Motif beads were aliquoted (1/8 of antibody kit per enrichment) and washed with PTMScan® IAP (immunoaffinity purification, both Cell Signalling Technology) buffer and ice-cold PBS. The dried peptides were reconstituted in 1 ml IAP buffer, mixed with the antibody beads, and incubated at 4°C for 1 h on an end-over-end rotator. Beads were pelleted at 2,000 xg and the unbound full proteome was retained for downstream analysis. The beads were further washed with IAP buffer and PBS. The acetylated peptides were eluted with 0.15% TFA and desalted using StageTips (3 disks of C18 material, 3M Empore™). Desalted peptides were dried and stored at -20°C until they were measured on the mass spectrometer.

The full proteome fractions were further fractionated via Trinity (TkDA3-4 cells) or high-pH reversed-phase fractionation (Ff-I01 cells) as described previously [170]. For Trinity fractionation, samples were reconstituted in 10 mM NH_4OAc (in water, pH 4.7) and loaded onto an Acclaim AmG C18 column (2.1x150 mm, Thermo Scientific) connected to a Dionex Ultimate 3000 HPLC system (Thermo Fisher). Peptides were eluted with an increasing gradient of elution buffer (10 mM NH_4OAc in ACN) and 32 fractions were collected. For high-pH reversed-phase fractionation, samples were reconstituted in 25 mM NH_4HCO_3 (pH 8) and loaded onto a C18 column (XBridge BEH130, 3.5 µm, 2.1x150mm, Waters Corp.) coupled to a Dionex Ultimate 3000 HPLC system (Thermo Fisher). Peptides were eluted with an increasing ACN concentration and 96 fractions were collected and further pooled to 48. Fractions were dried and stored at -20°C until measurement.

Data-dependent LC-MS acquisition

A nanoflow LC-MS setup was used for data acquisition by coupling a Dionex Ultimate 3000 UHPLC+ system to a Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific). Peptides were loaded onto an in-house packed trap column (75 µm x 2 cm, 5 µm C18 resin; Reprosil PUR AQ, Dr. Maisch) and washed for 10 min with 0.1% FA and 5% DMSO. Subsequently, peptides were separated using an analytical column (75 µm x 40 cm, packed in-house with 3 µm C18 resin; Reprosil PUR AQ) with a flow rate of 300 nl/min and an increasing ACN gradient. Measurements were performed in DDA and positive ionization mode.

For the TMT6-plex labeled full proteome, a linear 50 min gradient from 8% to 34% buffer B (0.1% FA, 5% DMSO in ACN) in LC buffer A (0.1% FA, 5% DMSO in MS-grade water) was used. Full scan MS1 spectra were recorded at 60,000 resolution and a scan range from 360-1300 m/z in the orbitrap. The AGC target was set to 4e5 charges and a maxIT of 20 ms was used. Isolated precursor from the MS1 scan were fragmented via CID and acquired with 15,000 resolution in the orbitrap. The AGC target was set to 5e4 charges and the maxIT to 22 ms. The MS3 spectra were obtained via SPS-MS3, which simultaneously selects 10 MS2 fragments that are further fragmented via HCD and read out in the orbitrap with 15,000 resolution. For this, the AGC target was set to 1e5 charges and the maxIT to 50 ms. The following minor changes were made for the TMT11-plex labeled FP

samples. MS2 spectra were acquired in the ion trap (rapid mode) with an AGC target of 2e4 charges and a maxIT of 60 ms. SPS-MS3 spectra were recorded with 50,000 resolution, an AGC target of 1.2e5 charges, and a maxIT of 120 ms.
For phoshopeptide analysis, a linear 80 min gradient from 4% to 32% LC buffer B was used. MS1 spectra were recorded with 60,000 resolution in the orbitrap, an AGC target of 4e5 charges, and maxIT of 20 ms (TMT6 samples) or 50 ms (TMT11 samples), respectively. MS1 precursors were isolated and fragmented via CID for subsequent MS2 spectra acquisition in the orbitrap. For this, 15,000 resolution and AGC target of 5e4 charges and a maxIT of 22 ms was used. The following MS3 scan was recorded at 15,000 resolution in the orbitrap for the TMT6-labeled samples and with 50,000 resolution for the TMT11-labeled samples.
Acetylated peptides were eluted in a linear 100 min gradient from 6% to 34% LC buffer B. The MS1 full scan was recorded in the orbitrap with 60,000 resolution within a scan range of 360-1,300 m/z, an AGC target of 4e5 charges, and a maxIT of 20 ms. MS1 precursors were fragmented via CID and subsequent MS2 spectra were acquired in the orbitrap with 15,000 resolution, an AGC target of 1e5 and a maxIT of 200 ms. For the SPS-MS3 spectra, 10 precursors were selected simultaneously and recorded in the orbitrap with 15,000 resolution, an AGC target of 1.2e5 charges, and a maxIT of 300 ms.

Database searching
Raw files were searched with the Maxquant software with its search Engine Andromeda [210, 237] (version [1.6.2.3]) against the UniProtKB human reference list (downloaded 22.07.2013). Default setting were applied, unless stated otherwise. The enzyme trypsin was specified as protease allowing for up to two missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification, while oxidation of methionine, and N-terminal protein acetylation were defined as variable modifications. In addition, serine, threonine, and tyrosine phosphorylation was set as variable modification for the phosphoproteome as well as lysine acetylation for the acetyl-IP. A target-decoy approach was used to adjust the data to 1% PSM and protein FDR. MS3-based TMT reporter ion quantification was enabled and the corresponding TMT correction factors were added.

Fullproteome data processing
For yielding protein quantification, the Maxquant output proteingroups.txt was used. Reversed hits and proteins that were only identified by site were removed. Then, the reporter ion intensities were log2 transformed and normalized to the averaged median of all 6 or 11 TMT channels. Batch effects between replicates were removed with ComBat [261] from the ´sva´package (version [3.30.1]) in RStudio (version [4.0.2]). Differentially expressed proteins were determined by analysis of variance (ANOVA) test with multiple testing correction according to the Benjamini-Hochberg (BH) procedure. A protein with a fold change >2 and an FDR <0.05 at one or more time points was definied as significant. Significantly changing proteins were row-wise z-scored and hierarchically clustered with the Perseus software [238]. Kyoto Encyclopedia of Gene and Genomes (KEGG) enrichment analysis was performed with the ´clusterProfiler´ [262] package (version [3.16.0]) using an FDR (BH corrected) threshold of 0.05. The GeneRatio is defined as the ratio of k/n, where k is the size of the overlap of the input proteinset with the specific KEGG

pathway and n is the size of the overlap of the input proteinset with all possible KEGG pathways. The principal component analysis (PCA) was performed with the 'factoextra' package (version [1.0.7]) and plotted with the 'ggplot2' package (version [3.3.2]). Transcription factor-target relationships were derived from TRRUST [263] and analysed for differentially expressed protein.

Phosphoproteomics and acetyl-IP data processing

Phosphoproteome quantification was deduced from the Maxquant output phosphosite.txt. Entries with a localization probability <0.75 were removed as well as reverse hits. Reporter ion intensities were log2 transformed and normalized with the correction factors of the corresponding full proteome data set. Batch effect correction was again removed with the ComBat package. Congruent to the full proteome, differentially expressed P-sites were defined by an FDR <0.05 (ANOVA with BH correction) and a fold change >2. Kinase-substrate relationships were predicted using the networkin web-tool [264]. The acetyl-IP data was processed like the phosphoproteome except that the quantitative analysis was performed on the acetyl(K)sites.txt.

scRNA-seq data processing

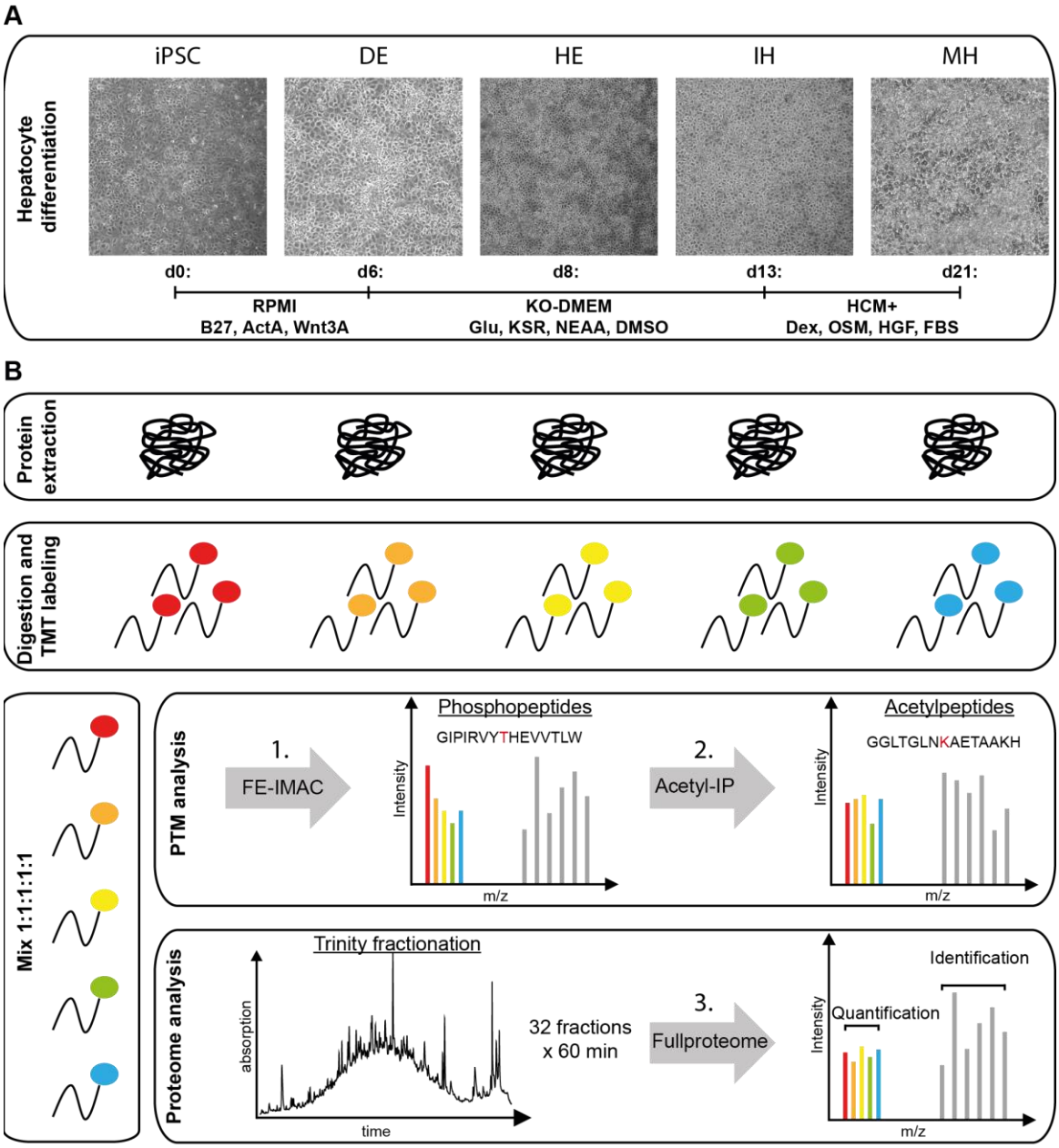
Transcriptomics data was derived from a previously published work [100] that studied hepatocyte differentiation with the same TkDA3-4 cell line. The transcriptomes of 425 single cells were acquired at the iPSC (n=80), DE (n=70), HE (n=113), IH (n=81), and MH (n=81) stages along this process. The normalized log2(FPKM) expression of all single cells at one time point were averaged to make it comparable to the bulk proteomics dataset.

4 Results and discussion

In order to elucidate biological processes during *in vitro* hepatocyte differentiation, temporally resolved protein expression, phosphorylation, and acetylation dynamics were acquired using quantitative MS.

4.1 Proteomics to study the temporal expression changes during hepatocyte differentiation

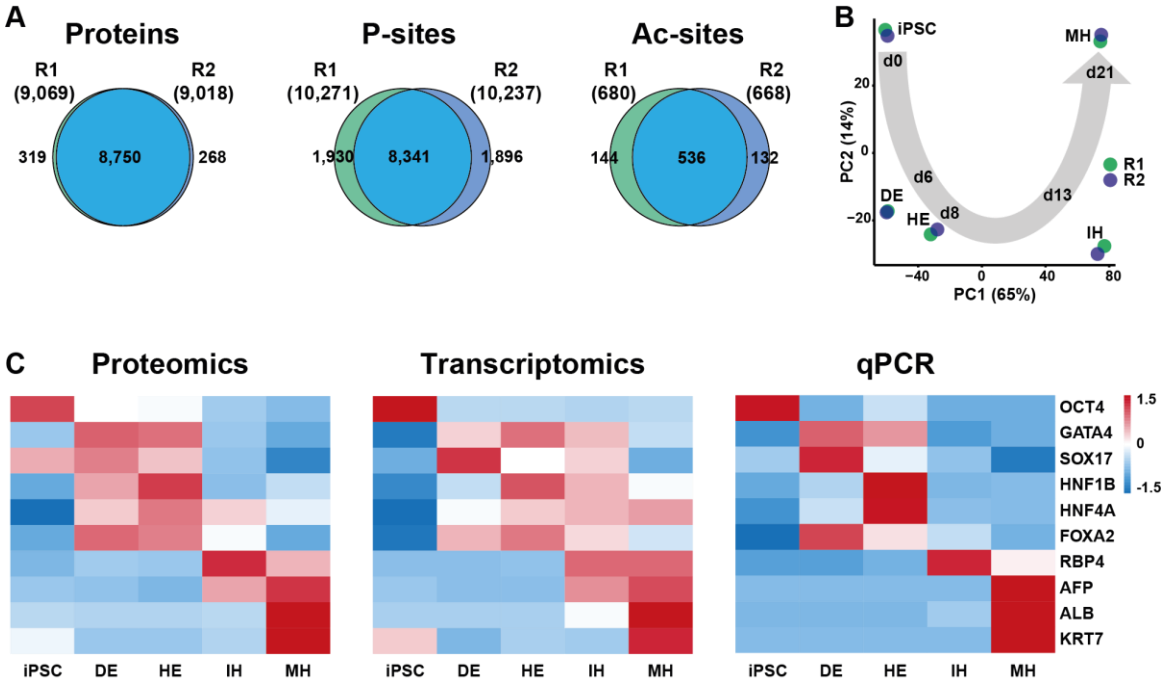
Hepatocyte differentiation was studied based on a previously published protocol [100] using human iPSCs (TkDA3-4 cells) in a 2D approach. First, stem cells were supplemented with activin A and WNT3a to form DE (III-Figure 1A), which is the common origin of multiple organs such as the liver, pancreas, and the intestines. After 6 days, the differentiation medium was modified and cells were further specified towards the hepatic endoderm (HE) for two more days. After additional 5 or 13 days, immature (IH) and mature hepatocytes (MH), respectively, have formed. To appreciate the differentiation process, samples from each developing stage were taken and used for subsequent proteomic sample preparation (III-Figure 1B). Although the workflow optimized previously (chapter II) would have been the preferred choice, it was not used for this experiment because the evaluation had not been completed when the experiment was started. After cell lysis, equal amounts of protein were employed for in-solution digested and the resulting peptides were labeled with a unique isobaric TMT reagent. Next, the samples were pooled and used for studying PTMs to expand the understanding and to add more information to the protein expression data. For this, phosphorylated peptides were enriched using a Fe-IMAC column and acetylated peptides via immunoprecipitation. To increase acquisition depth, the remaining fullproteome was fractionated before peptide spectra were recorded on a Lumos Fusion mass spectrometer. With this experimental setup around 9,000 proteins, 12,000 P-sites, and 800 acetylation sites (Ac-sites) were confidently quantified with a high overlap between replicates, which facilitates a detailed investigation (III-Figure 2A). Overlapping proteins which were identified in both replicates were further used for PCA (III-Figure 2B). Close clustering of replicate indicated high reproducibility of the differentiation as well as sample preparation. Moreover, a clear time-dependent separation along PC1 and PC2 was observed confirming sufficient depth of the dataset to study the underlying molecular mechanisms of the differentiation process.



III-Figure 1: Workflow for the proteomics experiment. (A) Microscopic characterization of the hepatocyte differentiation process from human iPSC. DE, definitive endoderm; HE, hepatic endoderm; IH, immature hepatocyte-like; MH, mature hepatocyte-like. d0-d21 denote the time in days and the information below the axis depict cell culture media and supplements. RPMI: RPMI medium; B27: B27™ supplement; ActA: activin A; KO-DMEM: KnockOut™ DMEM medium; Glu: GlutaMax™; KSR: KnockOut™ Serum Replacement; NEAA: non-essential amino acids; DMSO: dimethylsulfoxid; HCM: HCM™ Hepatocyte Culture Medium; Dex: dexamethasone; OSM: oncostatin M; HGF: hepatocyte growth factor; FBS: fetal bovine serum. (B) Proteomics workflow from protein extraction in urea buffer to digestion and labeling with TMT. The pooled TMT channels were enriched for phosphorylated and acetylated peptides, before the fullproteome was deep-fractionated and spectra were acquired via LC-MS3.

To evaluate the differentiation efficiency, a set of typical hepatocyte markers was quantified via qPCR and their expression was referenced by the averaged single-cell transcriptomes of a previous study [100] sharing the same experimental setup (III-Figure 2C). The temporal expression of all

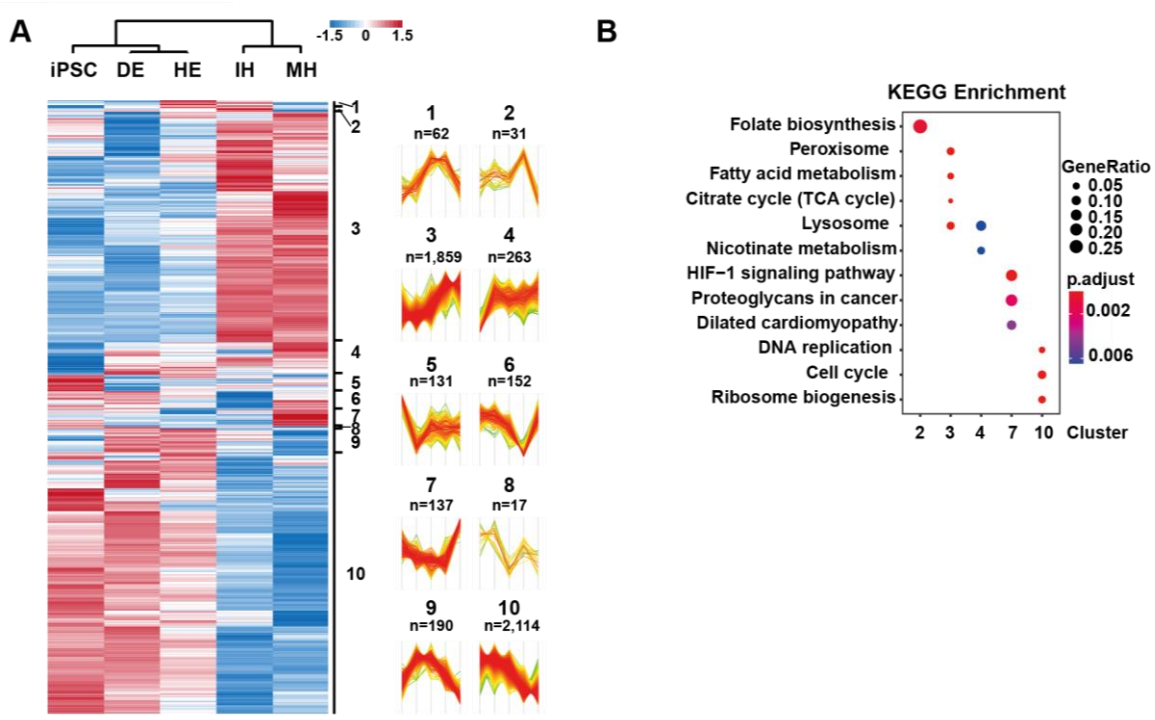
selected proteins revealed high congruence confirming the successful differentiation and high quality of the proteomics data.



III-Figure 2: Quality control and differentiation check. (A) Venn diagrams showing the number of identified proteins, P-sites, and Ac-sites of both independent replicates. (B) PCA of the proteomics experiment. (C) Z-scored normalized heatmaps of the quantities of known stage-specific protein markers. Left panel: proteomics data; middle panel: previously published single-cell RNA-seq data [100]; right panel: qPCR data.

4.2 Protein changes suggest wide-ranging metabolic switch from hepatic endoderm to immature hepatocytes

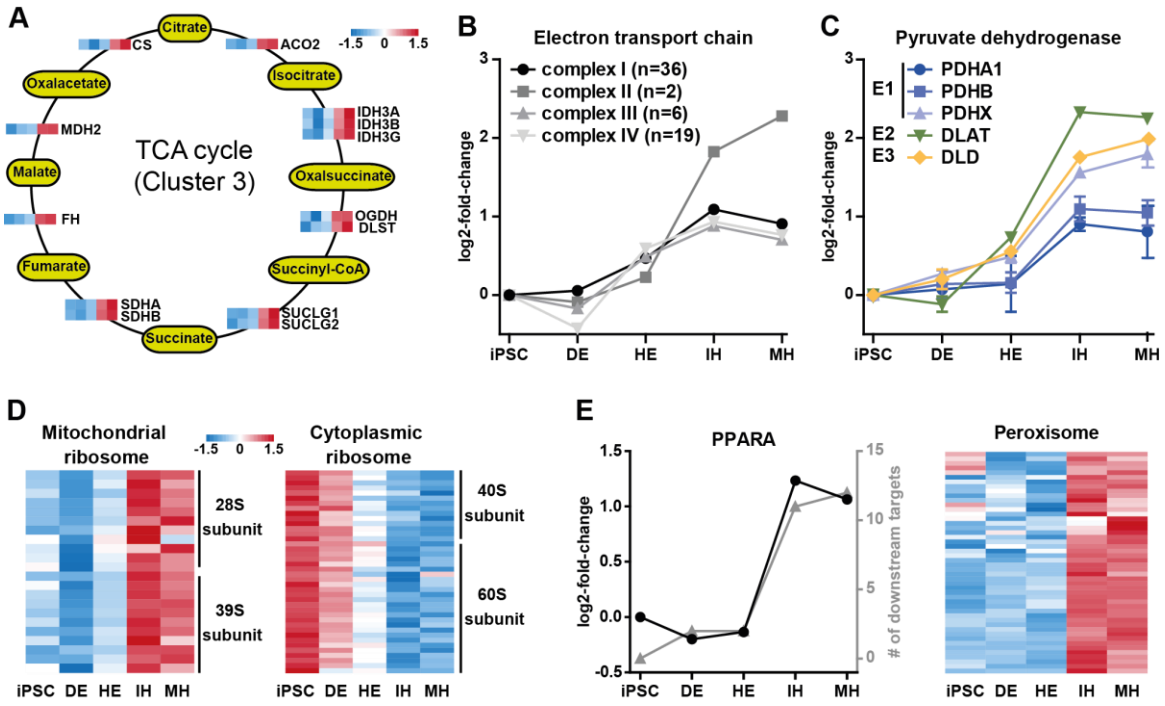
After the proteomics data quality was confirmed, the global protein dynamics were investigated in more detail. For this, the dataset was filtered for differentially expressed proteins (ANOVA: FDR <0.05, fold-change <2) in order to detect the most decisive changes. More than half of the quantified proteins, in total 4,956, showed significant expression changes at any of the investigated time points, which suggests broad protein rearrangements during the differentiation. To classify the apparent protein dynamics, hierarchical clustering was performed (III-Figure 3A, left panel), which allowed the discrimination of 10 distinct expression patterns (III-Figure 3A, right panel). With 1,859 or 2,114 members, most significantly regulated proteins were observed in cluster 3 or 10, reflecting an increase or decrease of protein expression between HE and IH, respectively. A KEGG analysis further suggested that mostly proteins related to metabolism were enriched in cluster 3, whereas cell cycle-related proteins appeared overrepresented in cluster 10 (III-Figure 3B).



III-Figure 3: Global proteomics analysis. (A) Z-scored heatmaps of hierarchical clustering from all 4,965 differentially expressed proteins (ANOVA, BH corrected: FDR <0.05 and fold change >2). Right panel shows the dynamics of all 10 distinct expression profiles (n depicts the number of proteins in this cluster). (B) KEGG enrichment of clusters defined in (A).

As multiple metabolic pathways were upregulated in cluster 3, a more detailed analysis was conducted. All 13 detected TCA cycle proteins had a synchrounous expression pattern, which showed a concerted upregulation between HE and IH (III-Figure 4A). Besides providing multiple important intermediates for biomolecules, the TCA cycle generates NADH, which is further submitted into the oxidative phosphorylation pathway and used for energy production. In this regard, a particularly interesting protein is the succinate dehydrogenase (SDHA/B), as it participates in the TCA cycle as well as the electron transport chain. The simultaneous upregulation of the succinate dehydrogenase together with multiple other proteins connected to the electron transport chain (III-Figure 4B) indicates a concerted metabolic switch towards the increased utilization of energy obtained from the oxidative phosphorylation. In addition, the three pyruvate dehydrogenase subunits responsible for converting pyruvate into acetyl-CoA, which is further fed into the TCA cycle, were upregulated from HE to IH (III-Figure 4C). Moreover, proteins associated with the mitochondrial ribosome, which mainly translates proteins from the electron transport chain, demonstrated elevated levels, whereas proteins related to the cytoplasmic ribosome decreased (III-Figure 4D). The peroxisome pathway, which possesses a key role in fatty acid metabolism, was another highly enriched pathway in cluster 3. 48 proteins associated with the peroxisome were detected in this cluster with very similar expression patterns (III-Figure 4E, right panel). Interestingly, mRNA levels of the transcription factor (TF) peroxisome proliferator-activated receptor alpha (PPARA) and the number of significantly regulated PPARA downstream targets followed the same trend (on protein expression level) pointing to an important role of this

TF in the metabolic switch (III-Figure 4E, left panel). To conclude, the numerous examples provide evidence for a metabolic switch between HE and IH. This switch includes the increased oxidation of fatty acids and higher utilization of oxidative phosphorylation. As the liver fulfills multiple metabolic functions, this metabolic switch also reflects the successful differentiation of hepatocyte-like cells. The highly synchronous switch combined with very pronounced expression changes within only 5 days was hereby striking, further suggesting the HE/IH transition as the key stage of *in vitro* hepatocyte differentiation.

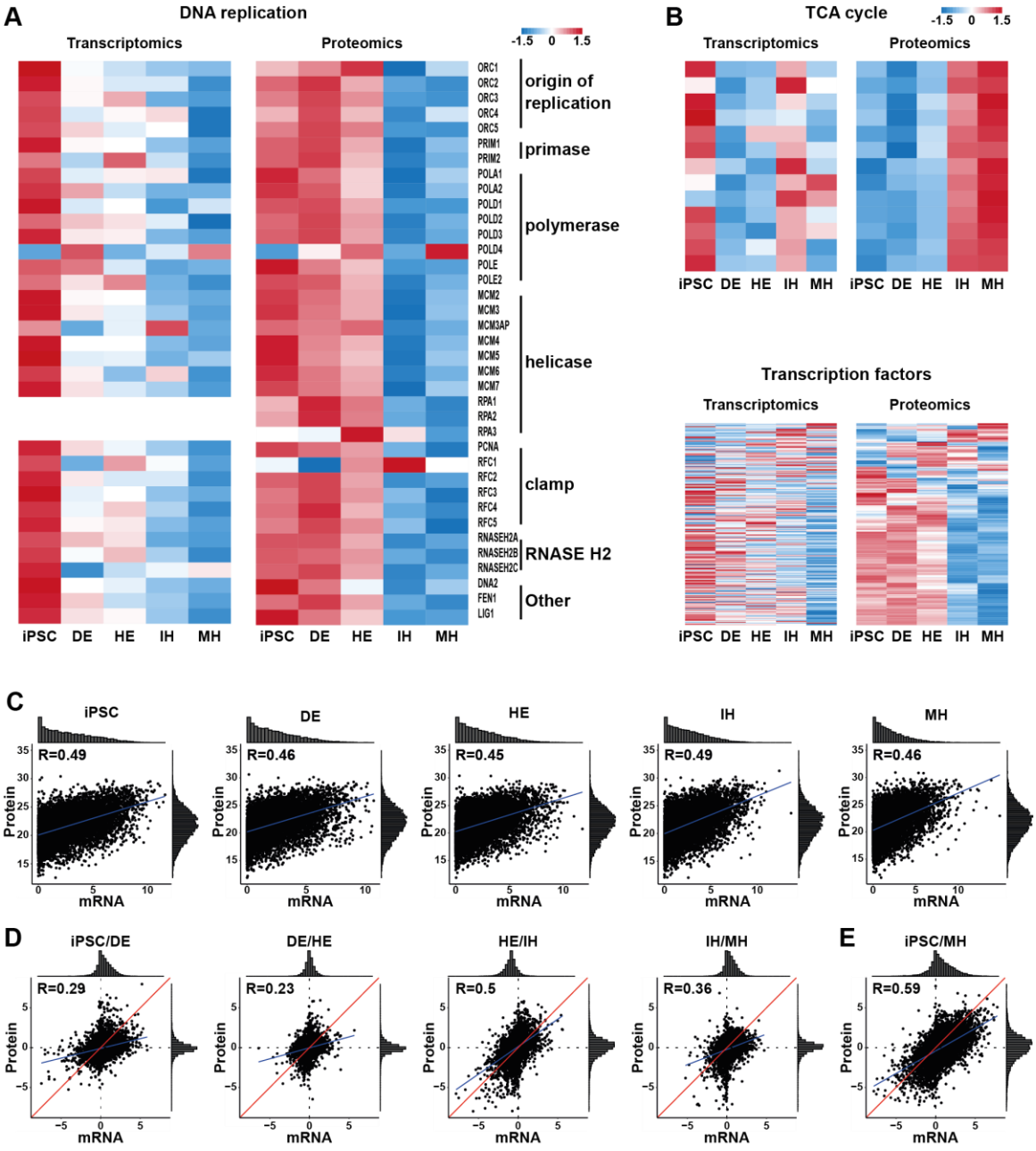


III-Figure 4: Metabolic switch between HE and IH. (A) Z-scored expression of proteins related to the tricarboxylic acid (TCA) cycle. (B) Temporal protein expression of the electron transport chain complexes I-IV. Data are normalized to iPSC and represent the median log2-fold-change (n depicts the number of proteins in each complex). (C) Temporal protein expression of the pyruvate dehydrogenase complex. Data are normalized to iPSC, represent the average log2-fold-change of both replicates, and error bars denote the range. (D) Heatmap showing the z-scored expression of proteins associated with the mitochondrial ribosome (left panel) and the cytoplasmic ribosome (right panel). (E) Left panel in black: temporal mRNA expression of PPARA. Data is normalized to iPSC and depicts the average of all cells quantified with scRNA-seq [100]. Left panel in gray: shows the number of significantly (ANOVA, BH corrected: FDR <0.05 and fold-change >2) regulated PPARA targets. Right panel: heatmap of z-scored proteins associated with the peroxisome.

4.3 Phosphorylation changes regulating cell cycle precede dynamics on proteome level

Furthermore, the KEGG enrichment revealed significant changes of multiple cell cycle-related proteins in cluster 10 (III-Figure 3B). A very consistent and concerted downregulation of the 37 identified proteins related to the DNA replication was observed between HE and IH (III-Figure 5A,

right panel). DNA replication changes were temporally coordinated with the metabolic switch, albeit with opposing direction. Single-cell transcriptomics data from the aforementioned study [100], using the same experimental setup, additionally confirmed decreasing levels of DNA replication during the differentiation (III-Figure 5A, left panel). However, the timing was different as mRNA levels were already decreased at the DE stage. This temporal difference of mRNA and protein expression was also apparent for the TCA cycle and the majority of differentially expressed transcription factors (III-Figure 5B). Noteworthy, this comparison indicated a more explicit as well as robust temporal behavior on protein level enabling an improved data interpretation. In this regard, the global correlation of mRNA and protein levels was analysed in more detail revealing a rather weak correlation at each of the 5 different developmental stages (III-Figure 5C). However, the experimental setup was not designed *a priori* for direct comparison of mRNA and protein levels because the analytes were extracted from different experiments, introducing additional variance due to the different differentiation batches. Nevertheless, the poor correlation is in accordance with previous studies reporting on the discrepancy of mRNA and protein level [140, 153]. Additionally, the ratios of proteins and transcripts between adjacent time points was calculated in order to see if similar conclusion can be drawn from either experiment. Interestingly, these data correlated even worse suggesting that the timely expression of mRNA and protein was quite different (III-Figure 5D). However, when looking at the ratio of the two stages that are furthest apart (iPSC and MH) the correlation was better (III-Figure 5E). This implies that the general trend can be recapitulated on protein and transcript level, however their temporal dynamics are discrepant.



III-Figure 5: Correlation of protein and mRNA expression levels. (A) Z-scored heatmaps of proteins related to DNA replication from the the previously published [100] transcriptomics (left panel) and proteomics (right panel) data. (B) Same as (A) but for TCA cycle proteins (upper panel) as well as transcription factors (lower panel). (C) Global correlation of mRNA and protein levels at each time point. (D) Scatterplot showing the log2-fold change between two adjacent time points on protein and mRNA level. The blue lines display the trendline, while the red lines depict the angle bisector. (E) Same as (D) but for iPSC and MH.

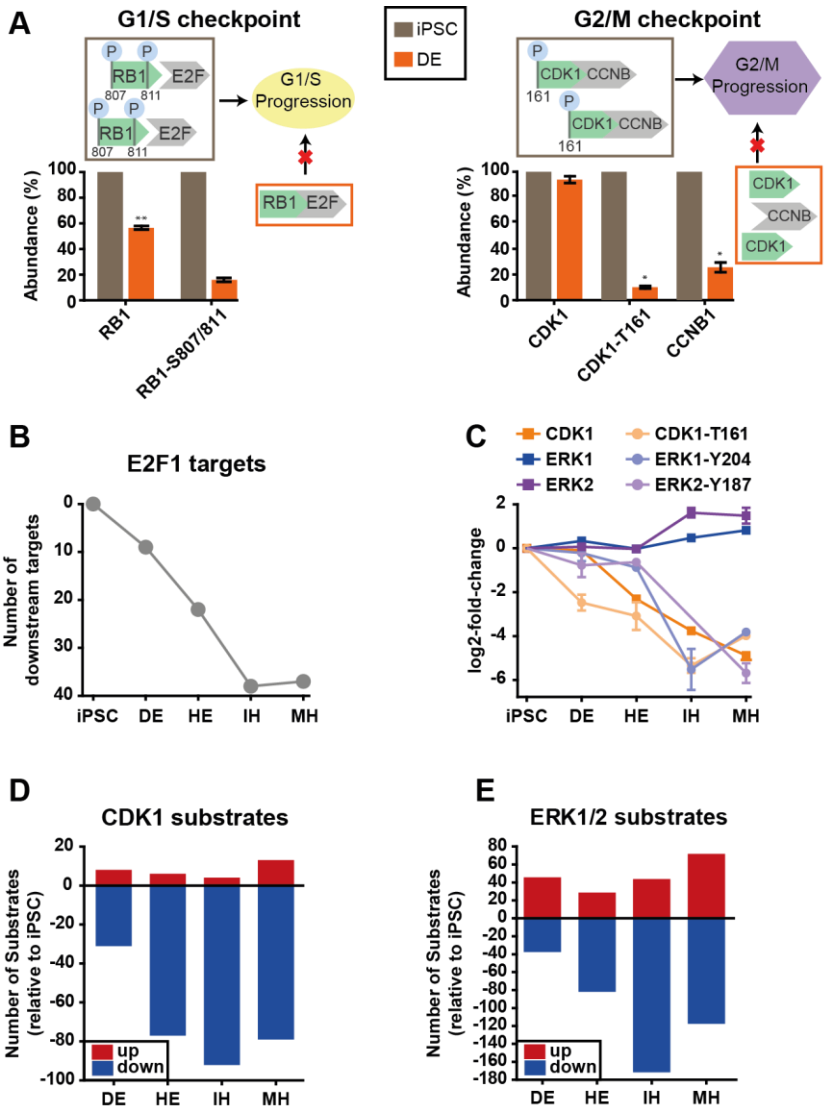
Since the cell cycle is mostly controlled via kinases and phosphatases at the two cell cycle checkpoints G1/S and G2/M, these were analyzed in more detail to elucidate on the substantial changes of the DNA replication. The retinoblastoma-associated protein 1 (RB1) is one key regulator for the G1/S checkpoint. Phosphorylation of RB1 inhibits the interaction with members of the E2F transcription factor family, which usually translocate into the nucleus to control protein expression, ultimately leading to G1/S progression (III-Figure 6A, left panel). On the contrary,

unphosphorylated RB1 binds E2Fs, resulting in impaired transclocation of the TF family and thereby impaired cell cycle progression. As the mRNA and protein levels of the DNA replication machinery revealed discrepant results, a phosphoproteomic experiment was conducted to elucidate the cell cycle alteration. The phosphoproteome analysis revealed a decrease of more than 80% of the two RB1 P-sites responsible for E2F binding, serine 807 and 811, between iPSC and DE. Noteworthy, the protein level of RB1 decreased similarly to the phospho data, although to a lesser extent. The effect of less RB1 abundance and phosphorylation levels was further confirmed by reduced levels of E2F1 downstream target proteins, which indicated a lower E2F1 activity due to scavenging of this TF by unphosphorylated RB1 (III-Figure 6B).

On the other hand, the G2/M checkpoint is mostly controlled by the cyclin-dependent kinase 1 (CDK1), whose activity is directly regulated by phosphorylation. High CDK1 activity enhances cell cycle progression, whereas low activity leads to its impairment (III-Figure 6A, right panel). While the CDK1 protein levels were constant between iPSC and DE stage, the phosphorylation at threonine 161, which induces kinase activity, dropped significantly (III-Figure 6C). In addition, cyclin B1 (CCNB1) levels were reduced, which is essential for CDK1 activity as it forms an active complex with the kinase. In agreement with lower kinase activity, phosphorylation of the downstream CDK1 substrates were diminished during the course of differentiation (III-Figure 6D).

In detail, all significantly changing P-sites (ANOVA, BH corrected: FDR <0.05 and fold-change relative to iPSC >2) were annotated with the Networkin tool [265], which uses an algorithm that combines the information from known substrates and motif-based predictions to retrieve kinase-substrate relationships.

To conclude, the detailed study of P-sites elucidated temporal dynamics of biological changes. For example, the combined analysis of protein expression and phosphorylation revealed that the changes at cell cycle checkpoints regulated by phosphorylation preceded the changes in protein expression involved in DNA replication. In general, alteration of the cell cycle are not unexpected during differentiation, as high proliferation has been previously reported as a feature of immature undifferentiated cells [266-268]. However, deciphering the discrete timing and mechanism of these processes can help to recapitulate the differentiation better and might be useful for the study of altered cell proliferation.



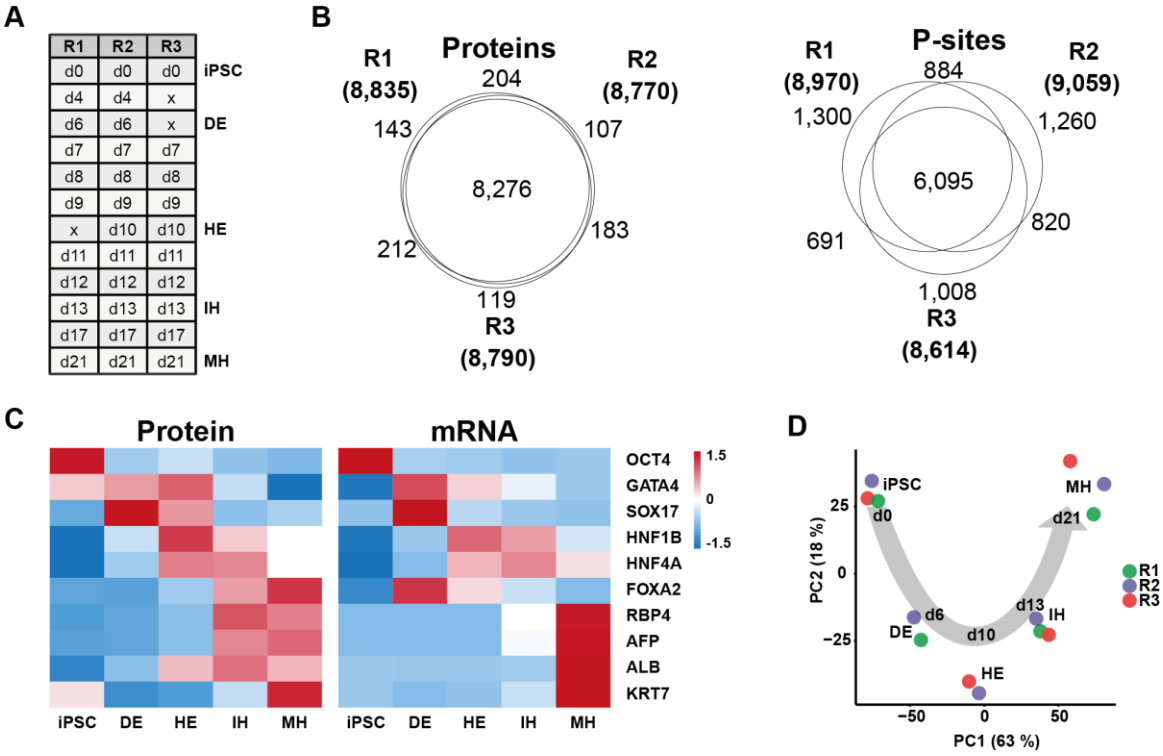
III-Figure 6: Cell cycle-related protein and phosphorylation changes. (A) Left panel: Model showing the regulation of the G1/S checkpoint. Hyperphosphorylation of RB1 impairs interaction with E2F, which leads to cell cycle progression. In contrast, unphosphorylated RB1 promotes interaction with E2F, which inhibits further G1/S progression. Right panel: Model showing CDK1-dependent G2/M checkpoint regulation. Phosphorylation at threonine 161 and binding of cyclin B (CCNB) activates CDK1, which enhances G2/M progression, while inactive CDK1 leads to the opposite effect. The y-axes show protein or phosphorylation, respectively, levels of DE relative to iPSC. Bars represent the average of two replicates and error bars denote the range. Asteriks show the significance (ANOVA, BH corrected: *FDR <0.05, ** FDR <0.01). (B) Number of significantly downregulated (ANOVA, BH corrected: FDR <0.05, fold-change <-2) E2F1 downstream targets relative to iPSC. (C) Temporal protein expression of CDK1 and ERK1/2 with their corresponding activity inducing P-sites. Data are normalized to iPSC, represent the average log2-fold-change of two replicates and error bars depict range. (D) Number of significantly (ANOVA, BH corrected: FDR <0.05, fold-change >2) up-/downregulated substrates of CDK1 relative to iPSC. (E) Same as (D) but with ERK1/2 substrates.

Another interesting finding was the diametrical opposing trend of ERK protein expression and phosphorylation. While the protein levels of ERK1 and ERK2 increased at IH, the phosphorylation at tyrosine 204 and 187 decreased substantially (III-Figure 6C). These two P-sites are positively

correlated with kinase activity, and indeed a reduction of downstream substrate phosphorylation could be confirmed by the networkin analysis (III-Figure 6E).
In summary, the two depicted examples illustrated differences between mRNA, protein, and phosphorylation levels. mRNA and protein levels can be divergent due to complex regulation of transcription, mRNA stability, translation, protein stability, and protein turn-over [269]. The addition of PTMs adds another layer of depth to comprehend the actual functionality of concrete proteins. The observed discrepancies of transcriptome, proteome, and phosphoproteome highlight the value of the herein generated datasets and the depth of information that can be retrieved. Hence, the global proteomic and phosphoproteomic datasets are important additions to the previously published single-cell transcriptomics experiment and allow detailed analysis of multiple biological processes during hepatocyte differentiation.

4.4 Biological replicate confirms hepatocyte-specific protein changes

To evaluate the concepts derived from TkDA3-4 cells, hepatocytes were differentiated using a second iPSC line (Ff-I01 cells), reprogrammed from a different donor and differentiated with a previously published protocol [253]. Hepatocyte differentiation with the Ff-I01 cells was performed with the collaborating research group of Keisuke Sekine. Compared to the previous experiment, the number of time points was increased to 12 to improve the temporal resolution (III-Figure 7A). Samples from day 0 (iPSC), day 6 (DE), day 10 (HE), day 13 (IH), and day 21 (MH) were applied for direct comparison to the previous experiment with TkDA3-4 cells. With the same proteomic sample preparation as depicted earlier around 9,000 proteins and 12,000 P-site were identified (III-Figure 7B), which is comparable to the previous experiment (III-Figure 2A). The success of the differentiation was again confirmed by the expression of well-established hepatocyte markers via proteomics and qPCR (III-Figure 7C). In addition, the PCA analysis indicated high reproducibility and a good time-dependent separation along PC1 and PC2 (III-Figure 7D).

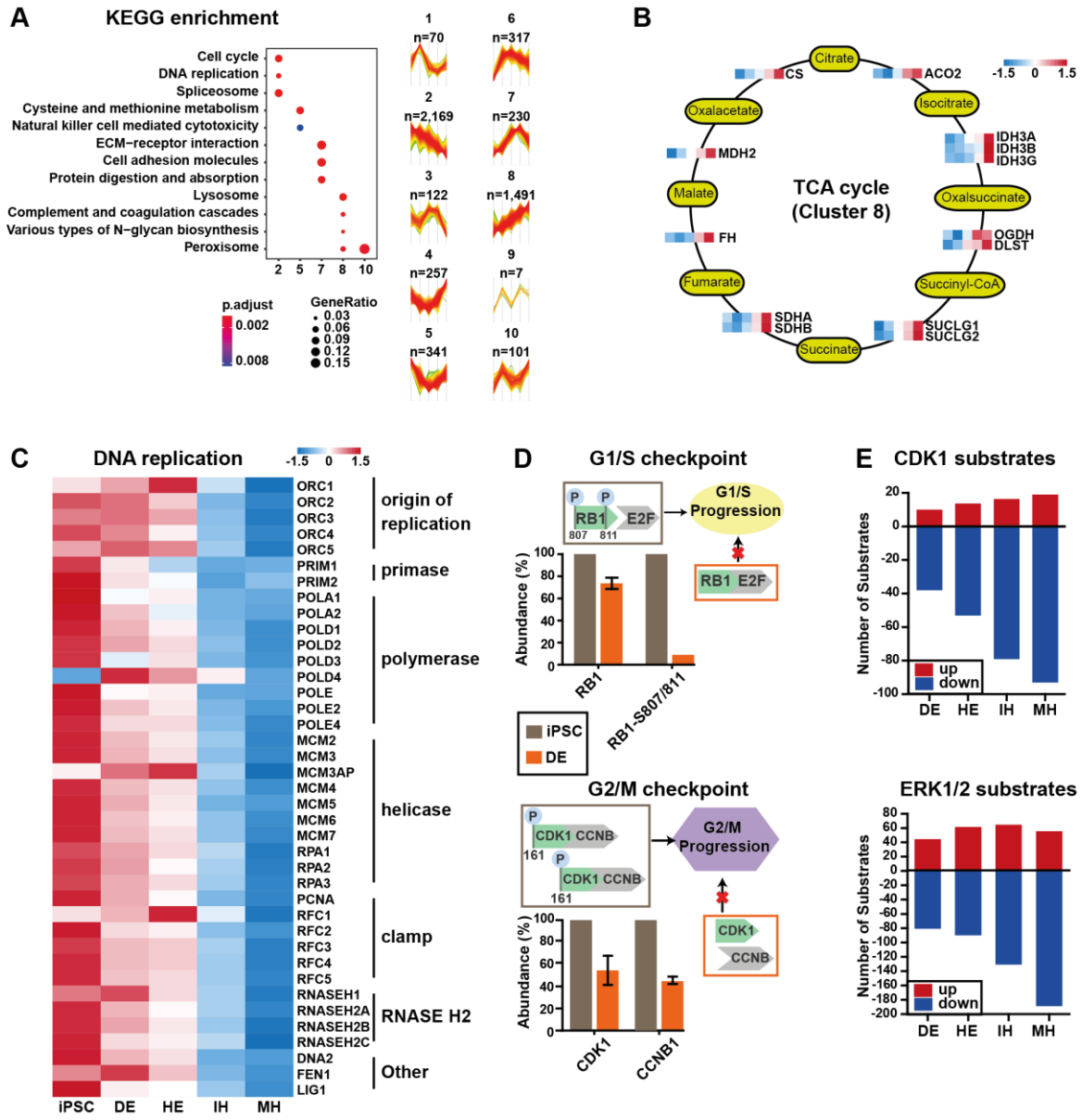


III-Figure 7: Experimental setup and quality control of Ff-I01 iPSC cell line. (A) Number of samples and time points (d=days) harvested for proteomic experiment. (B) Venn diagrams of identified proteins and P-sites in all three replicates. (C) Heatmap of z-score normalized markers quantified via proteomics (left) or qPCR (right). (D) PCA of the three replicates at iPSC (d0), DE (d6), HE (d10), IH (d13), and MH (d21).

Significantly changing proteins (n=5,105; ANOVA, BH corrected: FDR <0.05, fold-change >2) at one or more time points were selected and used for hierarchical clustering revealing 10 distinct expression profiles (III-Figure 8A, right panel). The majority of proteins was present in cluster 2 and 8 with a linear increasing or decreasing trend over time. To study the proteins in more detail, a KEGG analysis was performed for each cluster (III-Figure 8A, left panel). In cluster 8 multiple proteins related to metabolism, e.g. TCA cycle (III-Figure 8B), were enriched with a very congruent expression pattern. This upregulation is in line with the metabolic switch revealed from the previous experiment. At the same time, Cluster 2 revealed an overrepresentation of cell cycle-related processes. While proteins related to DNA replication decreased rather constantly over time (III-Figure 8C), the cell cycle checkpoints revealed substantial alterations on protein and phosphorylation levels already at the DE stage. For instance, phosphorylation of RB1 serine residues essential for binding to the E2F transcription factor family dropped to 10% relative to the iPSC stage (III-Figure 8D, upper panel). Moreover, the protein levels of cyclin B (CCNB1) were reduced, which are both indications of a diminished cell cycle progression. Noteworthy, although the activating P-sites of CDK1 and ERK1/2 were not detected, the downregulation of downstream substrates indicated decreasing activity (III-Figure 8E).

In general, the proteome and phosphoproteome analysis confirmed the previous results obtained from TkDA3-4 cells, establishing the observed changes as common characteristics of *in vitro* hepatocyte differentiation. However, the massive proteomic remodeling between HE and IH was less pronounced in the Ff-I01 compared to the TkDA3-4 differentiation. One reason for this could

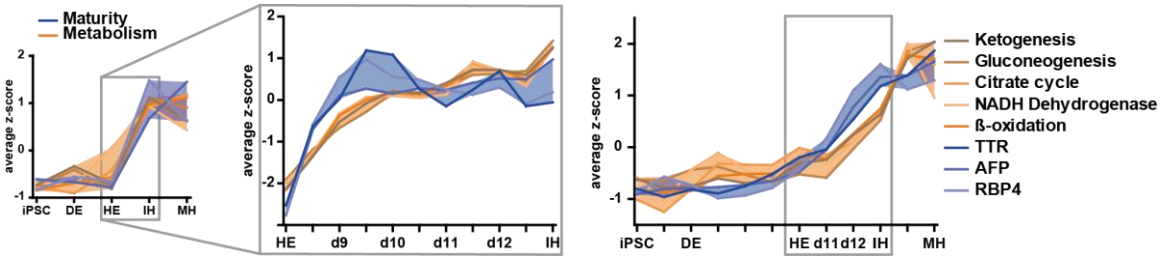
be the slightly modified protocol, as the HE samples were only taken at day 10 (Ff-I01 cells) and not at day 8 (TkDA3-4 cells). This extended HE differentiation in the first protocol might have further fostered the expression changes. Furthermore, since two independent iPSC lines with different baseline protein expression were used, some differences might be also due to cell line specificities.



III-Figure 8: Proteome and phosphoproteome analysis of Ff-I01 cells during hepatocyte differentiation. (A) Left panel: KEGG enrichment analysis of clustered proteins. Top three pathways are shown. Right panel: Expression profiles after hierarchical clustering (n=number of proteins in this cluster). (B) Dynamic expression (z-scored) of TCA cycle-related proteins. (C) Z-score normalized heatmap of proteins associated with DNA replication. (D) Upper panel: Model showing the regulation of the G1/S checkpoint. Hyperphosphorylated RB1 inhibits E2F interaction, which leads to cell cycle progression. In contrast, unphosphorylated RB1 interacts with E2F, which inhibits G1/S progression. Lower panel: Model showing CDK1-dependent G2/M checkpoint regulation. Phosphorylation at threonine 161 and binding of cyclin B (CCNB) activates CDK1, which enhances G2/M progression, while inactive CDK1 leads to the opposite effect.

The y-axes show protein or phosphorylation, levels of DE relative to iPSC. Bars represent the average of two replicates and error bars denote the range. (E) Number of significantly regulated (ANOVA, BH corrected: FDR <0.05; fold change >2) substrates of CDK1 and ERK1/2 relative to iPSC. Substrate predictions are based on Networkin.

In order to address if the metabolic switch fostered the differentiation into hepatocytes or vice versa, the dynamics of several metabolic pathways as well as some selected IH markers were studied. To increase the temporal resolution of the metabolic switch, the TkDA3-4 experiment (III-Figure 9, left panel) was repeated and samples were taken every 12 h between the HE and IH stage. Of note, the increase in maturation marker preceded the metabolic switch (III-Figure 9, middle panel). The timely separation was further supported by the differentiation of Ff-I01 cells (III-Figure 9, right panel) and has not been described before. Indeed, the timing of this switch is of interest considering that energy production of PSCs relies on aerobic glycolysis [270]. This phenomenon to maintain an anaerobic metabolism even in the presence of sufficient oxygen levels has previously been described in cancer cells and is called the Warburg-effect [271]. Despite elaborated research, the exact function of this effect remains unclear. With a decrease in pluripotency a metabolic switch from the less-energy efficient anaerobic glycolysis would be directly expected. However, the switch to aerobic glycolysis was only observed after 8-13 days.



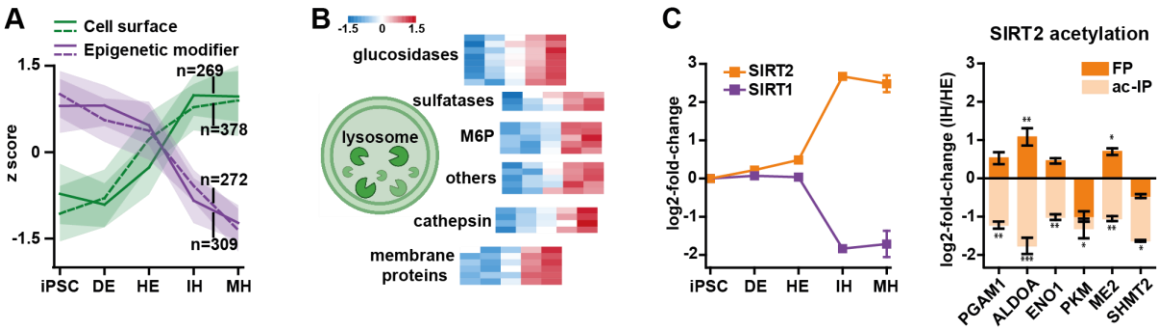
III-Figure 9: Metabolism vs. maturity. Temporal expression of selected markers for immature hepatocytes (blue) and metabolism pathways (orange) from TkDA3-4 cells (left panel), from TkDA3-4 cells with 12 h samples between HE and IH (middle panel), and from the Ff-I01 cells (right panel). Lines represent the median expression (z-scored) of all proteins related to the corresponding pathway.

4.5 Widespread changes in protein families accompany hepatocyte differentiation

To gain a more detailed knowledge of functional processes that are happening during the differentiation, the previously selected differentially expressed proteins (ANOVA, BH corrected: FDR <0.05; fold-change >2) from both cell line experiments (TkDA3-4 and Ff-I01 cells) were further divided into classes with distinct biological functions. Firstly, cell surface proteins (n=492 for TkDA3-4, n=518 for Ff-I01), derived from the cell surface protein atlas [272], and epigenetic modifiers (n=350 for TkDA3-4, n=383 for Ff-I01), retrieved from the Epifactor database [273], were hierarchically clustered (n=8 clusters) to detect proteins with similar expression profiles. From each protein class the temporal profiles of the cluster comprising the most proteins is illustrated in III-Figure 10A. Within the group of cell-surface proteins, lysosomal proteins were

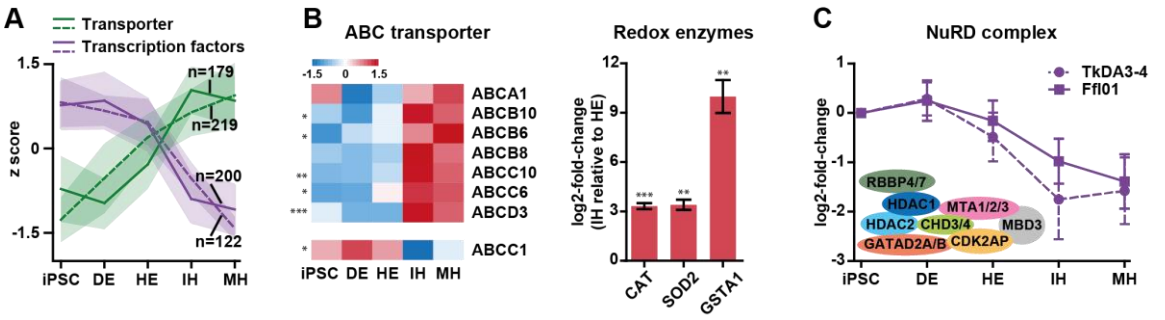
overrepresented in both biological replicates (III-Figure 10B and 0-Figure 1A). Since the lysosome is responsible for degradation of multiple biopolymers, like fatty acids or carbohydrates, its upregulation might be associated with the co-occuring metabolic switch. The impact of different lysosome levels were linked to metabolic changes in quiescent and primed hematopoetic stem cells before [274] and were suspected to be involved in epidermal differentiation [275]. There are additional studies claiming that the lysosome not solely executes metabolic functions, but is rather an organelle recognized with an emerging role in cell signaling [276, 277]. However, additional experiments are required to unravel the exact functions of lysosomes for hepatocyte differentiation. Noteworthy, from the 726 detected cell surface proteins, 97 were differentially expressed at at least one time point (0-Table 2). For example, the two fibroblast growth factor receptors (FGFR) 1 and 4 were significantly upregulated at the HE stage suggesting increased utilization of FGF signaling, which was previously proposed as an important pathway for early hepatocyte development [105, 278]. This list of stage-specific cell surface proteins can serve as resource for controlling future hepatocyte differentiations and for purifying specific cell populations, e.g. via fluorescence-activated cell scanning (FACS) sorting.

With 88% (TkDA3-4) or 71% (Ff-I01), the vast majority of proteins annotated as epigenetic modifiers substantially decreased during the differentiation suggesting a global epigenetic change of the cells (III-Figure 10A). One example of an epigenetic modification is the acetylation of histone, thereby regulating gene expression [279]. Acetylation, however, is not limited to histones, but is a PTM on lysine residues that can be actively added or removed by enzymes and can thereby alter biological functions. One group of deacetylating enzymes are the NAD^+-dependent Sirtuins (SIRTs), which have previously been associated with cell differentiation and metabolic remodeling [280]. Especially SIRT1 and SIRT2 have been thoroughly investigated in this regard [281-288]. Interestingly, these two proteins showed strikingly opposite dynamics in their protein expression (III-Figure 10C, left panel), which was further confirmed with higher temporal resolution (0-Figure 1B). To elucidate more on the deacetylase activity during the differentiation, an immunoprecipitation (IP) was performed to enrich acetylated peptides. Given that SIRT2 was the only significantly increasing deacetylase from HE to IH, increasingly downregulated Ac-sites in this time span were presumed potential SIRT2 substrates. Indeed, 93 significantly decreasing Ac-sites (ANOVA, BH corrected: FDR <0.05; fold-change <-2) were identified between HE and IH. Among these was the previously reported SIRT2 substrate phosphoglycerate mutase PGAM-K100, which has important roles in energy production [285]. In addition, several other metabolic enzymes were deacetylated, including ALDOA-K13, ENO1-K262, and PKM-K115 (III-Figure 10C, right panel), thereby indicating novel Ac-sites to the SIRT2 substrates previously identified by Cha *et al.* [281]. To conclude, SIRT1 and 2 were identified as potential key regulators for the observed metabolic switch. Moreover, novel putative SIRT2 substrates together with their dynamics during hepatocyte differentiation were identified. In order to verify the biological function of these substrates and to attribute them to SIRT2, additional experiments are required. For example, a knockout or a specific inhibition of the activity could elucidate more on the influence of SIRT2 as a key regulator.



III-Figure 10: Expression profiles of cell surface proteins and epigenetic modifiers. (A) Temporal expression of the majority of cell surface proteins and epigenetic modifiers. Median and standard deviations are shown as lines and range (n=number of proteins with this profile). Solid lines represent the TkDA3-4 experiment and dashed lines the Ff-I01 experiment. (B) Z-scored expression of lysosomal proteins (TkDA3-4 experiment) depicting a subset of the ´cell surface´group in (A). (C) Left panel: temporal expression of SIRT1 and SIRT2 as log2-fold-change relative to iPSC. Right panel: log2-fold-change of protein (FP, dark orange) and corresponding Ac-site (ac-IP, light orange) of known and putative SIRT2 substrates. Data in both plots represent the average of two replicates, the error bars denote the range, and asteriks show the significance (ANOVA, BH corrected: *FDR <0.05; **FDR <0.01; ***FDR<0.001).

The liver synthesizes large quantities of multiple biomolecules, such as bile, proteins, and lipids, with vital functions for the body. In order to secret these molecules, hepatocytes need to express high levels of transporter proteins, which makes this an interesting class of proteins to study in more detail. Therefore, from the list of previously selected differentially expressed proteins, transporter proteins were hierarchically clustered to detect common expression patterns. With 62% (179 of 289 detected) in TkDA3-4 and 65% (219 of 337 detected) in Ff-I01 cells, respectively, the expression of most transporter proteins showed a consistent increase over time (III-Figure 11A). Among these, multiple ATP-binding cassette (ABC) transporters were detected, which are in general highly expressed in the liver. This protein class is tightly connected to metabolism, efflux, and drug resistance in the liver making primary hepatocytes a frequently used cell system for preclinical drug toxicity tests [289]. The upregulation of ABC transporters is additional evidence of a successful differentiation towards hepatocytes (III-Figure 11B, left panel and 0-Figure 1C, left panel) and a prerequisite for the suitability of these cells for toxicity tests. Interestingly, expression of the ABCC1 transporter, which positively regulates efflux of the antioxidant glutathione, decreased along differentiation. However, the decrease of ABCC1 might be a consequence of the metabolic switch to oxidative phosphorylation for energy consumption during hepatocyte differentiation. The change to an aerobic energy production causes elevated levels of reactive oxygen species (ROS) as a byproduct, which is in excess harmful to the cell and therefore needs to be removed. Indeed, multiple enzymes for eliminating ROS were additionally upregulated (III-Figure 11B, right panel and 0-Figure 1C, right panel) and might constitute with the decrease in glutathione efflux the cell adaptations to aerobic metabolism.



III-Figure 11: Expression profiles of transporter proteins and transcription factors. (A) Temporal expression of the majority of transporter proteins and transcription factors. Median and standard deviation are shown as lines and range (n=number of proteins with this profile). Solid lines represent the TkDA3-4 experiment and dashed lines the Ff-I01 experiment. (B) Left panel: Heatmap of z-scored ABC transporter protein expression. Right panel: log2-fold-change of enzymes reducing ROS at IH relative to HE. Bar represents the average, error bars denote the range of two replicates, and asterisks show the significance (ANOVA, BH corrected: *FDR <0.05, **FDR <0.01, ***FDR <0.001). (C) Temporal expression of the NuRD complex. Data points depict the median of the relative log2-fold-change (normalized to iPSC) of all complex members shown in the lower left corner. Error bars represent the standard deviation.

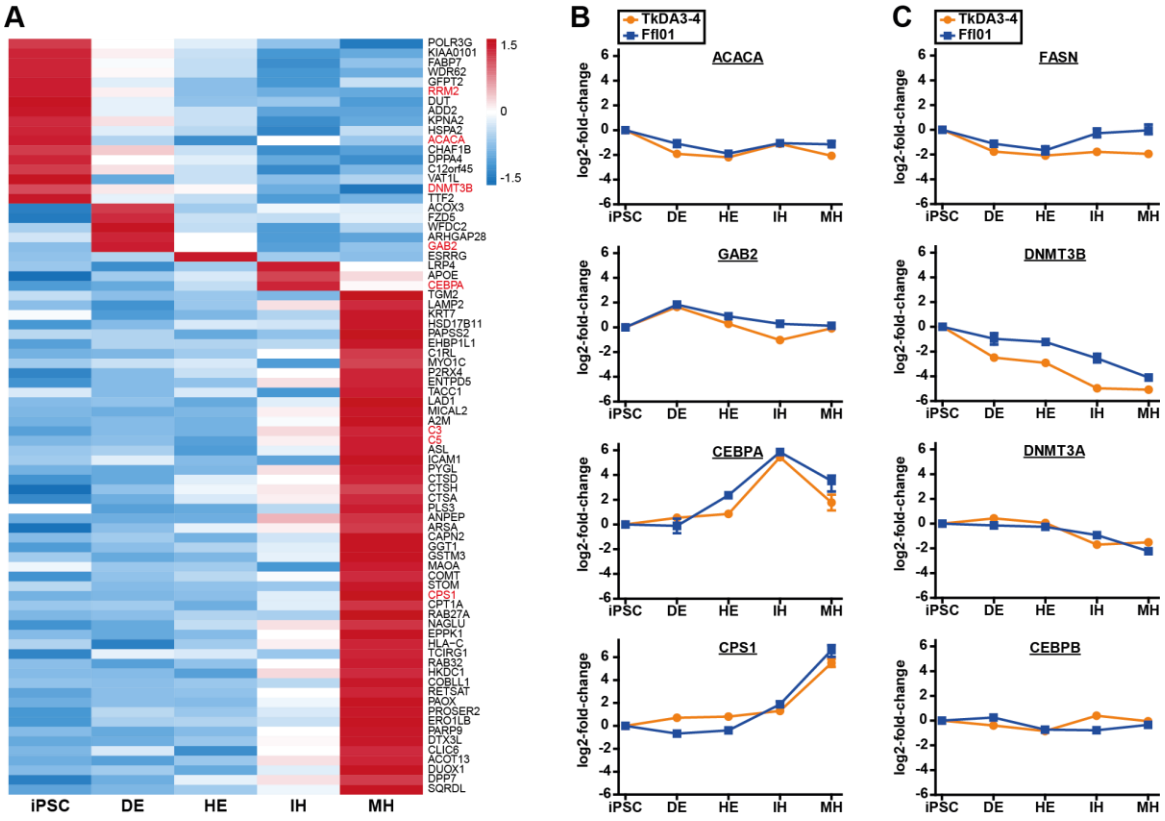
Transcription factors (TFs) are proteins that bind to specific gene regions regulating their transcription rates and thereby control cellular function and identity. TFs are of special interest for developing organs, as they orchestrate cell fade and lineage specification. To elucidate their influence on hepatocyte differentiation, the abundances of TFs at the development stages were investigated. For this, significantly expressed TFs (ANOVA, BH corrected: FDR <0.05; fold-change >2) were hierarchically clustered, as the protein classes before. The analysis revealed for the majority of TFs (76% for TkDA3-4 and 51% for Ff-I01) constantly high abundance levels in the beginning of the differentiation (III-Figure 11A). Co-expression of a broad spectrum of TFs might enable the rapid and dynamic response to different signaling cues to differentiate into different cell types. With increasing cell specification, a substantial drop of TF levels was observed (III-Figure 11A), which suggested that the cells removed most nonessential TFs during maturation, whereas some important TFs for hepatocyte differentiation were elevated. One example for this is the NuRD (nucleosome remodeling deacetylase) complex (III-Figure 11C), which is expected to fine-tune the expression of a variety of genes important for lineage commitment [290, 291] and maintains the heterogeneity in ESCs [292]. In addition, similar expression dynamics were observed by the multiprotein complex SIN3 (0-Figure 1D), which, like the NuRD complex, belongs to the HDAC1/2 complex family, which are known chromatin modifier that regulate transcription [293]. Since the majority of proteins altered their expression levels between HE and IH (III-Figure 3A and III-Figure 8A), substantial changes of TFs at this time can be expected.

4.6 Temporal protein profiling reveals novel stage-specific markers

With the depth of these proteomic datasets, an additional expression analysis was conducted to elucidate marker proteins that are specific for a certain time point. To obtain stage-specific markers, the differentially expressed proteins (ANOVA, BH corrected: FDR <0.05) were filtered for

proteins that were at least 2-fold higher than at all other time points. This analysis revealed 78 proteins that were quantified with high reproducibility and congruent expression profiles in both iPSC lines (III-Figure 12A).

Among the 17 novel iPSC markers was the acetyl-CoA carboxylase (ACACA), which is the rate-limiting enzyme in long fatty acids biosynthesis (III-Figure 12B). Another protein with similar biological function is the fatty acid synthase (FASN). Although FASN was not part of the marker list, the expression profiles were quite similar (III-Figure 12C). As PSCs proliferate rapidly, the high abundance of key enzymes for fatty acid production might be due to a high demand of fatty acids in processes like membrane synthesis. The DNA methyltransferase (DNMT) 3B was another interesting iPSC marker (III-Figure 12C) that emerged from this analysis. While DNMT3A and DNMT3B are required for *de novo* methylation, DNMT1 maintains the methylation state (0-Figure 2A). Although both DNMT3s are highly homologous and share multiple common functions [294], their temporal expression during hepatocyte differentiation diverge from each other. While DNMT3B levels dropped considerably (2 to 5-fold) between iPSC and DE, DNMT3A expression remained constant until the HE stage (III-Figure 12C). This trend was further supported by the transcriptomics analysis (0-Figure 2B) suggesting distinct functions of both proteins as previously described in hESCs [294]. Interestingly, DNMT1, but not DNMT3A/B, has been shown to be essential for cardiomyocyte and hematopoietic differentiation before [295], which implies that the maintenance of methylation levels might be more crucial for cardiac development than the *de novo* methylation [296]. The overall decreasing levels of all three DNMTs suggest a reduction of methylation during hepatocyte differentiation but further experiments are required in order to get a more detailed view on the methylation state.



III-Figure 12: Novel stage-specific protein marker. (A) Heatmap of proteins significantly upregulated (ANOVA, BH corrected: FDR <0.05) with more than 2-fold compared to any other time point. Data is z-scored and represents the average from both iPSC lines. (B) Temporal expression of selected examples for stage-specific protein markers from (A). (C) Temporal expression of proteins related to (A) or (B). Data points show the average and error bars denote the range of replicates.

PSCs and hepatocytes are both known for expressing high levels of the epithelial marker E-cadherin. However, instead of maintaining their epithelial phenotype throughout the differentiation, PSCs undergo an EMT for DE generation [297]. The transcriptomics and proteomics data supported this previous finding by a drop of E-cadherin levels upon DE differentiation followed by a steady increase towards MH (0-Figure 2C). Moreover, mRNA levels of the mesenchymal marker N-cadherin increased substantially at the DE-stage and maintained constant, while protein levels showed a steady increase over time. This EMT for DE development is an interesting finding considering that GRB2-associated binding protein 2 (GAB2) was detected as a novel DE marker (III-Figure 12B). While GAB2 was previously associated with EMT in cancer cell [298], the proteomics analysis suggested an additional connection to DE development.

Given that most TFs dropped upon differentiation (III-Figure 11A), the identification of the TF CCAAT/enhancer-binding protein alpha (CEBPA) as an IH marker was of particular interest. CEPBA was previously linked to several metabolic genes [299] and the knock-out in mice was lethal due to the inability to accumulate hepatic glycogen [300, 301]. Furthermore, CEBPA was shown to be highly abundant in early mice embryos and coexpressed with the IH marker alpha-fetoprotein (AFP, [302]). The high-resolution proteomics data adds more temporal information to these previous findings by showing the very high and specific upregulation of CEBPA at IH with a more

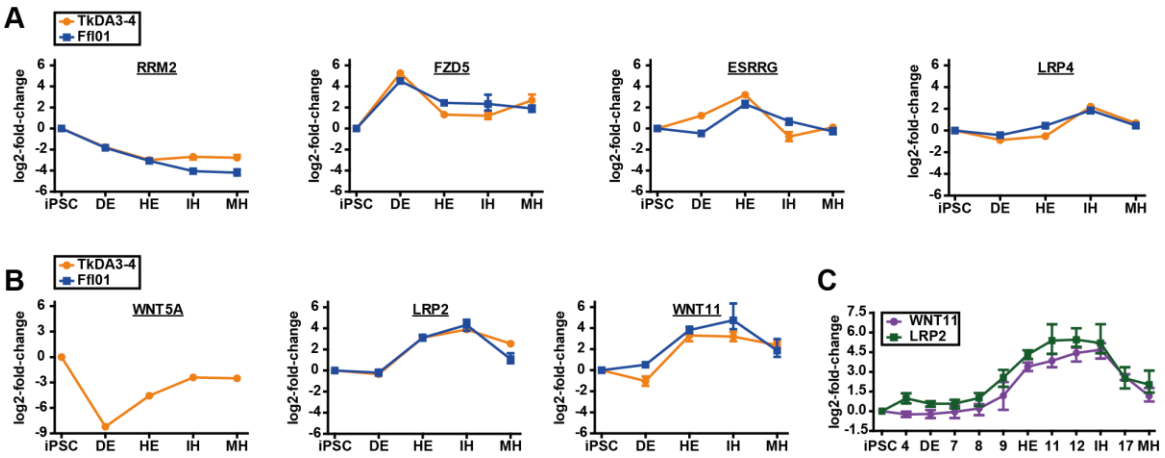
than 10-fold increase (III-Figure 12B) but a subsequent drop at MH stage, which indicates a very time-dependent role for hepatocyte development. As the CEBPA upregulation is co-occuring with the observed metabolic switch, this could explain the lethality of CEBPA knock-out in mice due to insufficient hepatic metabolism [299, 300]. Another important TF in the liver is CCAAT/enhancer-binding protein beta (CEBPB), which originates from the same basic region leucine zipper family and has demonstrated the induction of hepatocyte proliferation [303]. This TF is highly upregulated during liver regenerative response [304] and its overexpression enables the conversion of pancreatic cells into hepatocytes [136]. However, the temporal profiling of CEBPB revealed no significant expression changes along the hepatocyte differentiation (III-Figure 12C), suggesting different regulation and function of CEBPA and CEBPB during hepatocyte differentiation.

While many proteins that are specific for functional hepatocytes were already elevated prematurely at the IH stage, 52 of the 78 novel marker proteins were nevertheless upregulated at MH stage. Interestingly, 19 of these MH markers were related to the immune system, which is a substantial and often underestimated biological function of the liver [305, 306]. Two prominent examples of these marker proteins are the complement proteins C5 as well as C3 (III-Figure 12A), which is the most abundant complement component in the blood being primary synthesized in hepatocytes [305]. The more than 20-fold increase in MHs compared to iPSCs suggests the thruthful recapitulation by the *in vitro* model. Similarly, carbamoyl-phosphate synthase (CPS1), which is the rate-limiting enzyme of the urea cycle, was upregulated more than 16-times at MH. Such changes only occurring in late hepatocyte differentiation further substantiate the physiological relevance of the proteomics data on the *in vitro* hepatocyte differentiation.

4.7 New insights into hepatocyte differentiation by WNT signaling

The canonical WNT signaling is recognized as an essential driver for DE development [103, 307], which reasons the supplementation of WNT3A or the WNT activator CHIR99021, respectively, during the first 6 days of differentiation. The previous marker analysis (III-Figure 12A) revealed substantial alterations of multiple proteins associated with the WNT signaling pathway at specific time points during the differentiation. For example, the WNT inhibitor RRM2 was highly abundant in iPSC and subsequently decreased upon differentiation (III-Figure 12A and III-Figure 13A). Additionally, the WNT-receptor frizzled 5 (FZD5) increased more than 8-fold during the first 6 days suggesting FZD5 as an important marker for DE development (III-Figure 13A). At the same time WNT5A, which is a driver of the non-canonical WNT pathway [308-310] and antagonist of WNT3A [311], dropped considerably (III-Figure 13B). This implies that not only the canonical WNT signaling is upregulated for DE development, but also the non-canonical WNT pathways are downregulated. While FZD5 receptor levels dropped beyond DE differentiation, the WNT inhibitors estrogen-related receptor gamma (ESRRG) [312] and WNT11 [310] were increased at the HE stage (III-Figure 13A and III-Figure 13B). This increase was followed by the upregulation of another WNT inhibitor Low-density lipoprotein receptor-related protein 4 (LRP4) [313] at the IH stage, suggesting that these inhibitors might inhibit WNT in a different way or even different WNT proteins. The inhibition of WNT beyond the DE is in accordance with previous studies demonstrating negative effects of DE development due to prolonged WNT3A exposure [101, 103]. If WNT signaling was

not stopped after 7 days, DE cells had started losing their progenitor capacities and failed further differentiation towards the pancreatic or liver lineage.
In summary, the analysis of WNT signaling during hepatocyte differentiation identified several novel key regulators and shed further light on this not fully understood developmental process.



III-Figure 13: WNT signaling during hepatocyte differentiation. (A) Temporal expression of WNT-related proteins that were identified as stage-specific markers (see III-Figure 12A). (B) Temporal expression of proteins related to the WNT pathway. (C) High-temporal dynamics of WNT11 and LRP2 from the Ff-I01 experiment. Data points show the average and error bars denote the range of replicates.

Low-density lipoprotein receptor-related proteins (LRPs) are a family of co-receptors involved in WNT signaling. Upon WNT binding to the frizzled receptor, LRP5 and LRP6 are involved in the transduction of the canonical pathway [314]. So far only these two family members have been associated with the WNT pathway, while the others are mostly known for regulating lipid metabolism [315]. Interestingly, the expression profiles of LRP2 and WNT11 were highly analogous in both iPSC lines (III-Figure 13B) suggesting a novel receptor/ligand interaction. The enhanced resolution of the Ff-I01 experiment confirmed the high temporal correlation of both proteins (III-Figure 13C). In addition, the temporal dynamics were not only congruent, but with a relative fold-change to iPSCs of more than 20-fold all the more striking. So far, WNT11, a known WNT inhibitor, and LRP2, a multiligand receptor for mediating endocytosis of several molecules [316-318], have not been associated with each other. Owing to the prior knowledge of LRP5 and LRP6, it is tempting to hypothesize that LRP2 is a novel co-receptor of WNT11 which is involved in its inhibitory function. However, further experiments are indispensable to prove the direct functional interaction.

Chapter IV: Benchmarking *in vitro* hepatocytes and *in vivo* liver samples

The following chapter is largely based on the publication "High temporal resolution proteome and phosphoproteome profiling of stem cell-derived hepatocyte development" [247] published in Cell Reports on March 29, 2022.

Authors contributions for [247]:

Conceptualization, J.K., B.T., and B.K.; methodology, J.K., K.S., A.B., M.B., R.Y., A.K., H.T., B.T., J.G.C., and B.K.; software, J.K., P.S., and M.W.; validation, J.K., P.S., M.W., M.B., A.K., and B.K.; formal analysis, J.K., P.S., M.W., and B.K.; investigation, J.K., K.S., A.B., R.Y., and M.B.; resources, J.K., K.S., A.B., M.W., B.T., J.G.C., and B.K.; data curation, J.K., P.S., and M.W.; writing – original draft, J.K. and B.K.; visualization, J.K., P.S., and M.W.; supervision, B.T., J.G.C., and B.K.; project administration, J.K. and B.K.; funding acquisition, B.T. and B.K.

1 Summary

Primary human hepatocytes thruthfully preserve the functionality of their *in vivo* counterparts and are hence very valuable for multiple medical research fields such as drug discovery, toxicology studies, and regenerative medicine. However, primary hepatocytes are very scarce and show high donor variability, limiting their broader application and highlight the need for alternatives. Pluripotent stem cell-derived hepatocytes present such an alternative. Apart from the classical monolayer culture, hepatocyte can also be generated as 3D liver organoids thereby most likely recapitulating cellular functionality and organization of hepatocytes closer than 2D culture systems. But so far, most studies classifying the maturity of pluripotent stem cell-derived hepatocytes have been limited to transcriptomics data or a few selected marker proteins. For an improved evaluation of the complex biology of the developmental process, a global proteomic and phosphoproteomic analysis was performed. For this, stem cell-derived hepatocytes from a 2D and 3D model were compared and benchmarked to fetal liver and primary human hepatocytes. With 8,800 proteins and 12,700 phosphorylation sites, the up to now most comprehensive proteomic and phosphoproteomic datasets comparing *in vitro* and *in vivo* hepatocytes was acquired. A PCA and the upregulation of several liver-specific proteins verified the successful differentiation into hepatocytes of both *in vitro* models. However, the 3D differentiation into hepatocytes was indeed superior to the monolayer approach, as the purity of the derived 3D cultures was higher and various liver-specific metabolic pathways were upregulated. Nevertheless, multiple proteins related to crucial liver-specific functions, such as the metabolism of alcohol and aldehyde, or the absorption, distribution, metabolism, and excretion of bioactive compounds, were still substantially lower expressed than in the fetal liver and primary hepatocytes. In addition, analysing kinase-substrate relationships uncovered significantly altered activity of multiple kinases between the *in vitro* and *in vivo* samples. In summary, although the results demonstrated superior differentiation efficiencies of the 3D protocol compared to 2D, both *in vitro* models fell short in completely resembling the *in vivo* counterparts. Owing to the deep proteome and phosphoproteome datasets multiple starting points for future protocol improvements were discovered, which exemplifies the value of the resources collected in this thesis.

2 Introduction

Initial protocols for generating tissue-specific cell types from PSCs were utilizing 2D monolayer formats. This has changed over the last years as more and more 2D protocols have been supplemented or replaced by 3D models [42]. These 3D models enable the integration of multiple different cell types and are therefore capable of resembling the complex structure of an organ in a more physiological manner. For example, iPSCs can be differentiated into complex liver buds (LBs), comprising hepatic endodermal, mesenchymal, and endothelial cells [123, 253]. By co-culturing these three cell types, they self-aggregate to form vascularized 3D organoids and their transplantation partly rescued liver failure in mice [122]. The recapitulation of essential liver functions by the LBs demonstrates the great potential of this technology for regenerative medicine. Moreover, PSC-derived hepatocytes are extensively studied as a potential model system for assessing absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) in the context of drug discovery. So far the 'gold standard' for such experiments have been PHH [128, 319], which are less demanding concerning culturing and can be employed for high-throughput drug screening [320]. In recent years, this technology has also evolved from 2D monolayer to 3D models, which better reflects the conditions *in vivo* [321]. However, PHH come with some disadvantages, as they are scarce, show high donor-to-donor variability, and are prone to changing their characteristics upon *in vitro* cell culturing [322]. As stem cell differentiation overcomes the shortage of cells by providing a virtually unlimited source as well as the potential for standardized and high-throughput protocols, they are increasingly explored as a promising alternative [323, 324]. Although PSC-derived hepatocytes possess a great potential for various applications, a full characterization on protein levels is still missing, as previous studies were mostly based on transcriptomics data [100, 255] or covered only a limited number of proteins [123, 253, 254].
In order to evaluate how closely they recapitulate their *in vivo* counterpart, the global protein expression as well as the phosphorylation pattern of 2D and 3D derived hepatocytes were benchmarked against fetal liver and PHH.

3 Material and Methods

2D Hepatocyte differentiation

The *in vitro* generated 2D hepatocytes used for this experiment were kindly provided by our collaborators from the research group of Dr. Keisuke Sekine (Department of Regenrative Medicine, Yokohama City University, Japan; Laboratory of Cancer Cell System, Tokyo, Japan). Hepatocytes were generated from Ff-I01 cells as described previously [253]. Briefly, stem cells were cultured in StemFit™ Basic03 (Ajinomoto) supplemented with 80 ng/ml bFGF (Fujifilm Wako Pure Chemical). For initiating differentiation, the cells were dissociated with Accutase® (Sigma-Aldrich) and plated on laminin 511-E8 (iMatrix-511, Nippi)-coated dishes in the presence of the ROCK inhibitor Y-27632 (Fujifilm Wako Pure Chemical). Cells were differentiated in RPMI1640 (GIBCO™) supplemented with 20% StemFit™ For Differentiation and 100 ng/ml activin A (both Ajinomoto) for the first 6 days. Additionally, 2 µM CHIR99021 (Cayman Chemical) was supplemented during the first three days and 0.5 mM sodium butyrate (Sigma-Aldrich) was added from day 1 to day 4. This step led to the expression of the DE marker CXCR4 in almost 100% of the cells [253]. Next, cells were differentiated in StemFit™ Basic03 (Ajinomoto) medium supplemented with 1% DMSO (Sigma-Aldrich), 0.1 mM 2-mercaptoethanol, 0.5% L-glutamine, and 1% non-essential amino acids (all GIBCO™). During the first 13 days, the medium was exchanged daily and thereafter every second day. For the final 8 days, cells were cultured in DMEM medium (GIBCO™) supplemented with 5% StemFit™ For Differentiation (Ajinomoto) and 0.1 µM dexamethasone (Sigma-Aldrich). Samples were harvested at day 0 (iPSC), day 6 (DE), day 10 (HE), day 13 (IH), and day 21 (MH), and washed twice with PBS (w/o $CaCl_2$ and $MgCl_2$). The cell pellet was stored at -80°C until further processing for proteomics analysis.

Differentiation of endothelial cells (ECs) and mesenchymal cells (MCs)

The *in vitro* generated ECs and MCs used for this experiment were kindly provided by our collaborators from the research group of Dr. Keisuke Sekine (Department of Regenrative Medicine, Yokohama City University, Japan; Laboratory of Cancer Cell System, Tokyo, Japan). ECs and MCs were generated from iPSCs as described previously [123]. For EC differentiation, iPSCs were plated onto laminin 511-E8 (iMatrix-511, Nippi)-coated dishes in mesoderm induction medium consisting of DMEM/F-12 (1:1 mixture) medium supplemented with 1% GlutaMAX™, 1% B-27™ (all GIBCO™), 8 µM CHIR99021 (Cayman Chemical), and 25 ng/mL BMP4 (R&D Systems). For initial seeding, 10 µM ROCK inhibitor Y-27632 (Fujifilm Wako Pure Chemical) was added. From day 4 to day 10, cells were cultured in EC Induction Medium consisting of StemPro-34 SFM medium supplemented with 200 ng/ml VEGF (both GIBCO™), and 2 µM forskolin (Cayman Chemical). The medium was exchanged every second day.

For MC/STM differentiation, iPSCs were seeded onto laminin 511-E8 (iMatrix-511, Nippi)-coated dishes and cultured in mesoderm induction medium. After 4 days, cells were cultured with LPM induction medium consisting of DMEM/F-12 supplemented with 1% B-27™, 10 ng/ml PDGFBB, 2ng/ml activin A, and 1% GlutaMAX™ for 2 days. Then, the medium was changed to STM induction medium consisting of DMEM/F-12, 1% B-27™, 10 ng/ml bFGF, 12 ng/ml BMP4, and 1% GlutaMAX™.

3D liver buds generation

The 3D LBs used in this thesis were kindly provided by our collaborators from the research group of Dr. Keisuke Sekine (Department of Regenrative Medicine, Yokohama City University, Japan; Laboratory of Cancer Cell System, Tokyo, Japan). As described previously [325], LBs were generated by seeding HE, EC, and MC cells at a 10:7:2 ratio onto Matrigel (BD Bioscience)-coated dishes. Cells were cultured in a 1:1 mixture of DMEM (GIBCO™) and KBM-VEC1 basal medium (Fujifilm Wako Pure Chemical) supplemented with 2.5% StemFit™ For Differentiation (Ajinomoto) and 0.1 µM dexamethasone (Sigma-Aldrich). For initial seeding, 10 µM of the ROCK inhibitor Y-27632 was added and thereafter half of the medium was exchanged daily.

Adult and fetal liver cells

Adult liver cells were kindly provided from the Universitätsklinikum Leipzig with the donor´s informed consent. PHH were isolated by the research group of Dr. Georg Damm as described previously [326]. Briefly, liver tissues were dissociated using a two-step EGTA/collagenase P perfusion incubation and the PHH cells were separated from the non-parenchymal cells by centrifugation at 50 xg. PHH were subsequently washed twice with PBS (w/o $CaCl_2$ and $MgCl_2$) and the cell pellet was stored at -80°C until further processing.

The human fetal liver samples were provided by the Joint MRC/Wellcome Trust HDBR with ethical approval. In this work, human fetal liver samples from gestation week 16 and 17 were used for the proteomics study.

Protein extraction, digestion and labeling

Cell pellets of PHH and fetal liver samples as well as from the 2D and 3D hepatocyte differentiation were resuspended in lysis buffer containing 8M urea, 40 mM Tris/HCl (pH 7.6), EDTA-free protease inhibitor (cOmplete™, Roche), and 1x phosphatase inhibitor mix (prepared in-house according to the Phosphatase Inhibitor 1, 2, and 3 from Sigma) and stored at -80°C. Cell lysates were thawed on ice and centrifuged with 20.000 xg for 20 min at 4°C. Protein concentration of the supernatants was determined using the Pierce™ Coomassie Bradford solution (Thermo Scientific) according to the manufacturer´s protocol. 70 µg of protein was used for subsequent reduction with 10 mM DTT for 45 min at 37°C and alkylation using 55 mM CAA for 30 min at RT. After diluting the urea concentration below 1.6 M with 40 mM Tris/HCl (pH 7.6), trypsin at a 1:100 enzyme:protein ratio was added for pre-digestion at 37°C and 700 rpm. After 3 h, trypsin was added again at a 1:100 ratio and proteins were digested overnight. The digestion was stopped by adding FA to a final concentration of 1% and the acidified peptides were subsequently desalted by loading onto StageTips [164] (10 C18 disks, Empore™ 3M), washing with 0.1% FA, and eluting with 0.1% FA in 50% ACN. For TMT labeling, desalted peptides were reconstituted in 20 µl 50 mM HEPES buffer (pH 8.5) and mixed with 5 µl of 11.6 mM TMT reagent. After 1 h shaking with 400 rpm at RT, the labeling reaction was stopped by adding 2 µl of 5% hydroxylamine (Sigma-Aldrich). The 11 TMT channels were pooled and the reactions vessel were rinsed with 20 µl of 10% FA in 10% ACN and combined to the pool. The labeled peptides were dried and desalted using SepPak column (Water Corp.). For this, peptides were loaded and washed with 0.07% TFA before they were eluted using 0.07% TFA in 50% ACN.

Phosphopeptide enrichment and off-line fractionation

For phosphopeptide enrichment, desalted peptides were reconstituted in 0.07% TFA in 30% ACN and loaded onto a FE^{3+}-IMAC column (ProPac™ IMAC-10 4x50 mm, Thermo Fisher Scientific) as described previously [173]. The unbound flow-through consisting of non-phosphorylated peptides was dried and stored for off-line fractionation. The bound phosphopeptides were eluted with 0.315% NH_4OH and further fractionated on StageTip [236]. For this, StageTips with 5 C18 disks (Empore™ 3M) were constructed and peptides were eluted wit increasing ACN concentration (5%, 7.5%, 10%, 12.5%, 15%, 17.5%, and 50% ACN). To obtain six fractions, the 5% and 50% as well as the 17.5% and flow-through fractions were combined and subsequently dried.

The non-phosphorylated IMAC flow-through was further deep fractionated via hPH reverse phase fractionation as described previously [170]. Briefly, peptides were reconstituted in 25 mM NH_4HCO_3 (pH 8) and loaded onto a C18 column (XBridge BEH130, 3.5 µm, 2.1x150 mm, Waters Corp.) coupled to a Dionex 3000 HPLC system (Thermo Fisher). Peptides were eluted with an increasing ACN concentration in 25 mM NH_4HCO_3, collected, and pooled to 48 fractions.

Data-dependent LC-MS acquisition

Fullproteome and phosphoproteome were measured in DDA on a nanoflow system consisting of a Dionex 3000 UHPLC+ connected to a Fusion Tribrid mass spectrometer (both Thermo Fisher Scientific). Peptides were dissolved in 0.1% FA (fullproteome) or 0.1% FA with 50 mM citrate (phosphoproteome) and loaded onto an in-house packed trap column (75 µm x 2 cm, 5 µm C18 resin; Reprosil PUR AQ, Dr. Maisch). An in-house packed analytical column (75 um x 40 cm, packed in-house with 3 um C18 resin; Reprosil PUR AQ) was used to separate peptides, which were further injected into the Fusion Lumos and subsequently measured in positive ionization mode. The fullproteome was eluted with a linear 50 min gradient from 8% to 34% LC buffer B (0.1% FA and 5% DMSO in ACN) in LC buffer A (0.1% FA and 5% DMSO in water). MS1 spectra were acquired in the orbitrap with 60,000 resolution, an AGC target of 4e5 charges, a maxIT of 20 ms, and a scan range of 360-1,300 m/z. For the following MS2 scan, charge states between 2 and 6 were allowed with decreasing priority and fragmented via CID before they were recorded in the ion trap with an AGC target of 2e4 charges and a maxIT of 60 ms in rapid mode. The MS3 spectra were acquired from simultaneous precursor selection (SPS-MS3) of 10 precursors that were further fragmented with HCD and measured in the orbitrap with 50,000 resolution. The AGC target was set to 1.2e5 charges and the maxIT to 120 ms.

The phosphoproteome was measured with a slightly different acquisition method and an 80 min linear gradient of LC buffer B. Full scan MS1 spectra were recorded with 60,000 resolution in the orbitrap, a scan range of 360-1,300 m/z, an AGC target of 4e5 charges, and a maxIT of 50 ms. CID was used for precursor fragmentation and subsequent MS2 spectra were acquired at 15,000 resolution in the orbitrap with an AGC target of 5e4 charges and a maxIT of 22 ms. For quantification, an additional MS3 scan was performed in the orbitrap with 50,000 resolution, an AGC target of 1.2e5 charges, and a maxIT of 120 ms. Multi-notch isolation was utlilized for selecting fragment ions of 10 notches with subsequent HCD fragmentation.

Database searching

Fullproteome and phosphoproteome were searched together in separate parameter groups with the Maxquant software [209, 210] (version [1.6.2.3]) and against the human UniProtKB reference list (downloaded 22.07.2013). Unless stated otherwise, default setting were applied. Trypsin was defined as protease and up to two missed cleavages were allowed. Carbamidomethylation was set as a fixed modification and oxidation of methionine as well as N-terminal protein acetylation were selected as variable modifications. For the phosphoproteome, serine, threonine, and tyrosine phosphorylation were additionally set as variable modification. TMT11-plex was defined as the quantification type and the corresponding correction factors were specified.

Fullproteome data processing

Quantitative fullproteome analysis was performed on the proteingroups.txt output file. Before data was normalized, reversed hits and protein entries that were only identified by site were removed. The reporter ion intensities were log2 transformed and the median of each TMT channel was normalized to the averaged median of all channels (median centering). Then, batch effects between replicates were removed with ComBat [261] from the ´sva´package (version [3.30.1]). These normalized reporter intensities were used for further downstream analyses. The PCA analysis was compiled from the 'factoextra' package (version [1.0.7]) and plotted with the 'ggplot2' package (version [3.3.2]). To determine variance, ANOVA testing with BH correction was performed and proteins with a fold-change >2 and a FDR <0.05 were classified as significantly altered.

Phosphoproteome data processing

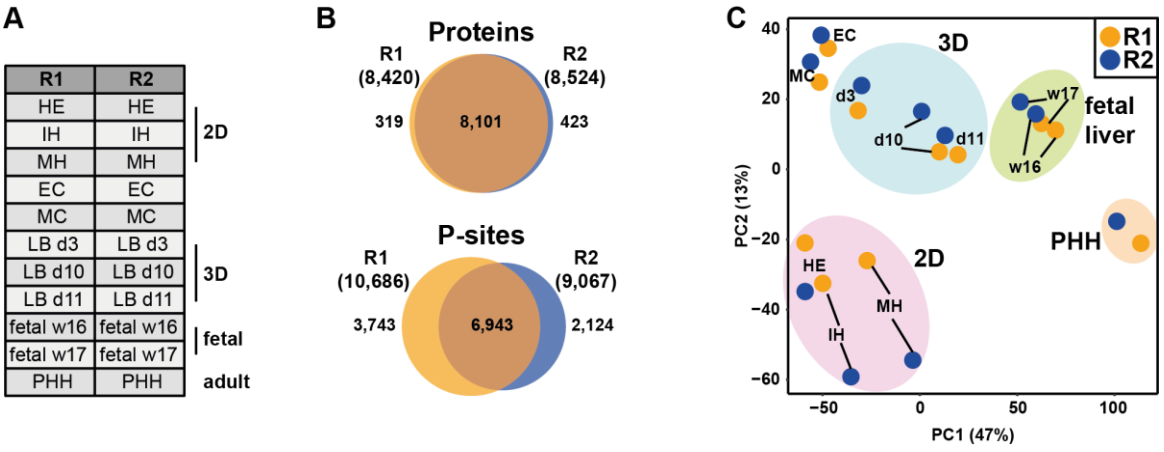
The phosphosite.txt was used for quantitative phosphoproteome analysis. Reverse hits and entries with a localization probability <0.75 were removed for further processing. Reporter intensities were log2 transformed and median-centered using the correction factors calculated from the corresponding fullproteome data. As for the fullproteome, batch effects between replicates were removed by ComBat and significantly regulated P-sites were calculated (ANOVA: FDR <0.05 and fold-change >2). For kinase-substrate relationships, the networkin prediction tool was implemented [264].

4 Results and discussion

In order to assess the hepatocyte differentiation of 2D and 3D culturing models, their proteomic and phosphoproteomic profiles were compared with profiles from *in vivo* fetal liver and PHH.

4.1 *In vitro* models generate rather immature hepatocytes

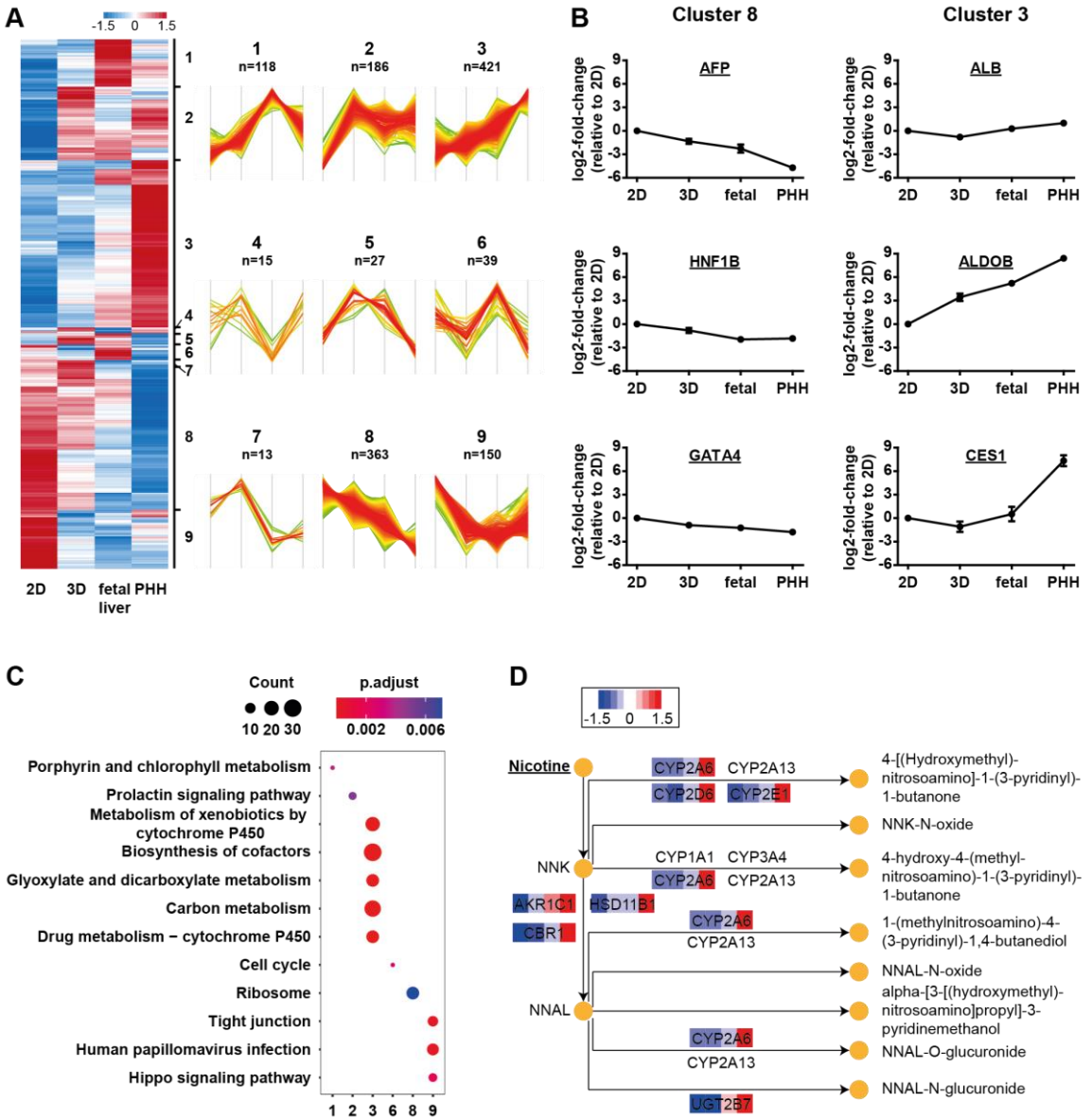
In the previous chapter, 2D monolayer-derived hepatocytes were examined to extrapolate molecular changes during epithelial cell differentiation. Apart from this approach, an enhanced culturing method for hepatocyte generation are LBs. These are 3D organ-like structures that consist of multiple cell types which self-organize upon co-culturing. For the LBs generated in the following study, cells from the HE, endothelial cells (EC), and mesenchymal cells (MC) were individually differentiated from iPSCs and then combined for hepatocyte maturation. The 2D and 3D derived hepatocytes for this experiment were generated in collaboration with the research group of Keisuke Sekine from Japan. For a direct comparison, proteomes from the 2D differentiation (HE, IH, MH), 3D differentiation (MC, EC, and LBs after pooling at day 3, 10, and 11), fetal liver (from week 16 and 17 after gestation), and PHH were gathered (IV-Figure 1A). The experiment was performed as dublicate (R1 and R2) with the previously described workflow (III-Figure 1). The deep (phospho)proteomic measurements led to 8,800 identified proteins and 12,700 P-sites (IV-Figure 1B). PCA analysis of the overlapping 8,100 proteins of both replicates revealed a clear separation between 2D, 3D, fetal liver, and PHH along PC1 and PC2, which implied that differences on protein levels are sufficient to discriminate the samples (IV-Figure 1C). Furthermore, the direction of 2D and 3D differentiation stages indicated a temporal maturation towards fetal liver and PHH, which suggested a successful development towards functional hepatocyte-like cells.



IV-Figure 1: Quality control of differentiation. (A) Overview of the samples used for the *in vitro* versus *in vivo* comparison. EC, endothelial cells; MC, mesenchymal cells. (B) Venn diagrams showing the number of identified proteins and phosphorylation-sites (P-sites) of both replicates, respectively. (C) PCA of the proteomics experiment.

Next, the MH, LB d11, fetal liver (average of week 16 and 17), and PHH were analysed in more detail and the differentially expressed proteins (ANOVA, BH corrected: FDR <0.05 and FC>2) were

subselected from the dataset. These 1,332 proteins were z-scored and hierarchically clustered leading to 9 distinct expression profiles of which cluster 3 and 8 comprised the majority of proteins (IV-Figure 2A). Cluster 8 contained progenitor proteins like AFP, hepatocyte nuclear factor 1-beta (HNF1B), and transcription factor GATA4, which were elevated in the 2D samples and decreased towards PHH (IV-Figure 2B, left panel). This suggested incomplete maturation of the 2D samples, as these proteins are associated with the developing liver rather than mature hepatocytes. In contrast, the expression of cluster 3 proteins was highest in PHH and comprised several liver-specific proteins, including albumin (ALB), fructose-bisphosphate aldolase B (ALDOB), and the liver carboxylesterase 1 (CES1). While the abundance levels of ALB did not diverge substantially between the samples, ALDOB and CES1 were 8-300 times higher expressed in PHH compared to the other samples (IV-Figure 2B, right panel). In addition, KEGG analysis revealed that multiple proteins related to metabolic pathways were enriched in cluster 3 (IV-Figure 2C). Among them were several pathways associated with cytochrome P450, which is a family of enzymes that is highly abundant in the liver and very important for drug metabolism. One significantly upregulated member of this family was for example CYP2A6, which is the primary enzyme involved in nicotine metabolism (IV-Figure 2D).

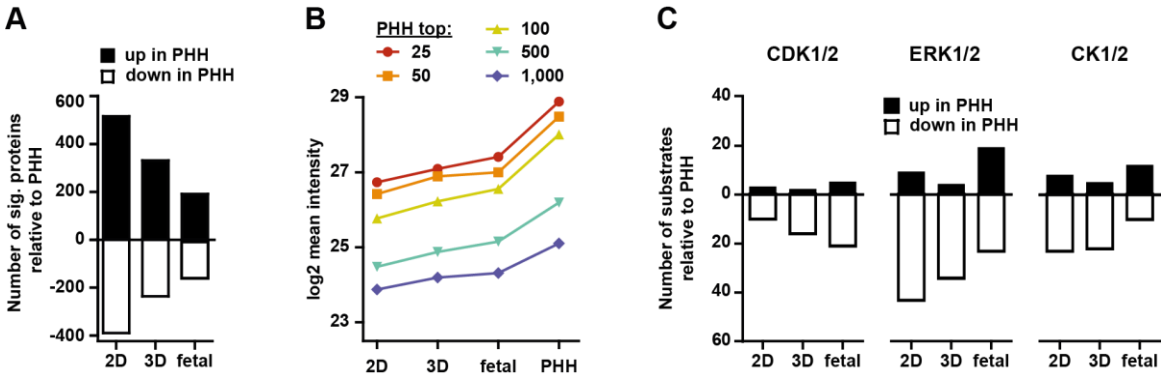


IV-Figure 2: Elevated levels of metabolic pathways in PHH. (A) The left panel displays the z-scored heatmap of hierarchical clustering of the 1,332 differentially expressed proteins (ANOVA, BH corrected: FDR<0.05 and FC>2). The right panel shows the expression profiles of proteins in all 9 clusters (n indicates the number of proteins in the cluster). (B) Scatter plots showing the protein expression of selected members from cluster 8 and 3 relative to the 2D differentiation. (C) KEGG enrichment analysis of the protein clusters from (A). The top five enriched pathways are shown. (D) Z-scored expression of proteins related to nicotine metabolism, which is part of the ´metabolism of xenobiotics by cytochrome P450´ from (C).

In summary, this analysis revealed that several progenitor markers were higher expressed in the 2D model than in the 3D model and the *in vivo* samples with a similar expression pattern of 3D and fetal liver samples. On the other hand, multiple liver-specific proteins were only highly expressed in PHH.

4.2 3D liver buds are superior to 2D hepatocytes

Since the above analysis indicated incomplete hepatocyte differentiations for either of the two models, a more detailed analysis was employed. To assess differences between the samples, the number of significantly altered proteins (ANOVA, BH corrected: FDR<0.05 and FC>2) relative to PHH were determined (IV-Figure 3A). Decreasing numbers of differentially expressed proteins from 2D to 3D to fetal liver suggested a convergence towards PHH. This trend was further confirmed by comparing the intensities of the top 25-1,000 most abundant proteins of PHH across all other samples (IV-Figure 3B). These analyses indicate that the 3D model better resembles PHH than the 2D model, albeit further maturation steps are required to better recapitulate the *in vivo* counterpart.

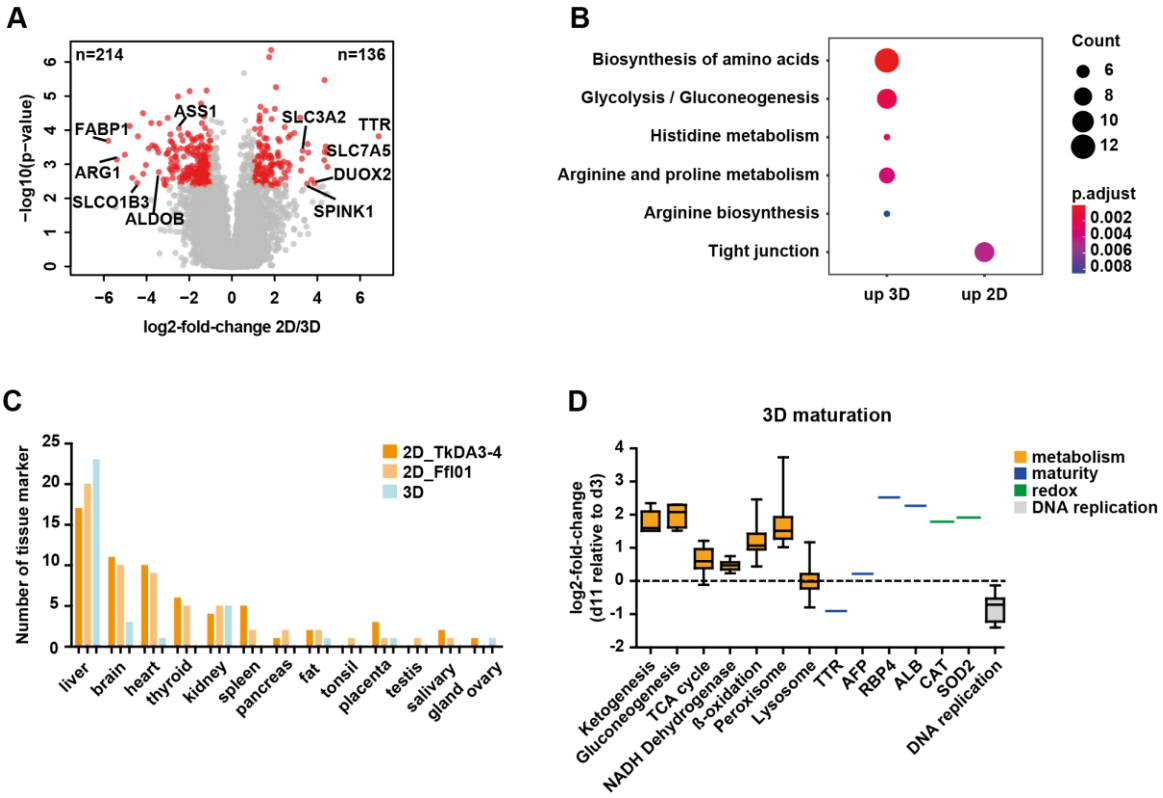


IV-Figure 3: Protein expression of 3D samples resemble PHH better than the 2D approach. (A) Number of significantly regulated (ANOVA, BH corrected: FDR<0.05 and FC>2) proteins relative to PHH. (B) Line plot showing the mean intensity of the top 25 to 1,000 PHH proteins in mature 2D, 3D, and fetal liver. (C) Number of significantly regulated (ANOVA, BH corrected: FDR<0.05 and FC>2) downstream substrates of CDK1/2 (left panel), ERK1/2 (middle panel), and CK1/2 alpha (right panel) relative to PHH.

To add additional information to the protein expression data, phosphorylation patterns were explored by analysing kinase-substrate relationships. This analysis was performed with Networkin [264], which integrates known and predicted relationships based on kinase recognition motifs. Previously, this type of analysis revealed a trend of decreasing CDK and ERK activity during the 2D hepatocyte differentiation (III-Figure 6D and 6E, III-Figure 8E). Indeed, the comparison of all samples further suggested low CDK1/2 and ERK1/2 activity in functional hepatocytes, as many substrates were downregulated in mature PHH (IV-Figure 3C, left and middle panel). A similar trend was also observed for the casein kinase (CK) 1 and 2 alpha (Figure 3C, right panel), which belong to the casein kinase family that is primarly implicated for its role in DNA repair, DNA transcription, circadian rhythm, and cell cycle control [327-330].

To elucidate on the divergence between the 2D and 3D models in more detail, a differential expression analysis was performed (IV-Figure 4A). The solute carrier SLCO1B3, which was among the 214 proteins upregulated in the 3D LBs, is for example involved in the transport of bile acids [331] and steroid conjugates in the liver [332]. Besides that, multiple metabolic pathways were enriched in the 3D samples according to KEGG analysis (IV-Figure 4B). For example, two key enzymes of the urea cycle, arginosuccinate synthase (ASS1) and arginase-1 (ARG1), were elevated

more than 5 or 35 times, respectively. Additionally, the previously mentioned liver specific aldolase (ALDOB) as well as the fatty acid-binding protein (FABP1), which plays an important role for cholesterol uptake in hepatocytes [333], were among the most upregulated proteins in LBs (IV-Figure 4A). In contrast, proteins upregulated in the 2D samples suggested that the differentiation towards hepatocytes was incomplete and a more heterogeneous cell population was obtained. For example, with transthyretin (TTR), dual oxidase 2 (DUOX2), and the two transporters SLC3A2 and SLC7A5 several proteins associated with the two thyroid hormones T3 and T4 were highly increased in the 2D differentiation [334-337]. In addition, SPINK1, a serine protease from the pancreas, was highly upregulated in the 2D samples. Since the thyroid gland as well as the pancreas share their endodermal origin with the liver, an incomplete commitment from the endoderm towards hepatocytes can be suspected. To elucidate on the differentiation specificity, upregulated proteins between d11 *versus* d3 of the LBs and MH *versus* HE of the 2D dataset were compared to a number of tissue-specific marker proteins. The marker list was retrieved from Wang *et al.* [153] where protein expression data of 29 human tissues were acquired in order to define their tissue specific abundance. By far the highest number of upregulated markers in the LBs were liver-specific (IV-Figure 4C). While also at the end point of the 2D differentiation mainly liver-specific markers were upregulated, the number of non-liver tissue-specific markers was higher than in LBs. While this comparison further illustrated the successful differentiation into liver-like cells, this analysis also confirmed the previous notion that the 3D approach is more specific leading to less lineage bifurcation.



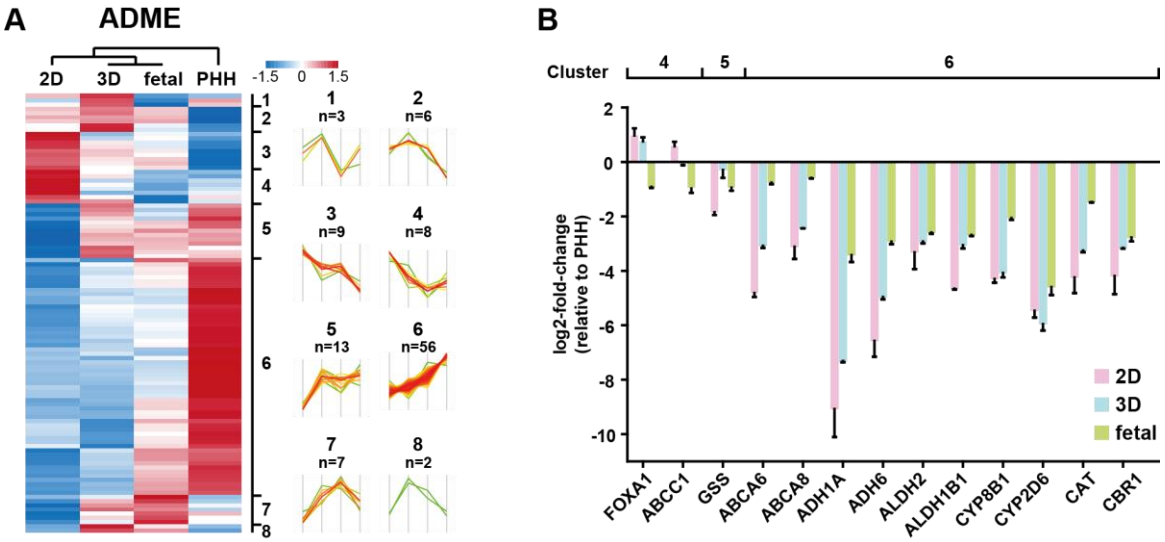
IV-Figure 4: Comparison of 2D and 3D-derived hepatocytes. (A) Volcano plot illustrating significantly regulated proteins (in red; ANOVA, BH corrected: FDR<0.05 and FC>2) between the 2D and 3D hepatocyte differentiations. (B) KEGG enrichment of proteins significantly up or downregulated from (A). Top five

enriched pathways are shown. (C) Bar plots depicting the number of upregulated tissue specific proteins between MH and HE for the 2D or LB d11 and LB d3 for the 3D differentiation. (D) Boxplots showing the fold-change (day 11 to day 3) of selected metabolic pathways, maturation marker, redox-reducing enzymes, and proteins related to DNA replication. For metabolic pathways and DNA replication, the fold-change of all associated proteins were taken into account.

Despite differences between both *in vitro* models, the main concepts that have been described from the previous 2D differentiation, such as the metabolic switch between the HE and IH stage (III-Figure 4 and III-Figure 8), could be confirmed in the 3D model. Also, essential hepatic pathways, like the TCA cycle, ketogenesis, and the gluconeogenesis were substantially increasing during the 3D maturation (IV-Figure 4D). Additionally, the early differentiation markers TTR and AFP showed modest changes during the LB maturation, whereas the liver-specific proteins RBP4 and ALB increased by more than 4-fold. Moreover, the expression of important redox-reducing enzymes was elevated and proteins associated with the DNA replication decreased similar as in the 2D differentiation (III-Figure 5, III-Figure 8, and III-Figure 11).

4.3 Expression of ADME/Tox-related proteins of *in vitro* models reflect fetal-like stage

One potential application of *in vitro* generated hepatocytes is the toxicity assessment of therapeutic drugs during their development. Critical for this process are proteins related to absorption, distribution, metabolism, and excretion as well as toxicity (ADME/Tox), which makes them an interesting protein class to analyse. Therefore, ADME/Tox-related proteins were retrieved from Schroder *et al.* [338] and considered for analysis. Differential expression analysis led to 104 ADME/Tox-related proteins that were significantly different (ANOVA, BH corrected: FDR<0.05 and FC>2) in either of the four samples (2D, 3D, fetal liver, or PHH). Further hierarchical clustering of these differentially regulated proteins revealed 8 distinct expression profiles (IV-Figure 5A). Cluster 3 and 4 comprised in total 17 proteins which were almost exclusively elevated in the 2D differentiation (IV-Figure 5B), such as the transcription factor FOXA1, which is essential for DE development and early lineage specification [339]. In cluster 5 the glutathione synthetase (GSS) was included, a protein that catalyses the second synthesis step of glutathione, which is an important antioxidant that is used in multiple assays to assess drug toxicity effects. Proteins that were most abundantly expressed in PHH were mainly found in cluster 6. Of note, Cluster 6 was with 56 members the biggest cluster showing that most ADME/Tox proteins were predominantly expressed in PHH and highlighting the need for further improvements in the current *in vitro* differentiation models. In this cluster various ABC transporters, metabolic proteins, including the alcohol and aldehyde dehydrogenase, and multiple liver-specific proteins were present (IV-Figure 5B). In addition, several proteins from the cytochrome P450 family were contained in cluster 6. Lastly, enzymes reducing ROS, such as CAT and CBR1, were highly abundant in PHH. As outlined in chapter 4.2 on page 62, this could be due to a high aerobic energy consumption of metabolic proteins driving the generation of ROS.



IV-Figure 5: Highly expressed ADME/Tox-related proteins in PHH. (A) Heatmap showing hierarchical clustering of all 104 significantly regulated proteins associated with ADME/Tox. (B) Bar graph of the relative expression of protein clusters from (A) normalized to PHH. Bars denote mean values and the error bars represent the range.

In conclusion, the majority of ADME/Tox proteins were siginificantly lower expressed in the *in vitro* samples and in the fetal liver than in PHH, which is in line with previous results claiming that PSC-derived hepatocytes rather mimic fetal than adult hepatocytes [258]. While the expression of FOXA1, ABCC1, and GSS were rather similar among all samples, multiple proteins from cluster 6 were up to 1,000 times more abundant in PHH. Therefore, additional experiments directly assessing the metabolization of distinct metabolites in iPSC-derived hepatocytes and PHH are required prior to a broader use of iPSC-derived hepatocytes in drug toxicity assessments.

4.4 Charting *in vitro* hepatocyte differentiation and future optimization strategies

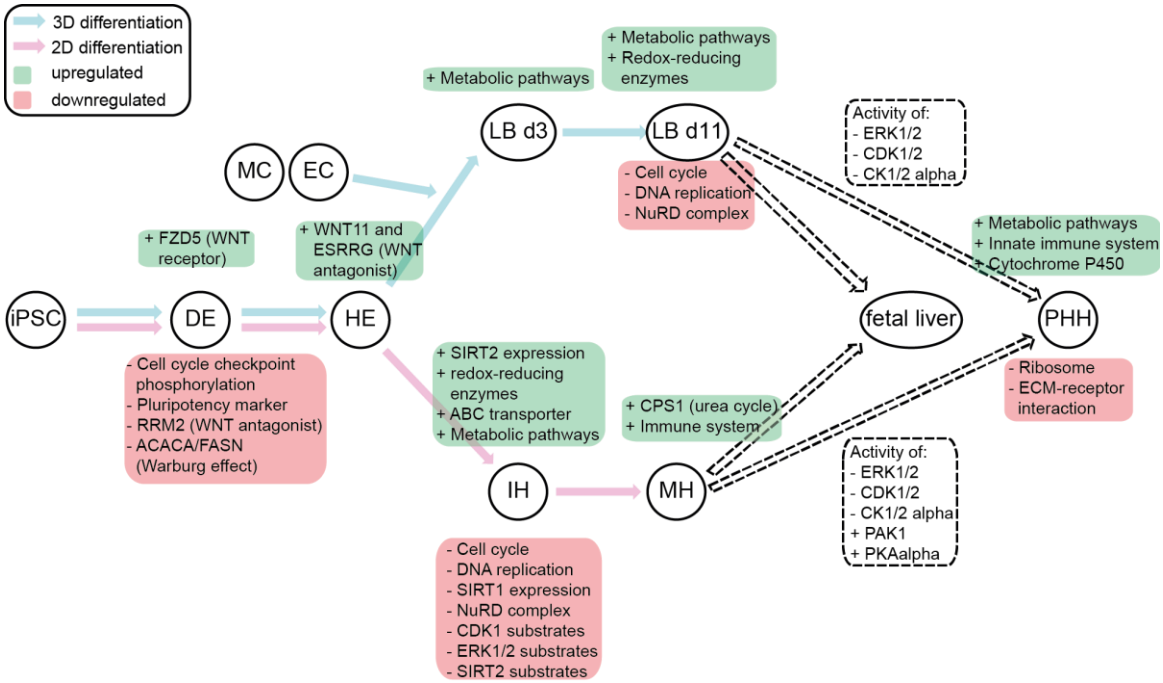
In the following section the molecular roadmap of *in vitro* hepatocyte differentiation will be delineated with a special emphasis on deducing strategies for future protocol improvements (IV-Figure 6).

Stem cells as well as cancer cells are rapidly dividing cells [266-268]. In agreement with a high demand for cellular membranes [340, 341], the fatty acid synthase (FASN), a key enzyme for *de novo* lipogenesis, was the most abundant protein in the TkDA3-4 cells and among the top20 in Ff-I01 cells. Upon differentiation, FASN levels dropped significantly at the DE stage. A similar expression pattern was observed for the acetyl-CoA carboxylase 1 (ACACA), the rate-limiting enzyme for *de novo* fatty acid biosynthesis. These results suggested an immediate change of fatty acid metabolism, similar to the upregulation of FASN during neural stem cell differentiation [342]. However, the most striking metabolic switch happened between HE and IH. Putative key regulators for this switch might be the two sirtuin enzymes SIRT 1 and 2. While high levels of SIRT1 have been positively correlated with pluripotency [281, 287], the reduced expression has

promoted differentiation of stem cells [284, 288]. In agreement with reduced cell proliferation during differentiation, low SIRT1 activity has been associated with accumulation of fatty acids and carbohydrates [283]. On the contrary, SIRT2 overexpression has been associated with spontaneous cell differentiation and diminished activity of glucolytic enzymes caused by deacetylation [281]. The detailed proteomic expression analysis in this thesis revealed a simultaneous switch of both proteins after the HE stage, in which SIRT1 levels decreased significantly and SIRT2 showed a congruent increase. These diametrically opposing expression profiles appeared to be highly regulated in a time-dependent manner in both iPSC lines. The high temporal resolution of the experiments increased the understanding of processes during hepatocyte development and generated new potential starting points for protocol improvements. For example, the temporal dissection of the SIRT1 downregulation suggests that the supplementation of an SIRT1 inhibitor at this stage might increase differentiation efficiencies. This is of special interest since the supplementation of SIRT1 inhibitors, for example, have promoted neural development and hematopoiesis before [282, 284, 343].

Moreover, unbiased expression analysis led to multiple novel and stage-specific marker proteins. Among these, various proteins have been associated with Wnt signalling, a known and essential pathway for cell development [98, 344]. One novel marker was the Wnt receptor frizzled-5 (FZD5) showing high protein levels at the DE stage with a subsequent drop during further differentiation. Although additional experiments are needed to explain the specific role of FZD5 in more detail, this striking expression profile suggested a key role for this development stage. More experiments are also required to validate the influence of the observed stage-specific upregulation of the Wnt inhibitors WNT11, ESRRG, and LRP4 and the assumption that Wnt signalling is not sustained after the DE stage. These coordinated expression changes in the Wnt signalling pathway illustrate a potential starting point for protocol optimization. Several studies have shown the benefit of using Wnt activators to guide cell development, such as the addition of Wnt3a for DE specification [97, 103].

Only very few studies have investigated phosphorylation of cell differentiation so far [62, 63, 256] and none has specifically focused on hepatocyte differentiation. Hence, in this global phosphoproteomic study multiple interesting findings complementing transcriptomics and proteomics data were retrieved. One example was the temporal unravelling of protein expression and phosphorylation levels of proteins related to the cell cycle, which suggested a preceding switch observed on phosphorylation level. Consistent with previous studies the CDK activity was elevated in PSCs [345, 346] and decreased upon differentiation. However, the activity was still higher than in the *in vivo* samples. Several studies have shown before that CDK1 activity is crucial to maintain pluripotency [346-349] and thus, it appears plausible that a partial downregulation or inhibition of CDK1 promotes the differentiation process.



IV-Figure 6: Roadmap of hepatocyte differentiation. Overview of multiple proteins and pathways significantly changing during *in vitro* hepatocyte differentiation. Dashed lines indicate pathways with distinct activity compared to the *in vivo* liver samples.

Another interesting set of targetable kinases that emerged from the analysis are ERK1/2, because they were linked to early lineage specification before [350-353] but their complex role has remained inconclusive. The protein expression analyses revealed increasing abundance levels of ERK during the differentiation, whereas the phosphorylation indicated decreased activity. In addition, the corresponding substrate phosphorylation decreased after the HE stage confirming that ERK activity was not sustained in later stages. However, compared to *in vivo* liver samples the activity was still elevated. A similar trend was observed for CK1/2 activity, which illustrates potential targets as ERK and CK inhibitors have previously been used to support differentiation [354-356] but not in the context of hepatocytes. Although additional experiments are needed to validate these findings, a recently published study [357] reported the beneficial application of CK1 inhibitors for differentiating pancreatic progenitors. As pancreas and liver share the DE as a common origin, the utilization of ERK and CK inhibitors might be a good starting point to improve *in vitro* hepatocyte differentiation.

Chapter V: General discussion and outlook

Omics technologies to study cell line differentiation - merits and limitations

The first systematic study of human embryonic development emerged in the 19th century when researchers started to collect and investigate human embryos derived from pregnancy abortions and maternal deaths. Big collections, such as the Carnegie Collection (more than 10,000 specimen) or the Kyoto Collection (more than 44,000 specimen), set the base for the understanding of early human development (reviewed in [358]). Their descriptive and morphological studies helped to gain insights into sequential developmental processes and to classify different embryonic stages. Since then, the possibilities of studying development has changed drastically. On the one hand, the ability to isolate and culture human hESCs in 1998 [5] as well as the discovery of reprogramming iPSCs in 2006 [7, 8] has faciliated the study of human development *in vitro*. On the other hand, the methods to follow cell lineage specification have developed from descriptive studies towards interventional studies allowing the dissection of the underlying molecular mechanisms.

Proteomics as the preeminent technology among protein-centric approaches

Fluorescence-activated cell sorting (FACS), western blot, and immuno staining are widely used technologies to analyze protein abundance during hepatocyte differentiation [123, 253, 254]. However, since these approaches are all based on antibodies, they have some drawbacks, such as their limitation to the low number of proteins for which specific antibodies exist or the low throughput of proteins that can be analysed in an experimental setting. Moreover, these semiquantitative approaches require prior knowledge or hypotheses of candidate proteins and thereby hamper the identification of unexpected protein hits. Mass spectrometry-based proteomics overcomes some of these limitations as it provides an unbiased way to identify and quantify thousands of proteins in one experiment. Hence, the analysis is not limited to a few protein markers but provides a global picture of protein expression, which is especially interesting for the comparison of different cell types. The comparison of *in vitro* versus *in vivo* liver samples in this thesis (chapter IV) showed how important such a thorough analysis can be. Although multiple hepatocyte-specific proteins were upregulated in the *in vitro* samples, the global protein expression allowed the discrimination in a fetal-like and an adult liver state suggesting that the current *in vitro* differentiations do not completely recapitulate the *in vivo* development.

A second example where high measuring depth proved useful was the study of *in vitro* hepatocyte differentiation with high temporal resolution (chapter III). Analysis of multiple time points during this process did not only allow validation of differentiation based on the expression of known protein markers, but also generation of a list of novel stage-specific proteins. These novel marker proteins can be used as validation sets for future differentiations or to enrich cells from a specific development stage using FACS. Moreover, the close investigation of such marker proteins will help to elucidate developmental concepts as shown in this thesis. For example, the substantial upregulation of FZD5 upon activation of the WNT pathway indicated that this is a crucial WNT receptor for DE specification. Several examples like this are presented in this work showing how proteomics can be utilized to retrieve new concepts and hypotheses. Although the demonstrated findings are based on statistically significant expression changes, further follow-up experiments

are desirable to test causality and functional relevance. Thus, the protein changes during hepatocyte differentiation were confirmed with a second iPSC line to exclude cell line-specific effects. Interventional follow-up experiments including an ablation of FZD5 in iPSCs would for example help to answer the question how important the single WNT receptor for proper DE specification is.

In comparison to genomics or transcriptomics, one drawback of protein-based studies is their inability to *in vitro* amplify analytes, which limits the detection of peptides/proteins to the sensitivity of the mass spectrometer. Therefore, proteins with low abundance may not be detected, although they could be of biological importance. Transcription factors (TFs) belong to such a key group of proteins that can define cell fate decision even at low abundance, making their study by proteomics difficult [359]. This issue is further enhanced by the DDA approach of bottom-up proteomics, as the selection of ions picked for fragmentation and subsequent MS2 scans is based on their MS1 intensity. Thus, high abundant ions/peptides are more likely to be measured. However, it is not always the low abundance of proteins that hamper their detection, but also the high abundance of co-eluting peptides that can mask the identification [360]. In this work, peptides were extensively fractionated prior to MS measurement to mitigate these issues. By fractionation, the sample complexicity is reduced and thus, the number of identified peptides is increased substantially [168]. However, for thorough sample fractionation, sufficient amount of starting material is required. Because large-scale cultivation of stem cells is demanding and supplements for differentiation are expensive, sample amounts of cell differentiation experiments are often scarce. Hence, in the first part of this thesis the downscaling of input material for proteomic sample preparation was optimized, which led to the identification of more than 10,000 peptides from just 2,000 cells (Chapter II). The optimized iST and SP3 beads approaches showed very promising results without requiring much additional equipment, making these workflows attractive for sample-limited applications.

Transcriptomics versus proteomics for studying cell differentiation

The amplification of mRNA in transcriptomics workflows allows the sensitive detection of the global transcriptome at a very high throughput and with great coverage. In the published scRNA-seq data integrated in this work [100], around 17,000 gene transcripts were covered in total. While a median of 5,500-7,500 transcripts were detected in a single cell, around 8,100 transcripts were covered across all five developmental stages. Interestingly, just 81 transcripts were identified in every single cell illustrating that missing values among single cells is a relevant drawback of this method. In contrast, the proteomics approach in this thesis prevented most missing values by labelling peptides with the isobaric TMT. This multiplexing strategy enables the concurrent identification and quantification of peptides from several samples, which reduces the issue of stochastic sampling of DDA and thus, decreases missing values within a single TMT batch. Thereby, for the vast majority of measurements either a peptide is detected in all TMT channels or it is missing completely. With this approach around 9,000 proteins were covered across all five time points, which is comparable to the 8,100 transcripts detected in the scRNA-seq dataset. Despite the comparable depth of proteomics and transcriptomics achieved in this thesis, *in vitro* differentiation protocols have been still mainly driven and characterized at the transcript level. Future PSC-based model systems will therefore highly benefit from focusing on proteins as the

functional unit of the cells instead of sole transcriptomic analysis. This becomes even more evident in view of a rather poor temporal correlation of mRNA and protein abundances as shown in this work (chapter III) and in previous studies [140, 153]. Interestingly, the correlation between the two most separated time points, iPSCs and MH, was acceptable indicating that mRNA levels can be used to infer overall protein abundances to some extent, but fail to delineate the exact temporal expression. Besides an inherent chronological sequence of transcription and translation, the diverging dynamics might be partly due to transcripts that are actually not translated into proteins and due to the on average 5 times higher stability of proteins compared to mRNA [361]. Especially in the dynamic environment of cell fate decisions, a high degree of fluctuations of mRNA might only partly mirror the actual presence of functional proteins in the respective developmental stage. This notion is further supported by the preceding decrease in transcript levels of proteins related to DNA replication (Chapter III). Nevertheless, the correlation between mRNA and protein levels is complex and several other factors, such as the vastly different dynamic range also contribute to the observed discrepancy. While transcripts quantified by RNA-seq span about four orders of magnitude, protein abundances recorded by MS span around eight orders of magnitude [153], further emphasizing the difficulties in direct correlation of the two omics approaches. To this notion, Schwanhausser and colleagues have shown that just 40% of the variance on protein levels can be explained from mRNA levels [361]. Thus, other processes, such as mRNA degradation or protein turnover, also influence actual protein levels in a cell. The buffering effect at protein level is yet another process that complicates this correlation. This effect describes the observation that the abundances of some proteins are not affected even with substantial changes on transcript level [140, 362]. Moreover, the cell lysate-based proteomics approach utilized in this study does not detect secreted proteins and thus biases the correlation to proteins residing in the cell. In any case, future developmental studies would benefit from the complementary study of secreted proteins as autocrine and paracrine signalling are crucial mediators of cell fate decisions.

In addition to detecting protein expression levels, mass spectrometry-based proteomics is also capable of analysing PTMs, which allows to derive information about the functionality of proteins, such as their enzymatic activity. This is of particular interest because the activity of proteins cannot always be inferred from their higher expression levels. So far only a limited number of studies exist that analysed PTMs in the context of cell differentiation on a large scale, although many PTMs regulate essential signalling pathways. For example, some studies have investigated phosphorylation changes during early stem cell differentiation, which revealed insights about kinases and signalling pathways involved in maintaining pluripotency as well as general differentiation processes [62, 63, 256]. Yet, so far a global survey of PTMs in the context of hepatocyte differentiation has been lacking. As discussed previously, sample amounts from differentiation experiments can be scarce, which is even more challenging for the study of PTMs as they usually require specific enrichment strategies due to their low abundance [363]. In this study, phosphorylated peptides were enriched, yielding around 12,000 P-sites, from limited sample material. However, others have identified more than 37,000 P-sites from 10 times more peptide quantity [174], demonstrating that large amounts of starting material are essential for comprehensive phosphorylation analyses. Although technological advances in the field of mass spectrometry-based proteomics have even led to the identification of nearly 300,000 distinct P-

sites [364] (PhosphoSitePlus® 12.01.2022), data mining remains one of the major bottlenecks for phosphoproteomics approaches. For more than 95% of these distinct P-sites no known kinase or biological function has been reported [365]. Furthermore, just 20% of the known kinases are responsible for phosphorylating more than 80% of the currently annotated substrates. This imbalance is presumably caused by the lack of highly selective tool compounds for most kinases as well as the bias of research studies towards a limited set of kinases [365]. However, as kinases and phosphatases recognize their substrates to some extent by specific amino acid sequences or motifs, algorithms can overlay this information to predict kinase-substrate relationships. In this thesis, the Networkin platform [264] was utilized to report known and predicted downstream substrates of CDK and ERK, as their phosphorylation pattern suggested significantly altered kinase activity. Indeed, the decreased phosphorylation of downstream substrates supported this finding. However, it is important to note that such an analysis requires the knowledge of the specific kinase motif, which again leads to a bias towards the low number of well-studied kinases.

Considering acetylation events, the availability of enzyme-substrate information in public databases is very poor. One approach to shed more light on HDAC substrates was the study of acetylation changes after treating cells with 19 different HDAC inhibitors [366]. Although high inhibitor selectivity is required to infer the exact enzyme-substrate relationships, such experiments are important starting points. A similiar approach analyzes the acetylation response upon knockdown of writer (acetylation) or eraser (deacetylation) proteins. By retrieving results from such studies [281, 285], some of the acetylation events in this work could be linked to SIRT2 and the observed metabolic switch. Since the biological function for most Ac-sites remains elusive, elaborated follow-up experiments are required. Therefore, the global analysis of acetylation dynamics are important resources, which will become increasingly valuable as soon as the functional annotations of Ac-sites can be improved.

Nevertheless, compared to single-cell transcriptomics, comprehensive proteomics approaches are so far limited to analyse the expression of bulk experiments. Indeed, the information from each single cell is lost by the proteomics approach because the averaged signal from all cells is recorded. Considering that *in vitro* differentiations often result in heterogenous cell populations, single-cell profiles would be however highly desirable. Therefore, the information obtained from the scRNA-seq experiment was exploited to investigate cell heterogeneity at the single cell level. Almost all cells commited successfully to the DE lineage and around 90% expressed IH specific marker genes. Only the HE and MH stages showed slightly higher heterogeneity [100]. Hence, the high differentiation efficiency allowed the measurement of protein expression of the target cells by bulk proteomics. Indeed, the high temporal separation in the PCA analysis derived from the proteomics data indicated that the differentiation purity as well as the analysis depth facilitated the dissection of the different development stages.

To conclude, both technologies have their pros and cons. While the high sensitivity of transcriptomics allows the study of single cells with good coverage, the measurement of the actual functional unit in proteomics better reflects the physiological function of the investigated cells. Taken together, both methods should be implemented in a complementary way for the investigation of lineage decisions and cellular functions.

Single-cell proteomics - future technology to study development?

Since the idea of single-cell proteomics emerged, it has mostly been devoted to antibody-based experiments. For example western blot analysis provides enough sensitivity to detect protein expression of single cells, but only for a very limited number of protein targets [367]. The same is true for fluorescence-activated cell sorting (FACS), which is a widely used technology to analyse and sort single cells. Combining the single cell resolution of flow-cytometry with MS led to the development of the CyTOF (cytometry by time-of-flight mass spectrometry). This technology has emerged in the field of single-cell proteomics in 2008 [368] and is based on the quantitative measurement of metal-conjugated antibodies (as reviewed in [369]). This technology has diminished one of the main drawbacks of FACS, the spectral overlap of fluorescent antibodies, and allows the simultaneous detection of more than 40 different proteins at once [370]. While these approaches are restricted in the numbers of protein targets that can be screened in parallel, several thousand cells can be analysed per second. However, major drawbacks of these technologies are the dependency on antibody availability and specificity as well as their targeted nature. In contrast, mass spectrometry-based approaches that do not rely on antibodies can study a much larger number of proteins, but with lower throughput due to the very time-consuming data acquisition. One attempt to overcome this limitation was SCOPE-MS (Single Cell ProtEomics) introduced by the Slavov group [371]. Here, the throughput was increased by multiplexing single cells with TMT, which enables the simultaneous analysis of up to 16 samples at the same time [372]. Moreover, a so-called carrier or booster channel, a sample with up to a few hundred or even a thousand times the protein amount of a single cell, is added to increase the signal intensity and thereby the identification rate of peptides. However, this approach is discussed controversial [373] because the dynamic range of mass spectrometers is limited and the signal from the carrier proteome overlays the single cell signals, which negatively affects the quantification accuracy (reviewed in [373, 374]). Another way to enhance sensitivity of single-cell proteomics is to improve methods for sample preparation, such as the nanoPOTS (nanodroplet processing in one pot for trace samples) technology [375], which utilizes a pipetting robot that is capable of handling nanoliter droplets. Thereby, the loss during sample preparation can be decreased, which resulted in the quantification of around 670 proteins from single cells [376]. Since its first publication in 2018, the idea of nanoPOTS was further developed and is now a commercially available picoliter dispensing platform from Cellenion (cellenONE F1.4, Cellenion, France). This system allows to perform the complete sample preparation from sorting and lysing of single cells to the digestion of proteins, and finally the labelling with TMT. Coupling this platform with a nanoPOTS autosampler [377] allows the direct injection of samples from the nanowells to the LC-MS system, thereby increasing sensitivity and throughput for single-cell analysis. Employing this workflow to 100 single cells led to the robust quantification of around 1,500 proteins from each single cell [378]. Although a carrier proteome was used in this study, the miniaturization and automatization is very promising for future single-cell proteomics applications. Apart from the sample preparation, several developments on the LC-MS side have also improved sensitivity. For example the FAIMS (Field Assymmetric Ion Mobility Spectrometry) device, which is able to remove singly charged background contaminants, gained 10 times more sensitivity compared to previous LC-MS setups [379]. With this ion mobility device the ion accumulation time can also be extended

resulting in enhanced signal intensity and thus, increased peptide identification [380]. Furthermore, decreasing the column diameter and flow rate to 20 µm and 20 nL/min, respectively, has shown augmented sensitivity compared to the standard proteomics setup, which is performed with 75 µm columns at 300 nL/min [381, 382].

In recent years, several innovations in sample preparation and instrumentation have increased sensitivity and led to major advances in the field of single-cell proteomics. This area of research has gained more interest, and single-cell analysis has made great strides in feasibility and comprehensiveness. However, some of the general problems of mass spectrometry-based proteomics discussed previously are even more pronounced at the single-cell level. For example, the dynamic range of proteins in a cell cannot be reduced by fractionation because of the limited sample amount. Since proteomics tends to identify the highly abundant proteins, it will be even more difficult to achieve significant depth. However, depth will need to be increased considerably to compete with scRNA-seq, which routinely measures transcripts from more than 5,000 genes. So far, mass spectrometry-based single-cell proteomics is still in its infancy, and the next years will show whether it can compete with scRNA-seq or whether other single-cell proteomics approaches, such as nanopore sequencing [383], will emerge as the leading technology. The development will be particularly interesting with regard to the analysis of PTMs at the single-cell level, thereby offering the opportunity to thruthfully reconstruct a complete picture of the effectors of cellular function and dysfunction in an unprecedented manner.

References

[1] W. Zakrzewski, M. Dobrzynski, M. Szymonowicz, and Z. Rybak, "Stem cells: past, present, and future," *Stem Cell Res Ther,* vol. 10, no. 1, p. 68, Feb 26 2019, doi: 10.1186/s13287-019-1165-5.

[2] F. Lu and Y. Zhang, "Cell totipotency: molecular features, induction, and maintenance," *Natl Sci Rev,* vol. 2, no. 2, pp. 217-225, Jun 2015, doi: 10.1093/nsr/nwv009.

[3] M. Boiani, E. Casser, G. Fuellen, and E. S. Christians, "Totipotency continuity from zygote to early blastomeres: a model under revision," (in eng), *Reproduction,* vol. 158, no. 2, pp. R49-r65, Aug 2019, doi: 10.1530/rep-18-0462.

[4] T. Vazin and W. J. Freed, "Human embryonic stem cells: derivation, culture, and differentiation: a review," *Restor Neurol Neurosci,* vol. 28, no. 4, pp. 589-603, 2010, doi: 10.3233/RNN-2010-0543.

[5] J. A. Thomson *et al.*, "Embryonic stem cell lines derived from human blastocysts," (in eng), *Science (New York, N.Y.),* vol. 282, no. 5391, pp. 1145-7, Nov 6 1998, doi: 10.1126/science.282.5391.1145.

[6] B. E. Reubinoff, M. F. Pera, C. Y. Fong, A. Trounson, and A. Bongso, "Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro," (in eng), *Nat Biotechnol,* vol. 18, no. 4, pp. 399-404, Apr 2000, doi: 10.1038/74447.

[7] K. Takahashi and S. Yamanaka, "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors," *Cell,* vol. 126, no. 4, pp. 663-76, Aug 25 2006, doi: 10.1016/j.cell.2006.07.024.

[8] K. Takahashi *et al.*, "Induction of pluripotent stem cells from adult human fibroblasts by defined factors," *Cell,* vol. 131, no. 5, pp. 861-72, Nov 30 2007, doi: 10.1016/j.cell.2007.11.019.

[9] Y. Kim, Y. A. Rim, H. Yi, N. Park, S. H. Park, and J. H. Ju, "The Generation of Human Induced Pluripotent Stem Cells from Blood Cells: An Efficient Protocol Using Serial Plating of Reprogrammed Cells by Centrifugation," *Stem Cells Int,* vol. 2016, p. 1329459, 2016, doi: 10.1155/2016/1329459.

[10] T. Aasen and J. C. Izpisua Belmonte, "Isolation and cultivation of human keratinocytes from skin or plucked hair for the generation of induced pluripotent stem cells," *Nature protocols,* vol. 5, no. 2, pp. 371-82, Feb 2010, doi: 10.1038/nprot.2009.241.

[11] J. Staerk *et al.*, "Reprogramming of human peripheral blood cells to induced pluripotent stem cells," *Cell stem cell,* vol. 7, no. 1, pp. 20-4, Jul 2 2010, doi: 10.1016/j.stem.2010.06.002.

[12] T. Zhou *et al.*, "Generation of human induced pluripotent stem cells from urine samples," *Nature protocols,* vol. 7, no. 12, pp. 2080-9, Dec 2012, doi: 10.1038/nprot.2012.115.

[13] Y. F. Jiang, M. Chen, N. N. Zhang, H. J. Yang, Q. Rui, and Y. F. Zhou, "In vitro and in vivo differentiation of induced pluripotent stem cells generated from urine-derived cells into cardiomyocytes," *Biol Open,* vol. 7, no. 1, Jan 8 2018, doi: 10.1242/bio.029157.

[14] L. Wang *et al.*, "Generation of integration-free neural progenitor cells from cells in human urine," *Nature methods,* vol. 10, no. 1, pp. 84-9, Jan 2013, doi: 10.1038/nmeth.2283.

[15] G. Liu, B. T. David, M. Trawczynski, and R. G. Fessler, "Advances in Pluripotent Stem Cells: History, Mechanisms, Technologies, and Applications," *Stem Cell Rev Rep,* vol. 16, no. 1, pp. 3-32, Feb 2020, doi: 10.1007/s12015-019-09935-x.

[16] N. Malik and M. S. Rao, "A review of the methods for human iPSC derivation," *Methods Mol Biol,* vol. 997, pp. 23-33, 2013, doi: 10.1007/978-1-62703-348-0_3.

[17] M. Stadtfeld, N. Maherali, D. T. Breault, and K. Hochedlinger, "Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse," *Cell stem cell,* vol. 2, no. 3, pp. 230-40, Mar 6 2008, doi: 10.1016/j.stem.2008.02.001.

[18] H. O. Li *et al.*, "A cytoplasmic RNA vector derived from nontransmissible Sendai virus with efficient gene transfer and expression," (in eng), *J Virol,* vol. 74, no. 14, pp. 6564-9, Jul 2000, doi: 10.1128/jvi.74.14.6564-6569.2000.

[19] N. Fusaki, H. Ban, A. Nishiyama, K. Saeki, and M. Hasegawa, "Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome," (in eng), *Proc Jpn Acad Ser B Phys Biol Sci,* vol. 85, no. 8, pp. 348-62, 2009, doi: 10.2183/pjab.85.348.

[20] K. Kaji, K. Norrby, A. Paca, M. Mileikovsky, P. Mohseni, and K. Woltjen, "Virus-free induction of pluripotency and subsequent excision of reprogramming factors," *Nature,* vol. 458, no. 7239, pp. 771-5, Apr 9 2009, doi: 10.1038/nature07864.

[21] P. Mali *et al.*, "Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes," *Stem cells (Dayton, Ohio),* vol. 28, no. 4, pp. 713-20, Apr 2010, doi: 10.1002/stem.402.

[22] J. Deinsberger, D. Reisinger, and B. Weber, "Global trends in clinical trials involving pluripotent stem cells: a systematic multi-database analysis," *NPJ Regen Med,* vol. 5, p. 15, 2020, doi: 10.1038/s41536-020-00100-4.

[23] Y. Qiao, O. S. Agboola, X. Hu, Y. Wu, and L. Lei, "Tumorigenic and Immunogenic Properties of Induced Pluripotent Stem Cells: a Promising Cancer Vaccine," *Stem Cell Rev Rep,* vol. 16, no. 6, pp. 1049-1061, Dec 2020, doi: 10.1007/s12015-020-10042-5.

[24] M. Yoshihara, A. Oguchi, and Y. Murakawa, "Genomic Instability of iPSCs and Challenges in Their Clinical Applications," (in eng), *Adv Exp Med Biol,* vol. 1201, pp. 23-47, 2019, doi: 10.1007/978-3-030-31206-0_2.

[25] Y. Jiang *et al.*, "Pluripotency of mesenchymal stem cells derived from adult marrow," (in eng), *Nature,* vol. 418, no. 6893, pp. 41-9, Jul 4 2002, doi: 10.1038/nature00870.

[26] D. Baksh, R. Yao, and R. S. Tuan, "Comparison of proliferative and multilineage differentiation potential of human mesenchymal stem cells derived from umbilical cord and bone marrow," *Stem cells (Dayton, Ohio),* vol. 25, no. 6, pp. 1384-92, Jun 2007, doi: 10.1634/stemcells.2006-0709.

[27] J. Saben *et al.*, "Distinct adipogenic differentiation phenotypes of human umbilical cord mesenchymal cells dependent on adipogenic conditions," *Exp Biol Med (Maywood),* vol. 239, no. 10, pp. 1340-51, Oct 2014, doi: 10.1177/1535370214539225.

[28] J. Y. Hsieh, Y. S. Fu, S. J. Chang, Y. H. Tsuang, and H. W. Wang, "Functional module analysis reveals differential osteogenic and stemness potentials in human mesenchymal stem cells from bone marrow and Wharton's jelly of umbilical cord," (in eng), *Stem Cells Dev,* vol. 19, no. 12, pp. 1895-910, Dec 2010, doi: 10.1089/scd.2009.0485.

[29] Y. Yang, H. Lin, H. Shen, B. Wang, G. Lei, and R. S. Tuan, "Mesenchymal stem cell-derived extracellular matrix enhances chondrogenic phenotype of and cartilage formation by encapsulated chondrocytes in vitro and in vivo," *Acta Biomater,* vol. 69, pp. 71-82, Mar 15 2018, doi: 10.1016/j.actbio.2017.12.043.

[30] S. M. Phadnis *et al.*, "Human bone marrow-derived mesenchymal cells differentiate and mature into endocrine pancreatic lineage in vivo," (in eng), *Cytotherapy,* vol. 13, no. 3, pp. 279-93, Mar 2011, doi: 10.3109/14653249.2010.523108.

[31] P. Stock, S. Bruckner, S. Winkler, M. M. Dollinger, and B. Christ, "Human bone marrow mesenchymal stem cell-derived hepatocytes improve the mouse liver after acute acetaminophen intoxication by preventing progress of injury," *Int J Mol Sci,* vol. 15, no. 4, pp. 7004-28, Apr 22 2014, doi: 10.3390/ijms15047004.

[32] A. Wilkins, K. Kemp, M. Ginty, K. Hares, E. Mallam, and N. Scolding, "Human bone marrow-derived mesenchymal stem cells secrete brain-derived neurotrophic factor which promotes neuronal survival in vitro," *Stem Cell Res,* vol. 3, no. 1, pp. 63-70, Jul 2009, doi: 10.1016/j.scr.2009.02.006.

[33] J. Dulak, K. Szade, A. Szade, W. Nowak, and A. Józkowicz, "Adult stem cells: hopes and hypes of regenerative medicine," (in eng), *Acta Biochim Pol,* vol. 62, no. 3, pp. 329-37, 2015, doi: 10.18388/abp.2015_1023.

[34] D. G. Phinney and D. J. Prockop, "Concise review: mesenchymal stem/multipotent stromal cells: the state of transdifferentiation and modes of tissue repair--current views," *Stem cells (Dayton, Ohio),* vol. 25, no. 11, pp. 2896-902, Nov 2007, doi: 10.1634/stemcells.2007-0637.

[35] Y. Han, X. Li, Y. Zhang, Y. Han, F. Chang, and J. Ding, "Mesenchymal Stem Cells for Regenerative Medicine," *Cells,* vol. 8, no. 8, Aug 13 2019, doi: 10.3390/cells8080886.

[36] K. Ko *et al.*, "Induction of pluripotency in adult unipotent germline stem cells," *Cell stem cell,* vol. 5, no. 1, pp. 87-96, Jul 2 2009, doi: 10.1016/j.stem.2009.05.025.

[37] K. Ko, M. J. Arauzo-Bravo, J. Kim, M. Stehling, and H. R. Scholer, "Conversion of adult mouse unipotent germline stem cells into pluripotent stem cells," *Nature protocols,* vol. 5, no. 5, pp. 921-8, May 2010, doi: 10.1038/nprot.2010.44.

[38] Z. Chen, Z. Li, and Z. He, "Plasticity of male germline stem cells and their applications in reproductive and regenerative medicine," *Asian J Androl,* vol. 17, no. 3, pp. 367-72, May-Jun 2015, doi: 10.4103/1008-682X.143739.

[39] M. Berdasco and M. Esteller, "DNA methylation in stem cell renewal and multipotency," (in eng), *Stem Cell Res Ther,* vol. 2, no. 5, p. 42, Oct 31 2011, doi: 10.1186/scrt83.

[40] D. E. Cohen and D. Melton, "Turning straw into gold: directing cell fate for regenerative medicine," *Nat Rev Genet,* vol. 12, no. 4, pp. 243-52, Apr 2011, doi: 10.1038/nrg2938.

[41] P. Cahan and G. Q. Daley, "Origins and implications of pluripotent stem cell variability and heterogeneity," *Nat Rev Mol Cell Biol,* vol. 14, no. 6, pp. 357-68, Jun 2013, doi: 10.1038/nrm3584.

[42] C. Jensen and Y. Teng, "Is It Time to Start Transitioning From 2D to 3D Cell Culture?," *Front Mol Biosci,* vol. 7, p. 33, 2020, doi: 10.3389/fmolb.2020.00033.

[43] D. Antoni, H. Burckel, E. Josset, and G. Noel, "Three-dimensional cell culture: a breakthrough in vivo," *Int J Mol Sci,* vol. 16, no. 3, pp. 5517-27, Mar 11 2015, doi: 10.3390/ijms16035517.

[44] M. A. Lancaster and J. A. Knoblich, "Organogenesis in a dish: modeling development and disease using organoid technologies," *Science (New York, N.Y.),* vol. 345, no. 6194, p. 1247125, Jul 18 2014, doi: 10.1126/science.1247125.

[45] M. A. Lancaster *et al.*, "Cerebral organoids model human brain development and microcephaly," *Nature,* vol. 501, no. 7467, pp. 373-9, Sep 19 2013, doi: 10.1038/nature12517.

[46] F. Antonica *et al.*, "Generation of functional thyroid from embryonic stem cells," *Nature,* vol. 491, no. 7422, pp. 66-71, Nov 1 2012, doi: 10.1038/nature11525.

[47] F. Sampaziotis *et al.*, "Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation," *Nat Biotechnol,* vol. 33, no. 8, pp. 845-852, Aug 2015, doi: 10.1038/nbt.3275.

[48] J. F. Dekkers *et al.*, "Characterizing responses to CFTR-modulating drugs using rectal organoids derived from subjects with cystic fibrosis," (in eng), *Sci Transl Med,* vol. 8, no. 344, p. 344ra84, Jun 22 2016, doi: 10.1126/scitranslmed.aad8278.

[49] Y. Du *et al.*, "Development of a miniaturized 3D organoid culture platform for ultra-high-throughput screening," *J Mol Cell Biol,* vol. 12, no. 8, pp. 630-643, Aug 1 2020, doi: 10.1093/jmcb/mjaa036.

[50] C. Mollinari, J. Zhao, L. Lupacchini, E. Garaci, D. Merlo, and G. Pei, "Transdifferentiation: a new promise for neurodegenerative diseases," *Cell Death Dis,* vol. 9, no. 8, p. 830, Aug 6 2018, doi: 10.1038/s41419-018-0891-4.

[51] R. L. Davis, H. Weintraub, and A. B. Lassar, "Expression of a single transfected cDNA converts fibroblasts to myoblasts," (in eng), *Cell,* vol. 51, no. 6, pp. 987-1000, Dec 24 1987, doi: 10.1016/0092-8674(87)90585-x.

[52] Q. Zhou, J. Brown, A. Kanarek, J. Rajagopal, and D. A. Melton, "In vivo reprogramming of adult pancreatic exocrine cells to beta-cells," *Nature,* vol. 455, no. 7213, pp. 627-32, Oct 2 2008, doi: 10.1038/nature07314.

[53] R. Zhang *et al.*, "In vivo cardiac reprogramming contributes to zebrafish heart regeneration," *Nature,* vol. 498, no. 7455, pp. 497-501, Jun 27 2013, doi: 10.1038/nature12322.

[54] Y.-C. Wang, S. E. Peterson, and J. F. Loring, "Protein post-translational modifications and regulation of pluripotency in human stem cells," *Cell Research,* vol. 24, no. 2, pp. 143-160, 2013, doi: 10.1038/cr.2013.151.

[55] S. M. Buckley *et al.*, "Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system," (in eng), *Cell stem cell,* vol. 11, no. 6, pp. 783-98, Dec 7 2012, doi: 10.1016/j.stem.2012.09.011.

[56] S. Tahmasebi *et al.*, "Sumoylation of Krüppel-like factor 4 inhibits pluripotency induction but promotes adipocyte differentiation," (in eng), *J Biol Chem,* vol. 288, no. 18, pp. 12791-804, May 3 2013, doi: 10.1074/jbc.M113.465443.

[57] Q. Yan *et al.*, "O-fucose modulates Notch-controlled blood lineage commitment," (in eng), *Am J Pathol,* vol. 176, no. 6, pp. 2921-34, Jun 2010, doi: 10.2353/ajpath.2010.090702.

[58] M. E. Torres-Padilla, D. E. Parfitt, T. Kouzarides, and M. Zernicka-Goetz, "Histone arginine methylation regulates pluripotency in the early mouse embryo," (in eng), *Nature,* vol. 445, no. 7124, pp. 214-8, Jan 11 2007, doi: 10.1038/nature05458.

[59] Y. L. Deribe, T. Pawson, and I. Dikic, "Post-translational modifications in signal integration," (in eng), *Nat Struct Mol Biol,* vol. 17, no. 6, pp. 666-72, Jun 2010, doi: 10.1038/nsmb.1842.

[60] A. Bononi *et al.*, "Protein kinases and phosphatases in the control of cell fate," (in eng), *Enzyme Res,* vol. 2011, p. 329098, 2011, doi: 10.4061/2011/329098.

[61] D. L. Swaney, C. D. Wenger, J. A. Thomson, and J. J. Coon, "Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry," (in eng), *Proceedings of the National Academy of Sciences of the United States of America,* vol. 106, no. 4, pp. 995-1000, Jan 27 2009, doi: 10.1073/pnas.0811964106.

[62] D. Van Hoof *et al.*, "Phosphorylation dynamics during early differentiation of human embryonic stem cells," *Cell stem cell,* vol. 5, no. 2, pp. 214-26, Aug 7 2009, doi: 10.1016/j.stem.2009.05.021.

[63] P. Yang *et al.*, "Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency," *Cell Syst,* vol. 8, no. 5, pp. 427-445 e10, May 22 2019, doi: 10.1016/j.cels.2019.03.012.

[64] T. Narita, B. T. Weinert, and C. Choudhary, "Functions and mechanisms of non-histone protein acetylation," *Nat Rev Mol Cell Biol,* vol. 20, no. 3, pp. 156-174, Mar 2019, doi: 10.1038/s41580-018-0081-3.

[65] V. G. Allfrey, R. Faulkner, and A. E. Mirsky, "ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS," (in eng), *Proceedings of the National Academy of Sciences of the United States of America,* vol. 51, no. 5, pp. 786-94, May 1964, doi: 10.1073/pnas.51.5.786.

[66] S. M. Görisch, M. Wachsmuth, K. F. Tóth, P. Lichter, and K. Rippe, "Histone acetylation increases chromatin accessibility," (in eng), *J Cell Sci,* vol. 118, no. Pt 24, pp. 5825-34, Dec 15 2005, doi: 10.1242/jcs.02689.

[67] S. Efroni *et al.*, "Global transcription in pluripotent embryonic stem cells," (in eng), *Cell stem cell,* vol. 2, no. 5, pp. 437-47, May 8 2008, doi: 10.1016/j.stem.2008.03.021.

[68] D. Huangfu *et al.*, "Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds," (in eng), *Nat Biotechnol,* vol. 26, no. 7, pp. 795-7, Jul 2008, doi: 10.1038/nbt1418.

[69] S. J. Kang, Y. I. Park, B. So, and H. G. Kang, "Sodium butyrate efficiently converts fully reprogrammed induced pluripotent stem cells from mouse partially reprogrammed cells," (in eng), *Cell Reprogram,* vol. 16, no. 5, pp. 345-54, Oct 2014, doi: 10.1089/cell.2013.0087.

[70] M. Borowiak *et al.*, "Small Molecules Efficiently Direct Endodermal Differentiation of Mouse and Human Embryonic Stem Cells," *Cell stem cell,* vol. 4, no. 4, pp. 348-358, 2009, doi: 10.1016/j.stem.2009.01.014.

[71] J. de Boer, R. Licht, M. Bongers, T. van der Klundert, R. Arends, and C. van Blitterswijk, "Inhibition of histone acetylation as a tool in bone tissue engineering," (in eng), *Tissue Eng,* vol. 12, no. 10, pp. 2927-37, Oct 2006, doi: 10.1089/ten.2006.12.2927.

[72] X. Yao, J. R. Zhang, H. R. Huang, L. C. Dai, Q. J. Liu, and M. Zhang, "Histone deacetylase inhibitor promotes differentiation of embryonic stem cells into neural cells in adherent monoculture," (in eng), *Chin Med J (Engl),* vol. 123, no. 6, pp. 734-8, Mar 20 2010.

[73] J. Hsieh, K. Nakashima, T. Kuwabara, E. Mejia, and F. H. Gage, "Histone deacetylase inhibition-mediated neuronal differentiation of multipotent adult neural progenitor cells," (in eng), *Proceedings of the National Academy of Sciences of the United States of America,* vol. 101, no. 47, pp. 16659-64, Nov 23 2004, doi: 10.1073/pnas.0407643101.

[74] C. Elizalde *et al.*, "Histone deacetylase 3 modulates the expansion of human hematopoietic stem cells," (in eng), *Stem Cells Dev,* vol. 21, no. 14, pp. 2581-91, Sep 20 2012, doi: 10.1089/scd.2011.0698.

[75] C. Haumaitre, O. Lenoir, and R. Scharfmann, "Histone deacetylase inhibitors modify pancreatic cell fate determination and amplify endocrine progenitors," (in eng), *Mol Cell Biol,* vol. 28, no. 20, pp. 6373-83, Oct 2008, doi: 10.1128/mcb.00413-08.

[76] L. Huang *et al.*, "Commitment and oncogene-induced plasticity of human stem cell-derived pancreatic acinar and ductal organoids," (in eng), *Cell stem cell,* vol. 28, no. 6, pp. 1090-1104.e6, Jun 3 2021, doi: 10.1016/j.stem.2021.03.022.

[77] H. S. Han, G. Kang, J. S. Kim, B. H. Choi, and S. H. Koo, "Regulation of glucose metabolism from a liver-centric perspective," *Exp Mol Med,* vol. 48, p. e218, Mar 11 2016, doi: 10.1038/emm.2015.122.

[78] R. C. Nordlie, J. D. Foster, and A. J. Lange, "Regulation of glucose production by the liver," (in eng), *Annu Rev Nutr,* vol. 19, pp. 379-406, 1999, doi: 10.1146/annurev.nutr.19.1.379.

[79] G. Schreiber, "The synthesis and secretion of plasma proteins in the liver," *Pathology,* vol. 10, no. 4, p. 394, 1978/01/01/ 1978, doi: https://doi.org/10.1016/S0031-3025(16)39817-8.

[80] P. B. Watkins, "Drug safety sciences and the bottleneck in drug development," (in eng), *Clin Pharmacol Ther,* vol. 89, no. 6, pp. 788-90, Jun 2011, doi: 10.1038/clpt.2011.63.

[81] I. J. Onakpoya, C. J. Heneghan, and J. K. Aronson, "Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature," (in eng), *BMC Med,* vol. 14, p. 10, Feb 4 2016, doi: 10.1186/s12916-016-0553-2.

[82] W. Seo and W. I. Jeong, "Hepatic non-parenchymal cells: Master regulators of alcoholic liver disease?," *World J Gastroenterol,* vol. 22, no. 4, pp. 1348-56, Jan 28 2016, doi: 10.3748/wjg.v22.i4.1348.

[83] S. Lemoinne, A. Cadoret, H. El Mourabit, D. Thabut, and C. Housset, "Origins and functions of liver myofibroblasts," *Biochim Biophys Acta,* vol. 1832, no. 7, pp. 948-54, Jul 2013, doi: 10.1016/j.bbadis.2013.02.019.

[84] T. Higashi, S. L. Friedman, and Y. Hoshida, "Hepatic stellate cells as key target in liver fibrosis," *Adv Drug Deliv Rev,* vol. 121, pp. 27-42, Nov 1 2017, doi: 10.1016/j.addr.2017.05.007.

[85] L. D. Deleve, X. Wang, and Y. Guo, "Sinusoidal endothelial cells prevent rat stellate cell activation and promote reversion to quiescence," (in eng), *Hepatology (Baltimore, Md.),* vol. 48, no. 3, pp. 920-30, Sep 2008, doi: 10.1002/hep.22351.

[86] L. D. DeLeve and A. C. Maretti-Mira, "Liver Sinusoidal Endothelial Cell: An Update," *Semin Liver Dis,* vol. 37, no. 4, pp. 377-387, Nov 2017, doi: 10.1055/s-0037-1617455.

[87] R. Malik, C. Selden, and H. Hodgson, "The role of non-parenchymal cells in liver growth," (in eng), *Semin Cell Dev Biol,* vol. 13, no. 6, pp. 425-31, Dec 2002, doi: 10.1016/s1084952102001301.

[88] C. E. Murry and G. Keller, "Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development," (in eng), *Cell,* vol. 132, no. 4, pp. 661-80, Feb 22 2008, doi: 10.1016/j.cell.2008.02.008.

[89] A. M. Zorn, "Liver development," in *StemBook*. Cambridge (MA): Harvard Stem Cell Institute

[90] S. Yagi, M. Hirata, Y. Miyachi, and S. Uemoto, "Liver Regeneration after Hepatectomy and Partial Liver Transplantation," (in eng), *Int J Mol Sci,* vol. 21, no. 21, Nov 9 2020, doi: 10.3390/ijms21218414.

[91] K. Abe *et al.*, "Endoderm-specific gene expression in embryonic stem cells differentiated to embryoid bodies," (in eng), *Exp Cell Res,* vol. 229, no. 1, pp. 27-34, Nov 25 1996, doi: 10.1006/excr.1996.0340.

[92] A. A. Palakkan, J. Nanda, and J. A. Ross, "Pluripotent stem cells to hepatocytes, the journey so far," *Biomed Rep,* vol. 6, no. 4, pp. 367-373, Apr 2017, doi: 10.3892/br.2017.867.

[93] K. D. Tremblay, P. A. Hoodless, E. K. Bikoff, and E. J. Robertson, "Formation of the definitive endoderm in mouse is a Smad2-dependent process," (in eng), *Development (Cambridge, England),* vol. 127, no. 14, pp. 3079-90, Jul 2000.

[94] S. D. Vincent, N. R. Dunn, S. Hayashi, D. P. Norris, and E. J. Robertson, "Cell fate decisions within the mouse organizer are governed by graded Nodal signals," *Genes & development,* vol. 17, no. 13, pp. 1646-62, Jul 1 2003, doi: 10.1101/gad.1100503.

[95] K. A. D'Amour, A. D. Agulnick, S. Eliazer, O. G. Kelly, E. Kroon, and E. E. Baetge, "Efficient differentiation of human embryonic stem cells to definitive endoderm," *Nat Biotechnol,* vol. 23, no. 12, pp. 1534-41, Dec 2005, doi: 10.1038/nbt1163.

[96] M. Nakanishi *et al.*, "Directed induction of anterior and posterior primitive streak by Wnt from embryonic stem cells cultured in a chemically defined serum-free medium," *FASEB J,* vol. 23, no. 1, pp. 114-22, Jan 2009, doi: 10.1096/fj.08-111203.

[97] S. Engert, I. Burtscher, W. P. Liao, S. Dulev, G. Schotta, and H. Lickert, "Wnt/beta-catenin signalling regulates Sox17 expression and is essential for organizer and endoderm formation in the mouse," *Development (Cambridge, England),* vol. 140, no. 15, pp. 3128-38, Aug 2013, doi: 10.1242/dev.088765.

[98] D. C. Hay *et al.*, "Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling," (in eng), *Proceedings of the National Academy of Sciences of the United States of America,* vol. 105, no. 34, pp. 12301-6, Aug 26 2008, doi: 10.1073/pnas.0806522105.

[99] P. Roelandt *et al.*, "Human embryonic and rat adult stem cells with primitive endoderm-like phenotype can be fated to definitive endoderm, and finally hepatocyte-like cells," *PLoS One,* vol. 5, no. 8, p. e12101, Aug 11 2010, doi: 10.1371/journal.pone.0012101.

[100] J. G. Camp *et al.*, "Multilineage communication regulates human liver bud development from pluripotency," (in eng), *Nature,* vol. 546, no. 7659, pp. 533-538, Jun 22 2017, doi: 10.1038/nature22796.

[101] A. G. Lade and S. P. Monga, "Beta-catenin signaling in hepatic development and progenitors: which way does the WNT blow?," *Dev Dyn,* vol. 240, no. 3, pp. 486-500, Mar 2011, doi: 10.1002/dvdy.22522.

[102] A. M. Zorn and J. M. Wells, "Vertebrate endoderm development and organ formation," *Annual review of cell and developmental biology,* vol. 25, pp. 221-51, 2009, doi: 10.1146/annurev.cellbio.042308.113344.

[103] S. Toivonen *et al.*, "Activin A and Wnt-dependent specification of human definitive endoderm cells," *Exp Cell Res,* vol. 319, no. 17, pp. 2535-44, Oct 15 2013, doi: 10.1016/j.yexcr.2013.07.007.

[104] R. Siller, S. Greenhough, E. Naumovska, and G. J. Sullivan, "Small-molecule-driven hepatocyte differentiation of human pluripotent stem cells," *Stem cell reports,* vol. 4, no. 5, pp. 939-52, May 12 2015, doi: 10.1016/j.stemcr.2015.04.001.

[105] J. Jung, M. Zheng, M. Goldfarb, and K. S. Zaret, "Initiation of mammalian liver development from endoderm by fibroblast growth factors," (in eng), *Science (New York, N.Y.),* vol. 284, no. 5422, pp. 1998-2003, Jun 18 1999, doi: 10.1126/science.284.5422.1998.

[106] J. M. Rossi, N. R. Dunn, B. L. Hogan, and K. S. Zaret, "Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm," *Genes & development,* vol. 15, no. 15, pp. 1998-2009, Aug 1 2001, doi: 10.1101/gad.904601.

[107] K. Si-Tayeb *et al.*, "Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells," (in eng), *Hepatology (Baltimore, Md.),* vol. 51, no. 1, pp. 297-305, Jan 2010, doi: 10.1002/hep.23354.

[108] L. K. Kanninen *et al.*, "Laminin-511 and laminin-521-based matrices for efficient hepatic specification of human pluripotent stem cells," *Biomaterials,* vol. 103, pp. 86-100, Oct 2016, doi: 10.1016/j.biomaterials.2016.06.054.

[109] C. Du *et al.*, "Highly efficient and expedited hepatic differentiation from human pluripotent stem cells by pure small-molecule cocktails," *Stem Cell Res Ther,* vol. 9, no. 1, p. 58, Mar 9 2018, doi: 10.1186/s13287-018-0794-4.

[110] F. Sampaziotis, M. C. de Brito, I. Geti, A. Bertero, N. R. Hannan, and L. Vallier, "Directed differentiation of human induced pluripotent stem cells into functional cholangiocyte-like cells," *Nature protocols,* vol. 12, no. 4, pp. 814-827, Apr 2017, doi: 10.1038/nprot.2017.011.

[111] S. Tamura, Y. Morikawa, A. Miyajima, and E. Senba, "Expression of oncostatin M in hematopoietic organs," (in eng), *Dev Dyn,* vol. 225, no. 3, pp. 327-31, Nov 2002, doi: 10.1002/dvdy.10156.

[112] A. Kamiya, T. Kinoshita, and A. Miyajima, "Oncostatin M and hepatocyte growth factor induce hepatic maturation via distinct signaling pathways," (in eng), *FEBS Lett,* vol. 492, no. 1-2, pp. 90-4, Mar 9 2001, doi: 10.1016/s0014-5793(01)02140-8.

[113] I. S. Behbahan *et al.*, "New approaches in the differentiation of human embryonic stem cells and induced pluripotent stem cells toward hepatocytes," *Stem Cell Rev Rep,* vol. 7, no. 3, pp. 748-59, Sep 2011, doi: 10.1007/s12015-010-9216-4.

[114] T. Kinoshita *et al.*, "Hepatic differentiation induced by oncostatin M attenuates fetal liver hematopoiesis," (in eng), *Proceedings of the National Academy of Sciences of the United States of America,* vol. 96, no. 13, pp. 7265-70, Jun 22 1999, doi: 10.1073/pnas.96.13.7265.

[115] C. C. Benoist *et al.*, "The procognitive and synaptogenic effects of angiotensin IV-derived peptides are dependent on activation of the hepatocyte growth factor/c-met system," *J*

Pharmacol Exp Ther, vol. 351, no. 2, pp. 390-402, Nov 2014, doi: 10.1124/jpet.114.218735.

[116] G. K. Michalopoulos, W. C. Bowen, K. Mulè, and J. Luo, "HGF-, EGF-, and dexamethasone-induced gene expression patterns during formation of tissue in hepatic organoid cultures," (in eng), *Gene Expr,* vol. 11, no. 2, pp. 55-75, 2003, doi: 10.3727/000000003108748964.

[117] S. S. Chen, W. Fitzgerald, J. Zimmerberg, H. K. Kleinman, and L. Margolis, "Cell-cell and cell-extracellular matrix interactions regulate embryonic stem cell differentiation," *Stem cells (Dayton, Ohio),* vol. 25, no. 3, pp. 553-61, Mar 2007, doi: 10.1634/stemcells.2006-0419.

[118] H. Wang, X. Luo, and J. Leighton, "Extracellular Matrix and Integrins in Embryonic Stem Cell Differentiation," *Biochem Insights,* vol. 8, no. Suppl 2, pp. 15-21, 2015, doi: 10.4137/BCI.S30377.

[119] E. Luce, A. Messina, J. C. Duclos-Vallee, and A. Dubart-Kupperschmitt, "Advanced Techniques and Awaited Clinical Applications for Human Pluripotent Stem Cell Differentiation into Hepatocytes," *Hepatology (Baltimore, Md.),* vol. 74, no. 2, pp. 1101-1116, Aug 2021, doi: 10.1002/hep.31705.

[120] J. C. Wong, S. Y. Gao, J. G. Lees, M. B. Best, R. Wang, and B. E. Tuch, "Definitive endoderm derived from human embryonic stem cells highly express the integrin receptors alphaV and beta5," *Cell Adh Migr,* vol. 4, no. 1, pp. 39-45, Jan-Mar 2010, doi: 10.4161/cam.4.1.10627.

[121] B. R. Lawton, J. A. Sosa, S. Roman, and D. S. Krause, "Effect of a matrigel sandwich on endodermal differentiation of human embryonic stem cells," (in eng), *Stem Cell Rev Rep,* vol. 9, no. 5, pp. 578-85, Oct 2013, doi: 10.1007/s12015-013-9447-2.

[122] T. Takebe *et al.*, "Vascularized and functional human liver from an iPSC-derived organ bud transplant," (in eng), *Nature,* vol. 499, no. 7459, pp. 481-4, Jul 25 2013, doi: 10.1038/nature12271.

[123] T. Takebe *et al.*, "Massive and Reproducible Production of Liver Buds Entirely from Human Pluripotent Stem Cells," *Cell reports,* vol. 21, no. 10, pp. 2661-2670, Dec 5 2017, doi: 10.1016/j.celrep.2017.11.005.

[124] W. Zhou, K. Graham, B. Lucendo-Villarin, O. Flint, D. C. Hay, and P. Bagnaninchi, "Combining stem cell-derived hepatocytes with impedance sensing to better predict human drug toxicity," (in eng), *Expert Opin Drug Metab Toxicol,* vol. 15, no. 1, pp. 77-83, Jan 2019, doi: 10.1080/17425255.2019.1558208.

[125] G. Holmgren *et al.*, "Long-term chronic toxicity testing using human pluripotent stem cell-derived hepatocytes," (in eng), *Drug Metab Dispos,* vol. 42, no. 9, pp. 1401-6, Sep 2014, doi: 10.1124/dmd.114.059154.

[126] J. H. Kim *et al.*, "Enhanced Metabolizing Activity of Human ES Cell-Derived Hepatocytes Using a 3D Culture System with Repeated Exposures to Xenobiotics," *Toxicol Sci,* vol. 147, no. 1, pp. 190-206, Sep 2015, doi: 10.1093/toxsci/kfv121.

[127] R. L. Gieseck, 3rd *et al.*, "Maturation of induced pluripotent stem cell derived hepatocytes by 3D-culture," *PLoS One,* vol. 9, no. 1, p. e86372, 2014, doi: 10.1371/journal.pone.0086372.

[128] J. L. Corbett and S. A. Duncan, "iPSC-Derived Hepatocytes as a Platform for Disease Modeling and Drug Discovery," *Front Med (Lausanne),* vol. 6, p. 265, 2019, doi: 10.3389/fmed.2019.00265.

[129] R. Jing *et al.*, "A Screen Using iPSC-Derived Hepatocytes Reveals NAD(+) as a Potential Treatment for mtDNA Depletion Syndrome," *Cell reports,* vol. 25, no. 6, pp. 1469-1484 e5, Nov 6 2018, doi: 10.1016/j.celrep.2018.10.036.

[130] S. M. Choi *et al.*, "Efficient drug screening and gene correction for treating liver disease using patient-specific stem cells," *Hepatology (Baltimore, Md.),* vol. 57, no. 6, pp. 2458-68, Jun 2013, doi: 10.1002/hep.26237.

[131] A. Miyajima, M. Tanaka, and T. Itoh, "Stem/progenitor cells in liver development, homeostasis, regeneration, and reprogramming," *Cell stem cell,* vol. 14, no. 5, pp. 561-74, May 1 2014, doi: 10.1016/j.stem.2014.04.010.

[132] H. Abdellatif, "Oval Cells: Potential Role in Liver Regeneration," *Biomedical Journal of Scientific & Technical Research,* vol. 2, no. 1, 2017, doi: 10.26717/bjstr.2018.02.000665.

[133] S. J. Awan, M. T. Baig, F. Yaqub, A. Tayyeb, and G. Ali, "In vitro differentiated hepatic oval-like cells enhance hepatic regeneration in CCl(4) -induced hepatic injury," (in eng), *Cell Biol Int,* vol. 41, no. 1, pp. 51-61, Jan 2017, doi: 10.1002/cbin.10699.

[134] D. Campard, P. A. Lysy, M. Najimi, and E. M. Sokal, "Native umbilical cord matrix stem cells express hepatic markers and differentiate into hepatocyte-like cells," (in eng), *Gastroenterology,* vol. 134, no. 3, pp. 833-48, Mar 2008, doi: 10.1053/j.gastro.2007.12.024.

[135] S. Z. Lin *et al.*, "Transplantation of human Wharton's Jelly-derived stem cells alleviates chemically induced liver fibrosis in rats," *Cell Transplant,* vol. 19, no. 11, pp. 1451-63, 2010, doi: 10.3727/096368910X514198.

[136] C. N. Shen, J. M. Slack, and D. Tosh, "Molecular basis of transdifferentiation of pancreas to liver," (in eng), *Nature cell biology,* vol. 2, no. 12, pp. 879-87, Dec 2000, doi: 10.1038/35046522.

[137] G. Song *et al.*, "Direct Reprogramming of Hepatic Myofibroblasts into Hepatocytes In Vivo Attenuates Liver Fibrosis," *Cell stem cell,* vol. 18, no. 6, pp. 797-808, Jun 2 2016, doi: 10.1016/j.stem.2016.01.010.

[138] P. Huang *et al.*, "Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors," *Nature,* vol. 475, no. 7356, pp. 386-9, May 11 2011, doi: 10.1038/nature10116.

[139] S. Sekiya and A. Suzuki, "Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors," *Nature,* vol. 475, no. 7356, pp. 390-3, Jun 29 2011, doi: 10.1038/nature10263.

[140] Y. Liu, A. Beyer, and R. Aebersold, "On the Dependency of Cellular Protein Levels on mRNA Abundance," *Cell,* vol. 165, no. 3, pp. 535-50, Apr 21 2016, doi: 10.1016/j.cell.2016.03.014.

[141] A. D. Catherman, O. S. Skinner, and N. L. Kelleher, "Top Down proteomics: facts and perspectives," *Biochem Biophys Res Commun,* vol. 445, no. 4, pp. 683-93, Mar 21 2014, doi: 10.1016/j.bbrc.2014.02.041.

[142] M. S. Kim *et al.*, "A draft map of the human proteome," *Nature,* vol. 509, no. 7502, pp. 575-81, May 29 2014, doi: 10.1038/nature13302.

[143] M. Wilhelm *et al.*, "Mass-spectrometry-based draft of the human proteome," (in eng), *Nature,* vol. 509, no. 7502, pp. 582-7, May 29 2014, doi: 10.1038/nature13319.

[144] V. Mayya and D. K. Han, "Phosphoproteomics by mass spectrometry: insights, implications, applications and limitations," *Expert Rev Proteomics,* vol. 6, no. 6, pp. 605-18, Dec 2009, doi: 10.1586/epr.09.84.

[145] M. Shehadul Islam, A. Aryasomayajula, and P. Selvaganapathy, "A Review on Macroscale and Microscale Cell Lysis Methods," *Micromachines,* vol. 8, no. 3, 2017, doi: 10.3390/mi8030083.

[146] A. K. Bhuyan, "On the mechanism of SDS-induced protein denaturation," *Biopolymers,* vol. 93, no. 2, pp. 186-99, Feb 2010, doi: 10.1002/bip.21318.

[147] K. L. Rundlett and D. W. Armstrong, "Mechanism of signal suppression by anionic surfactants in capillary electrophoresis-electrospray ionization mass spectrometry," (in eng), *Analytical chemistry,* vol. 68, no. 19, pp. 3493-7, Oct 1 1996, doi: 10.1021/ac960472p.

[148] R. R. Loo, N. Dales, and P. C. Andrews, "Surfactant effects on protein structure examined by electrospray ionization mass spectrometry," (in eng), *Protein Sci,* vol. 3, no. 11, pp. 1975-83, Nov 1994, doi: 10.1002/pro.5560031109.

[149] A. Das and C. Mukhopadhyay, "Urea-mediated protein denaturation: a consensus view," (in eng), *J Phys Chem B,* vol. 113, no. 38, pp. 12816-24, Sep 24 2009, doi: 10.1021/jp906350s.

[150] L. Switzar, M. Giera, and W. M. Niessen, "Protein digestion: an overview of the available techniques and recent developments," *Journal of proteome research,* vol. 12, no. 3, pp. 1067-77, Mar 1 2013, doi: 10.1021/pr301201x.

[151] J. M. Burkhart, C. Schumbrutzki, S. Wortelkamp, A. Sickmann, and R. P. Zahedi, "Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics," (in eng), *J Proteomics,* vol. 75, no. 4, pp. 1454-62, Feb 2 2012, doi: 10.1016/j.jprot.2011.11.016.

[152] P. Giansanti, T. T. Aye, H. van den Toorn, M. Peng, B. van Breukelen, and A. J. Heck, "An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas," *Cell reports,* vol. 11, no. 11, pp. 1834-43, Jun 23 2015, doi: 10.1016/j.celrep.2015.05.029.

[153] D. Wang *et al.*, "A deep proteome and transcriptome abundance atlas of 29 healthy human tissues," *Molecular systems biology,* vol. 15, no. 2, p. e8503, Feb 18 2019, doi: 10.15252/msb.20188503.

[154] A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, and M. Mann, "In-gel digestion for mass spectrometric characterization of proteins and proteomes," *Nature protocols,* vol. 1, no. 6, pp. 2856-60, 2006, doi: 10.1038/nprot.2006.468.

[155] J. Havlis and A. Shevchenko, "Absolute quantification of proteins in solutions and in polyacrylamide gels by mass spectrometry," (in eng), *Analytical chemistry,* vol. 76, no. 11, pp. 3029-36, Jun 1 2004, doi: 10.1021/ac035286f.

[156] M. HaileMariam *et al.*, "S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics," *Journal of proteome research,* vol. 17, no. 9, pp. 2917-2924, Sep 7 2018, doi: 10.1021/acs.jproteome.8b00505.

[157] J. R. Wisniewski, A. Zougman, N. Nagaraj, and M. Mann, "Universal sample preparation method for proteome analysis," *Nature methods,* vol. 6, no. 5, pp. 359-62, May 2009, doi: 10.1038/nmeth.1322.

[158] N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, and M. Mann, "Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells," (in eng), *Nature methods,* vol. 11, no. 3, pp. 319-24, Mar 2014, doi: 10.1038/nmeth.2834.

[159] M. Sielaff *et al.*, "Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range," (in eng), *Journal of proteome research,* Oct 11 2017, doi: 10.1021/acs.jproteome.7b00433.

[160] C. S. Hughes, S. Foehr, D. A. Garfield, E. E. Furlong, L. M. Steinmetz, and J. Krijgsveld, "Ultrasensitive proteome analysis using paramagnetic bead technology," (in eng), *Molecular systems biology,* vol. 10, p. 757, Oct 30 2014, doi: 10.15252/msb.20145625.

[161] T. S. Batth *et al.*, "Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation," *Molecular & cellular proteomics : MCP,* vol. 18, no. 5, pp. 1027-1035, May 2019, doi: 10.1074/mcp.TIR118.001270.

[162] T. Muller, M. Kalxdorf, R. Longuespee, D. N. Kazdal, A. Stenzinger, and J. Krijgsveld, "Automated sample preparation with SP3 for low-input clinical proteomics," *Molecular systems biology,* vol. 16, no. 1, p. e9111, Jan 2020, doi: 10.15252/msb.20199111.

[163] B. Ruprecht, D. Wang, R. Z. Chiozzi, L. H. Li, H. Hahne, and B. Kuster, "Hydrophilic Strong Anion Exchange (hSAX) Chromatography Enables Deep Fractionation of Tissue Proteomes," (in eng), *Methods Mol Biol,* vol. 1550, pp. 69-82, 2017, doi: 10.1007/978-1-4939-6747-6_7.

[164] J. Rappsilber, M. Mann, and Y. Ishihama, "Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips," (in eng), *Nature protocols,* vol. 2, no. 8, pp. 1896-906, 2007, doi: 10.1038/nprot.2007.261.

[165] B. Manadas, V. M. Mendes, J. English, and M. J. Dunn, "Peptide fractionation in proteomics approaches," *Expert Rev Proteomics,* vol. 7, no. 5, pp. 655-63, Oct 2010, doi: 10.1586/epr.10.46.

[166] M. Gilar, P. Olivova, A. E. Daly, and J. C. Gebler, "Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions," (in eng), *J Sep Sci,* vol. 28, no. 14, pp. 1694-703, Sep 2005, doi: 10.1002/jssc.200500116.

[167] M. S. Ritorto, K. Cook, K. Tyagi, P. G. Pedrioli, and M. Trost, "Hydrophilic strong anion exchange (hSAX) chromatography for highly orthogonal peptide separation of complex proteomes," *Journal of proteome research,* vol. 12, no. 6, pp. 2449-57, Jun 7 2013, doi: 10.1021/pr301011r.

[168] S. Di Palma, P. J. Boersema, A. J. Heck, and S. Mohammed, "Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC and ZIC-cHILIC) provide high resolution separation and increase sensitivity in proteome analysis," *Analytical chemistry,* vol. 83, no. 9, pp. 3440-7, May 1 2011, doi: 10.1021/ac103312e.

[169] Y. Ishihama, J. Rappsilber, and M. Mann, "Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics," (in eng), *Journal of proteome research,* vol. 5, no. 4, pp. 988-94, Apr 2006, doi: 10.1021/pr050385q.

[170] P. Yu *et al.*, "Trimodal Mixed Mode Chromatography That Enables Efficient Offline Two-Dimensional Peptide Fractionation for Proteome Analysis," (in eng), *Analytical chemistry,* vol. 89, no. 17, pp. 8884-8891, Sep 5 2017, doi: 10.1021/acs.analchem.7b01356.

[171] B. Macek, M. Mann, and J. V. Olsen, "Global and site-specific quantitative phosphoproteomics: principles and applications," *Annu Rev Pharmacol Toxicol,* vol. 49, pp. 199-221, 2009, doi: 10.1146/annurev.pharmtox.011008.145606.

[172] L. Andersson and J. Porath, "Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography," (in eng), *Anal Biochem,* vol. 154, no. 1, pp. 250-4, Apr 1986, doi: 10.1016/0003-2697(86)90523-3.

[173] B. Ruprecht, H. Koch, P. Domasinska, M. Frejno, B. Kuster, and S. Lemeer, "Optimized Enrichment of Phosphoproteomes by Fe-IMAC Column Chromatography," (in eng), *Methods Mol Biol,* vol. 1550, pp. 47-60, 2017, doi: 10.1007/978-1-4939-6747-6_5.

[174] P. Mertins *et al.*, "Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry," *Nature protocols,* vol. 13, no. 7, pp. 1632-1661, Jul 2018, doi: 10.1038/s41596-018-0006-9.

[175] B. Ruprecht, H. Koch, G. Medard, M. Mundt, B. Kuster, and S. Lemeer, "Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns," *Molecular & cellular proteomics : MCP,* vol. 14, no. 1, pp. 205-15, Jan 2015, doi: 10.1074/mcp.M114.043109.

[176] H. Post *et al.*, "Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons," *Journal of proteome research,* vol. 16, no. 2, pp. 728-737, Feb 3 2017, doi: 10.1021/acs.jproteome.6b00753.

[177] J. R. Murillo, M. Kuras, M. Rezeli, T. Miliotis, L. Betancourt, and G. Marko-Varga, "Automated phosphopeptide enrichment from minute quantities of frozen malignant melanoma tissue," *PLoS One,* vol. 13, no. 12, p. e0208562, 2018, doi: 10.1371/journal.pone.0208562.

[178] I. Arribas Diez, I. Govender, P. Naicker, S. Stoychev, J. Jordaan, and O. N. Jensen, "Zirconium(IV)-IMAC Revisited: Improved Performance and Phosphoproteome Coverage

by Magnetic Microparticles for Phosphopeptide Affinity Enrichment," *Journal of proteome research,* vol. 20, no. 1, pp. 453-462, Jan 1 2021, doi: 10.1021/acs.jproteome.0c00508.

[179] L. Liu *et al.*, "p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage," (in eng), *Mol Cell Biol,* vol. 19, no. 2, pp. 1202-9, Feb 1999, doi: 10.1128/mcb.19.2.1202.

[180] J. Boyes, P. Byfield, Y. Nakatani, and V. Ogryzko, "Regulation of activity of the transcription factor GATA-1 by acetylation," (in eng), *Nature,* vol. 396, no. 6711, pp. 594-8, Dec 10 1998, doi: 10.1038/25166.

[181] M. Esmaeili, S. A. Blythe, J. W. Tobias, K. Zhang, J. Yang, and P. S. Klein, "Chromatin accessibility and histone acetylation in the regulation of competence in early development," (in eng), *Dev Biol,* vol. 462, no. 1, pp. 20-35, Jun 1 2020, doi: 10.1016/j.ydbio.2020.02.013.

[182] N. Z. Saraiva, C. S. Oliveira, and J. M. Garcia, "Histone acetylation and its role in embryonic stem cell differentiation," *World J Stem Cells,* vol. 2, no. 6, pp. 121-6, Dec 26 2010, doi: 10.4252/wjsc.v2.i6.121.

[183] S. Legartová *et al.*, "Cell differentiation along multiple pathways accompanied by changes in histone acetylation status," (in eng), *Biochem Cell Biol,* vol. 92, no. 2, pp. 85-93, Apr 2014, doi: 10.1139/bcb-2013-0082.

[184] T. Kawamura *et al.*, "Acetylation of GATA-4 is involved in the differentiation of embryonic stem cells into cardiac myocytes," *J Biol Chem,* vol. 280, no. 20, pp. 19682-8, May 20 2005, doi: 10.1074/jbc.M412428200.

[185] C. Choudhary *et al.*, "Lysine acetylation targets protein complexes and co-regulates major cellular functions," (in eng), *Science (New York, N.Y.),* vol. 325, no. 5942, pp. 834-40, Aug 14 2009, doi: 10.1126/science.1175371.

[186] S. C. Kim *et al.*, "Substrate and functional diversity of lysine acetylation revealed by a proteomics survey," *Mol Cell,* vol. 23, no. 4, pp. 607-18, Aug 2006, doi: 10.1016/j.molcel.2006.06.026.

[187] T. Svinkina *et al.*, "Deep, Quantitative Coverage of the Lysine Acetylome Using Novel Anti-acetyl-lysine Antibodies and an Optimized Proteomic Workflow," *Molecular & cellular proteomics : MCP,* vol. 14, no. 9, pp. 2429-40, Sep 2015, doi: 10.1074/mcp.O114.047555.

[188] H. Steen and M. Mann, "The ABC's (and XYZ's) of peptide sequencing," (in eng), *Nat Rev Mol Cell Biol,* vol. 5, no. 9, pp. 699-711, Sep 2004, doi: 10.1038/nrm1468.

[189] A. Ducret, I. Van Oostveen, J. K. Eng, J. R. Yates, 3rd, and R. Aebersold, "High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry," (in eng), *Protein Sci,* vol. 7, no. 3, pp. 706-19, Mar 1998, doi: 10.1002/pro.5560070320.

[190] A. Makarov and M. Scigelova, "Coupling liquid chromatography to Orbitrap mass spectrometry," *J Chromatogr A,* vol. 1217, no. 25, pp. 3938-45, Jun 18 2010, doi: 10.1016/j.chroma.2010.02.022.

[191] A. B. Chakraborty and S. J. Berger, "Optimization of reversed-phase peptide liquid chromatography ultraviolet mass spectrometry analyses using an automated blending methodology," (in eng), *J Biomol Tech,* vol. 16, no. 4, pp. 327-35, Dec 2005.

[192] A. J. Link *et al.*, "Direct analysis of protein complexes using mass spectrometry," (in eng), *Nat Biotechnol,* vol. 17, no. 7, pp. 676-82, Jul 1999, doi: 10.1038/10890.

[193] Y. Bian *et al.*, "Robust Microflow LC-MS/MS for Proteome Analysis: 38000 Runs and Counting," *Analytical chemistry,* vol. 93, no. 8, pp. 3686-3690, Mar 2 2021, doi: 10.1021/acs.analchem.1c00257.

[194] Y. Bian *et al.*, "Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC-MS/MS," *Nature communications,* vol. 11, no. 1, p. 157, Jan 9 2020, doi: 10.1038/s41467-019-13973-x.

[195] M. Wilm, "Principles of electrospray ionization," *Molecular & cellular proteomics : MCP,* vol. 10, no. 7, p. M111 009407, Jul 2011, doi: 10.1074/mcp.M111.009407.

[196] A. Gomez and K. Tang, "Charge and fission of droplets in electrostatic sprays," *Physics of Fluids,* vol. 6, no. 1, pp. 404-414, 1994, doi: 10.1063/1.868037.

[197] T. G., "Disintegration of Water Drops in an Electric Field," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 280, no. 1382, pp. 383-397, 1964, doi: 10.1098/rspa.1964.0151.

[198] B. A. Thomson and J. V. Iribarne, "Field induced ion evaporation from liquid surfaces at atmospheric pressure," *The Journal of Chemical Physics,* vol. 71, no. 11, pp. 4451-4463, 1979, doi: 10.1063/1.438198.

[199] J. V. Iribarne, "On the evaporation of small ions from charged droplets," *The Journal of Chemical Physics,* vol. 64, no. 6, 1976, doi: 10.1063/1.432536.

[200] M. Dole, L. L. Mack, R. L. Hines, R. C. Mobley, L. D. Ferguson, and M. B. Alice, "Molecular Beams of Macroions," *The Journal of Chemical Physics,* vol. 49, no. 5, pp. 2240-2249, 1968, doi: 10.1063/1.1670391.

[201] J. P. Savaryn, T. K. Toby, and N. L. Kelleher, "A researcher's guide to mass spectrometry-based proteomics," *Proteomics,* vol. 16, no. 18, pp. 2435-43, Sep 2016, doi: 10.1002/pmic.201600113.

[202] D. J. Douglas, A. J. Frank, and D. Mao, "Linear ion traps in mass spectrometry," (in eng), *Mass spectrometry reviews,* vol. 24, no. 1, pp. 1-29, Jan-Feb 2005, doi: 10.1002/mas.20004.

[203] M. Scigelova and A. Makarov, "Orbitrap mass analyzer--overview and applications in proteomics," (in eng), *Proteomics,* vol. 6 Suppl 2, pp. 16-21, Sep 2006, doi: 10.1002/pmic.200600528.

[204] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, and R. Graham Cooks, "The Orbitrap: a new mass spectrometer," (in eng), *J Mass Spectrom,* vol. 40, no. 4, pp. 430-43, Apr 2005, doi: 10.1002/jms.856.

[205] M. W. Senko *et al.*, "Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates," *Analytical chemistry,* vol. 85, no. 24, pp. 11710-4, Dec 17 2013, doi: 10.1021/ac403115c.

[206] J. Seidler, N. Zinn, M. E. Boehm, and W. D. Lehmann, "De novo sequencing of peptides by MS/MS," *Proteomics,* vol. 10, no. 4, pp. 634-49, Feb 2010, doi: 10.1002/pmic.200900459.

[207] A. Devabhaktuni and J. E. Elias, "Application of de Novo Sequencing to Large-Scale Complex Proteomics Data Sets," *Journal of proteome research,* vol. 15, no. 3, pp. 732-42, Mar 4 2016, doi: 10.1021/acs.jproteome.5b00861.

[208] J. Allmer, "Algorithms for the de novo sequencing of peptides from tandem mass spectra," (in eng), *Expert Rev Proteomics,* vol. 8, no. 5, pp. 645-57, Oct 2011, doi: 10.1586/epr.11.54.

[209] J. Cox and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nat Biotechnol,* vol. 26, no. 12, pp. 1367-72, Dec 2008, doi: 10.1038/nbt.1511.

[210] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann, "Andromeda: a peptide search engine integrated into the MaxQuant environment," *Journal of proteome research,* vol. 10, no. 4, pp. 1794-805, Apr 1 2011, doi: 10.1021/pr101065j.

[211] J. E. Elias and S. P. Gygi, "Target-decoy search strategy for mass spectrometry-based proteomics," *Methods Mol Biol,* vol. 604, pp. 55-71, 2010, doi: 10.1007/978-1-60761-444-9_5.

[212] J. V. Olsen *et al.*, "Global, in vivo, and site-specific phosphorylation dynamics in signaling networks," *Cell,* vol. 127, no. 3, pp. 635-48, Nov 3 2006, doi: 10.1016/j.cell.2006.09.026.

[213] M. M. Savitski *et al.*, "Confident phosphorylation site localization using the Mascot Delta Score," *Molecular & cellular proteomics : MCP,* vol. 10, no. 2, p. M110 003830, Feb 2011, doi: 10.1074/mcp.M110.003830.

[214] J. Grossmann *et al.*, "Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods," (in eng), *J Proteomics,* vol. 73, no. 9, pp. 1740-6, Aug 5 2010, doi: 10.1016/j.jprot.2010.05.011.

[215] J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann, "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ," (in eng), *Molecular & cellular proteomics : MCP,* vol. 13, no. 9, pp. 2513-26, Sep 2014, doi: 10.1074/mcp.M113.031591.

[216] A. B. Arul and R. A. S. Robinson, "Sample Multiplexing Strategies in Quantitative Proteomics," *Analytical chemistry,* vol. 91, no. 1, pp. 178-189, Jan 2 2019, doi: 10.1021/acs.analchem.8b05626.

[217] J. Li *et al.*, "TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing," *Journal of proteome research,* vol. 20, no. 5, pp. 2964-2972, May 7 2021, doi: 10.1021/acs.jproteome.1c00168.

[218] N. A. Karp, W. Huber, P. G. Sadowski, P. D. Charles, S. V. Hester, and K. S. Lilley, "Addressing accuracy and precision issues in iTRAQ quantitation," *Molecular & cellular proteomics : MCP,* vol. 9, no. 9, pp. 1885-97, Sep 2010, doi: 10.1074/mcp.M900628-MCP200.

[219] M. M. Savitski *et al.*, "Measuring and managing ratio compression for accurate iTRAQ/TMT quantification," *Journal of proteome research,* vol. 12, no. 8, pp. 3586-98, Aug 2 2013, doi: 10.1021/pr400098r.

[220] M. M. Savitski *et al.*, "Delayed fragmentation and optimized isolation width settings for improvement of protein identification and accuracy of isobaric mass tag quantification on Orbitrap-type mass spectrometers," *Analytical chemistry,* vol. 83, no. 23, pp. 8959-67, Dec 1 2011, doi: 10.1021/ac201760x.

[221] C. D. Wenger *et al.*, "Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging," *Nature methods,* vol. 8, no. 11, pp. 933-5, Oct 2 2011, doi: 10.1038/nmeth.1716.

[222] M. Wuhr *et al.*, "Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster," *Analytical chemistry,* vol. 84, no. 21, pp. 9214-21, Nov 6 2012, doi: 10.1021/ac301962s.

[223] L. Ting, R. Rad, S. P. Gygi, and W. Haas, "MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics," *Nature methods,* vol. 8, no. 11, pp. 937-40, Oct 2 2011, doi: 10.1038/nmeth.1714.

[224] G. C. McAlister *et al.*, "MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes," *Analytical chemistry,* vol. 86, no. 14, pp. 7150-8, Jul 15 2014, doi: 10.1021/ac502040v.

[225] I. Virant-Klun, S. Leicht, C. Hughes, and J. Krijgsveld, "Identification of Maturation-Specific Proteins by Single-Cell Proteomics of Human Oocytes," (in eng), *Molecular & cellular proteomics : MCP,* vol. 15, no. 8, pp. 2616-27, Aug 2016, doi: 10.1074/mcp.M115.056887.

[226] P. Cohen, "The origins of protein phosphorylation," (in eng), *Nature cell biology,* vol. 4, no. 5, pp. E127-30, May 2002, doi: 10.1038/ncb0502-e127.

[227] H. Zhou *et al.*, "Zirconium phosphonate-modified porous silicon for highly specific capture of phosphopeptides and MALDI-TOF MS analysis," (in eng), *Journal of proteome research,* vol. 5, no. 9, pp. 2431-7, Sep 2006, doi: 10.1021/pr060162f.

[228] H. Zhou *et al.*, "Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis," (in eng), *Journal of proteome research,* vol. 7, no. 9, pp. 3957-67, Sep 2008, doi: 10.1021/pr800223m.

[229] T. Y. Low, M. A. Mohtar, P. Y. Lee, N. Omar, H. Zhou, and M. Ye, "WIDENING THE BOTTLENECK OF PHOSPHOPROTEOMICS: EVOLVING STRATEGIES FOR PHOSPHOPEPTIDE

ENRICHMENT," (in eng), *Mass spectrometry reviews,* vol. 40, no. 4, pp. 309-333, Jul 2021, doi: 10.1002/mas.21636.

[230] S. J. Humphrey, O. Karayel, D. E. James, and M. Mann, "High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform," *Nature protocols,* vol. 13, no. 9, pp. 1897-1916, 2018, doi: 10.1038/s41596-018-0014-9.

[231] S. J. Humphrey, S. B. Azimifar, and M. Mann, "High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics," *Nature Biotechnology,* vol. 33, no. 9, pp. 990-995, 2015, doi: 10.1038/nbt.3327.

[232] W. Chen, L. Chen, and R. Tian, "An integrated strategy for highly sensitive phosphoproteome analysis from low micrograms of protein samples," *Analyst,* vol. 143, no. 15, pp. 3693-3701, Jul 23 2018, doi: 10.1039/c8an00792f.

[233] W. Chen *et al.*, "Simple and Integrated Spintip-Based Technology Applied for Deep Proteome Profiling," (in eng), *Analytical chemistry,* vol. 88, no. 9, pp. 4864-71, May 3 2016, doi: 10.1021/acs.analchem.6b00631.

[234] J. Zecha *et al.*, "TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach," (in eng), *Molecular & cellular proteomics : MCP,* vol. 18, no. 7, pp. 1468-1478, Jul 2019, doi: 10.1074/mcp.TIR119.001385.

[235] C. S. Hughes, S. Moggridge, T. Muller, P. H. Sorensen, G. B. Morin, and J. Krijgsveld, "Single-pot, solid-phase-enhanced sample preparation for proteomics experiments," *Nature protocols,* vol. 14, no. 1, pp. 68-85, Jan 2019, doi: 10.1038/s41596-018-0082-x.

[236] B. Ruprecht, J. Zecha, D. P. Zolg, and B. Kuster, "High pH Reversed-Phase Micro-Columns for Simple, Sensitive, and Efficient Fractionation of Proteome and (TMT labeled) Phosphoproteome Digests," (in eng), *Methods Mol Biol,* vol. 1550, pp. 83-98, 2017, doi: 10.1007/978-1-4939-6747-6_8.

[237] S. Tyanova, T. Temu, and J. Cox, "The MaxQuant computational platform for mass spectrometry-based shotgun proteomics," (in eng), *Nature protocols,* vol. 11, no. 12, pp. 2301-2319, Dec 2016, doi: 10.1038/nprot.2016.136.

[238] S. Tyanova *et al.*, "The Perseus computational platform for comprehensive analysis of (prote)omics data," *Nature methods,* vol. 13, no. 9, pp. 731-40, Sep 2016, doi: 10.1038/nmeth.3901.

[239] J. Cox and M. Mann, "1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data," (in eng), *BMC Bioinformatics,* vol. 13 Suppl 16, no. Suppl 16, p. S12, 2012, doi: 10.1186/1471-2105-13-s16-s12.

[240] S. Moggridge, P. H. Sorensen, G. B. Morin, and C. S. Hughes, "Extending the Compatibility of the SP3 Paramagnetic Bead Processing Approach for Proteomics," (in eng), *Journal of proteome research,* vol. 17, no. 4, pp. 1730-1740, Apr 6 2018, doi: 10.1021/acs.jproteome.7b00913.

[241] H. Steen, J. A. Jebanathirajah, J. Rush, N. Morrice, and M. W. Kirschner, "Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements," *Molecular & cellular proteomics : MCP,* vol. 5, no. 1, pp. 172-81, Jan 2006, doi: 10.1074/mcp.M500135-MCP200.

[242] M. Marcantonio, M. Trost, M. Courcelles, M. Desjardins, and P. Thibault, "Combined enzymatic and data mining approaches for comprehensive phosphoproteome analyses: application to cell signaling events of interferon-gamma-stimulated macrophages," *Molecular & cellular proteomics : MCP,* vol. 7, no. 4, pp. 645-60, Apr 2008, doi: 10.1074/mcp.M700383-MCP200.

[243] Q. R. Li, Z. B. Ning, J. S. Tang, S. Nie, and R. Zeng, "Effect of peptide-to-TiO2 beads ratio on phosphopeptide enrichment selectivity," (in eng), *Journal of proteome research,* vol. 8, no. 11, pp. 5375-81, Nov 2009, doi: 10.1021/pr900659n.

[244] C. F. Tsai *et al.*, "Immobilized metal affinity chromatography revisited: pH/acid control toward high selectivity in phosphoproteomics," (in eng), *Journal of proteome research,* vol. 7, no. 9, pp. 4058-69, Sep 2008, doi: 10.1021/pr800364d.

[245] M. Leutert, R. A. Rodriguez-Mias, N. K. Fukuda, and J. Villen, "R2-P2 rapid-robotic phosphoproteomics enables multidimensional cell signaling studies," *Molecular systems biology,* vol. 15, no. 12, p. e9021, Dec 2019, doi: 10.15252/msb.20199021.

[246] C. J. Tape *et al.*, "Reproducible automated phosphopeptide enrichment using magnetic TiO2 and Ti-IMAC," *Analytical chemistry,* vol. 86, no. 20, pp. 10296-302, Oct 21 2014, doi: 10.1021/ac5025842.

[247] J. Krumm *et al.*, "High temporal resolution proteome and phosphoproteome profiling of stem cell-derived hepatocyte development," (in eng), *Cell reports,* vol. 38, no. 13, p. 110604, Mar 29 2022, doi: 10.1016/j.celrep.2022.110604.

[248] P. W. Burridge and E. T. Zambidis, "Highly efficient directed differentiation of human induced pluripotent stem cells into cardiomyocytes," (in eng), *Methods Mol Biol,* vol. 997, pp. 149-61, 2013, doi: 10.1007/978-1-62703-348-0_12.

[249] R. S. Kadzik and E. E. Morrisey, "Directing lung endoderm differentiation in pluripotent stem cells," (in eng), *Cell stem cell,* vol. 10, no. 4, pp. 355-61, Apr 6 2012, doi: 10.1016/j.stem.2012.03.013.

[250] J. Tchieu *et al.*, "A Modular Platform for Differentiation of Human PSCs into All Major Ectodermal Lineages," (in eng), *Cell stem cell,* vol. 21, no. 3, pp. 399-410.e7, Sep 7 2017, doi: 10.1016/j.stem.2017.08.015.

[251] J. R. Spence *et al.*, "Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro," (in eng), *Nature,* vol. 470, no. 7332, pp. 105-9, Feb 3 2011, doi: 10.1038/nature09691.

[252] Y. Kunisada, N. Tsubooka-Yamazoe, M. Shoji, and M. Hosoya, "Small molecules induce efficient differentiation into insulin-producing cells from human induced pluripotent stem cells," (in eng), *Stem Cell Res,* vol. 8, no. 2, pp. 274-84, Mar 2012, doi: 10.1016/j.scr.2011.10.002.

[253] K. Sekine *et al.*, "Generation of human induced pluripotent stem cell-derived liver buds with chemically defined and animal origin-free media," *Scientific reports,* vol. 10, no. 1, p. 17937, Oct 21 2020, doi: 10.1038/s41598-020-73908-1.

[254] D. Vyas *et al.*, "Self-assembled liver organoids recapitulate hepatobiliary organogenesis in vitro," *Hepatology (Baltimore, Md.),* vol. 67, no. 2, pp. 750-761, Feb 2018, doi: 10.1002/hep.29483.

[255] L. T. Ang *et al.*, "A Roadmap for Human Liver Differentiation from Pluripotent Stem Cells," (in eng), *Cell reports,* vol. 22, no. 8, pp. 2190-2205, Feb 20 2018, doi: 10.1016/j.celrep.2018.01.087.

[256] K. T. Rigbolt *et al.*, "System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation," (in eng), *Science signaling,* vol. 4, no. 164, p. rs3, Mar 15 2011, doi: 10.1126/scisignal.2001570.

[257] D. van Hoof, J. Krijgsveld, and C. Mummery, "Proteomic analysis of stem cell differentiation and early development," (in eng), *Cold Spring Harbor perspectives in biology,* vol. 4, no. 3, Mar 1 2012, doi: 10.1101/cshperspect.a008177.

[258] M. Baxter *et al.*, "Phenotypic and functional analyses show stem cell-derived hepatocyte-like cells better mimic fetal rather than adult hepatocytes," (in eng), *Journal of hepatology,* vol. 62, no. 3, pp. 581-9, Mar 2015, doi: 10.1016/j.jhep.2014.10.016.

[259] C. Rowe *et al.*, "Proteome-wide analyses of human hepatocytes during differentiation and dedifferentiation," (in eng), *Hepatology (Baltimore, Md.),* vol. 58, no. 2, pp. 799-809, Aug 2013, doi: 10.1002/hep.26414.

[260] N. Takayama *et al.*, "Transient activation of c-MYC expression is critical for efficient platelet generation from human induced pluripotent stem cells," (in eng), *The Journal of*

experimental medicine, vol. 207, no. 13, pp. 2817-30, Dec 20 2010, doi: 10.1084/jem.20100844.

[261] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical Bayes methods," *Biostatistics,* vol. 8, no. 1, pp. 118-27, Jan 2007, doi: 10.1093/biostatistics/kxj037.

[262] G. Yu and Q. Y. He, "ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization," (in eng), *Mol Biosyst,* vol. 12, no. 2, pp. 477-9, Feb 2016, doi: 10.1039/c5mb00663e.

[263] H. Han *et al.*, "TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions," *Nucleic Acids Res,* vol. 46, no. D1, pp. D380-D386, Jan 4 2018, doi: 10.1093/nar/gkx1013.

[264] H. Horn *et al.*, "KinomeXplorer: an integrated platform for kinome biology studies," *Nature methods,* vol. 11, no. 6, pp. 603-4, Jun 2014, doi: 10.1038/nmeth.2968.

[265] R. Linding *et al.*, "Systematic discovery of in vivo phosphorylation networks," (in eng), *Cell,* vol. 129, no. 7, pp. 1415-26, Jun 29 2007, doi: 10.1016/j.cell.2007.05.052.

[266] P. N. Ghule *et al.*, "Reprogramming the pluripotent cell cycle: restoration of an abbreviated G1 phase in human induced pluripotent stem (iPS) cells," *J Cell Physiol,* vol. 226, no. 5, pp. 1149-56, May 2011, doi: 10.1002/jcp.22440.

[267] K. A. Becker, J. L. Stein, J. B. Lian, A. J. van Wijnen, and G. S. Stein, "Human embryonic stem cells are pre-mitotically committed to self-renewal and acquire a lengthened G1 phase upon lineage programming," *J Cell Physiol,* vol. 222, no. 1, pp. 103-10, Jan 2010, doi: 10.1002/jcp.21925.

[268] C. Hindley and A. Philpott, "The cell cycle and pluripotency," *Biochem J,* vol. 451, no. 2, pp. 135-43, Apr 15 2013, doi: 10.1042/BJ20121627.

[269] R. de Sousa Abreu, L. O. Penalva, E. M. Marcotte, and C. Vogel, "Global signatures of protein and mRNA expression levels," *Mol Biosyst,* vol. 5, no. 12, pp. 1512-26, Dec 2009, doi: 10.1039/b908315d.

[270] S. Varum *et al.*, "Energy metabolism in human pluripotent stem cells and their differentiated counterparts," *PLoS One,* vol. 6, no. 6, p. e20914, 2011, doi: 10.1371/journal.pone.0020914.

[271] O. Warburg, F. Wind, and E. Negelein, "THE METABOLISM OF TUMORS IN THE BODY," (in eng), *J Gen Physiol,* vol. 8, no. 6, pp. 519-30, Mar 7 1927, doi: 10.1085/jgp.8.6.519.

[272] D. Bausch-Fluck *et al.*, "A mass spectrometric-derived cell surface protein atlas," *PLoS One,* vol. 10, no. 3, p. e0121314, 2015, doi: 10.1371/journal.pone.0121314.

[273] Y. A. Medvedeva *et al.*, "EpiFactors: a comprehensive database of human epigenetic factors and complexes," *Database (Oxford),* vol. 2015, p. bav067, 2015, doi: 10.1093/database/bav067.

[274] R. Liang *et al.*, "Restraining Lysosomal Activity Preserves Hematopoietic Stem Cell Quiescence and Potency," *Cell stem cell,* vol. 26, no. 3, pp. 359-376 e7, Mar 5 2020, doi: 10.1016/j.stem.2020.01.013.

[275] S. Mahanty *et al.*, "Keratinocyte differentiation promotes ER stress-dependent lysosome biogenesis," *Cell Death & Disease,* vol. 10, no. 4, 2019, doi: 10.1038/s41419-019-1478-4.

[276] R. E. Lawrence and R. Zoncu, "The lysosome as a cellular centre for signalling, metabolism and quality control," *Nature cell biology,* vol. 21, no. 2, pp. 133-142, Feb 2019, doi: 10.1038/s41556-018-0244-7.

[277] S. M. Ferguson, "Beyond indigestion: emerging roles for lysosome-based signaling in human disease," *Curr Opin Cell Biol,* vol. 35, pp. 59-68, Aug 2015, doi: 10.1016/j.ceb.2015.04.014.

[278] T. Berg *et al.*, "Fibroblast growth factor 10 is critical for liver growth during embryogenesis and controls hepatoblast survival via beta-catenin activation," *Hepatology (Baltimore, Md.),* vol. 46, no. 4, pp. 1187-97, Oct 2007, doi: 10.1002/hep.21814.

[279] A. Drazic, L. M. Myklebust, R. Ree, and T. Arnesen, "The world of protein acetylation," (in eng), *Biochim Biophys Acta,* vol. 1864, no. 10, pp. 1372-401, Oct 2016, doi: 10.1016/j.bbapap.2016.06.007.

[280] J. E. Choi and R. Mostoslavsky, "Sirtuins, metabolism, and DNA repair," *Curr Opin Genet Dev,* vol. 26, pp. 24-32, Jun 2014, doi: 10.1016/j.gde.2014.05.005.

[281] Y. Cha *et al.*, "Metabolic control of primed human pluripotent stem cell fate and function by the miR-200c-SIRT2 axis," (in eng), *Nature cell biology,* vol. 19, no. 5, pp. 445-456, May 2017, doi: 10.1038/ncb3517.

[282] B. S. Kim, C. H. Lee, G. E. Chang, E. Cheong, and I. Shin, "A potent and selective small molecule inhibitor of sirtuin 1 promotes differentiation of pluripotent P19 cells into functional neurons," *Scientific reports,* vol. 6, p. 34324, Sep 29 2016, doi: 10.1038/srep34324.

[283] T. Tobita *et al.*, "SIRT1 Disruption in Human Fetal Hepatocytes Leads to Increased Accumulation of Glucose and Lipids," *PLoS One,* vol. 11, no. 2, p. e0149344, 2016, doi: 10.1371/journal.pone.0149344.

[284] B. Hu *et al.*, "Repression of SIRT1 promotes the differentiation of mouse induced pluripotent stem cells into neural stem cells," (in eng), *Cell Mol Neurobiol,* vol. 34, no. 6, pp. 905-12, Aug 2014, doi: 10.1007/s10571-014-0071-8.

[285] Y. Xu *et al.*, "Oxidative stress activates SIRT2 to deacetylate and stimulate phosphoglycerate mutase," *Cancer Res,* vol. 74, no. 13, pp. 3630-42, Jul 1 2014, doi: 10.1158/0008-5472.CAN-13-3615.

[286] X. Si *et al.*, "Activation of GSK3beta by Sirt2 is required for early lineage commitment of mouse embryonic stem cell," *PLoS One,* vol. 8, no. 10, p. e76699, 2013, doi: 10.1371/journal.pone.0076699.

[287] Y. L. Lee *et al.*, "Sirtuin 1 facilitates generation of induced pluripotent stem cells from mouse embryonic fibroblasts through the miR-34a and p53 pathways," *PLoS One,* vol. 7, no. 9, p. e45633, 2012, doi: 10.1371/journal.pone.0045633.

[288] V. Calvanese *et al.*, "Sirtuin 1 regulation of developmental genes during differentiation of stem cells," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 107, no. 31, pp. 13736-41, Aug 3 2010, doi: 10.1073/pnas.1001399107.

[289] M. J. Gómez-Lechón, L. Tolosa, I. Conde, and M. T. Donato, "Competency of different cell models to predict human hepatotoxic drugs," (in eng), *Expert Opin Drug Metab Toxicol,* vol. 10, no. 11, pp. 1553-68, Nov 2014, doi: 10.1517/17425255.2014.967680.

[290] J. Basta and M. Rauchman, "The nucleosome remodeling and deacetylase complex in development and disease," *Transl Res,* vol. 165, no. 1, pp. 36-47, Jan 2015, doi: 10.1016/j.trsl.2014.05.003.

[291] S. L. Kloet *et al.*, "NuRD-interacting protein ZFP296 regulates genome-wide NuRD localization and differentiation of mouse embryonic stem cells," *Nature communications,* vol. 9, no. 1, p. 4588, Nov 2 2018, doi: 10.1038/s41467-018-07063-7.

[292] N. Reynolds *et al.*, "NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment," *Cell stem cell,* vol. 10, no. 5, pp. 583-94, May 4 2012, doi: 10.1016/j.stem.2012.02.020.

[293] R. D. Kelly and S. M. Cowley, "The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts," (in eng), *Biochem Soc Trans,* vol. 41, no. 3, pp. 741-9, Jun 2013, doi: 10.1042/bst20130010.

[294] G. A. Challen *et al.*, "Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells," *Cell stem cell,* vol. 15, no. 3, pp. 350-364, Sep 4 2014, doi: 10.1016/j.stem.2014.06.018.

[295] M. Jackson *et al.*, "Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells," *Mol Cell Biol,* vol. 24, no. 20, pp. 8862-71, Oct 2004, doi: 10.1128/MCB.24.20.8862-8871.2004.

[296] Z. D. Smith and A. Meissner, "DNA methylation: roles in mammalian development," *Nat Rev Genet,* vol. 14, no. 3, pp. 204-20, Mar 2013, doi: 10.1038/nrg3354.

[297] Q. Li *et al.*, "A sequential EMT-MET mechanism drives the differentiation of human embryonic stem cells towards hepatocytes," *Nature communications,* vol. 8, p. 15166, May 3 2017, doi: 10.1038/ncomms15166.

[298] C. Ding *et al.*, "Gab2 facilitates epithelial-to-mesenchymal transition via the MEK/ERK/MMP signaling in colorectal cancer," *J Exp Clin Cancer Res,* vol. 35, p. 5, Jan 12 2016, doi: 10.1186/s13046-015-0280-0.

[299] K. Matsusue *et al.*, "Hepatic CCAAT/enhancer binding protein alpha mediates induction of lipogenesis and regulation of glucose homeostasis in leptin-deficient mice," (in eng), *Mol Endocrinol,* vol. 18, no. 11, pp. 2751-64, Nov 2004, doi: 10.1210/me.2004-0213.

[300] N. D. Wang *et al.*, "Impaired energy homeostasis in C/EBP alpha knockout mice," (in eng), *Science (New York, N.Y.),* vol. 269, no. 5227, pp. 1108-12, Aug 25 1995, doi: 10.1126/science.7652557.

[301] L. E. Olofsson *et al.*, "CCAAT/enhancer binding protein alpha (C/EBPalpha) in adipose tissue regulates genes in lipid and glucose metabolism and a genetic variation in C/EBPalpha is associated with serum levels of triglycerides," (in eng), *J Clin Endocrinol Metab,* vol. 93, no. 12, pp. 4880-6, Dec 2008, doi: 10.1210/jc.2008-0574.

[302] A. Westmacott, Z. D. Burke, G. Oliver, J. M. Slack, and D. Tosh, "C/EBPalpha and C/EBPbeta are markers of early liver development," *Int J Dev Biol,* vol. 50, no. 7, pp. 653-7, 2006, doi: 10.1387/ijdb.062146aw.

[303] B. Wang, C. Gao, and K. P. Ponder, "C/EBPbeta contributes to hepatocyte growth factor-induced replication of rodent hepatocytes," (in eng), *Journal of hepatology,* vol. 43, no. 2, pp. 294-302, Aug 2005, doi: 10.1016/j.jhep.2005.02.029.

[304] L. E. Greenbaum, D. E. Cressman, B. A. Haber, and R. Taub, "Coexistence of C/EBP alpha, beta, growth-induced proteins and DNA synthesis in hepatocytes during liver regeneration. Implications for maintenance of the differentiated state during liver growth," (in eng), *J Clin Invest,* vol. 96, no. 3, pp. 1351-65, Sep 1995, doi: 10.1172/jci118170.

[305] C. A. Alper, A. M. Johnson, A. G. Birtch, and F. D. Moore, "Human C'3: evidence for the liver as the primary site of synthesis," (in eng), *Science (New York, N.Y.),* vol. 163, no. 3864, pp. 286-8, Jan 17 1969, doi: 10.1126/science.163.3864.286.

[306] Z. Zhou, M. J. Xu, and B. Gao, "Hepatocytes: a key cell type for innate immunity," *Cell Mol Immunol,* vol. 13, no. 3, pp. 301-15, May 2016, doi: 10.1038/cmi.2015.97.

[307] A. B. McLean *et al.*, "Activin a efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed," (in eng), *Stem cells (Dayton, Ohio),* vol. 25, no. 1, pp. 29-38, Jan 2007, doi: 10.1634/stemcells.2006-0219.

[308] E. H. Lee *et al.*, "Disruption of the non-canonical WNT pathway in lung squamous cell carcinoma," (in eng), *Clin Med Oncol,* vol. 2008, no. 2, pp. 169-179, Apr 1 2008.

[309] A. Sato, H. Yamamoto, H. Sakane, H. Koyama, and A. Kikuchi, "Wnt5a regulates distinct signalling pathways by binding to Frizzled2," *EMBO J,* vol. 29, no. 1, pp. 41-54, Jan 6 2010, doi: 10.1038/emboj.2009.322.

[310] J. A. Bisson, B. Mills, J. C. Paul Helt, T. P. Zwaka, and E. D. Cohen, "Wnt5a and Wnt11 inhibit the canonical Wnt pathway and promote cardiac progenitor development via the Caspase-dependent degradation of AKT," *Dev Biol,* vol. 398, no. 1, pp. 80-96, Feb 1 2015, doi: 10.1016/j.ydbio.2014.11.015.

[311] M. J. Nemeth, L. Topol, S. M. Anderson, Y. Yang, and D. M. Bodine, "Wnt5a inhibits canonical Wnt signaling in hematopoietic stem cells and enhances repopulation," (in eng), *Proceedings of the National Academy of Sciences of the United States of America,* vol. 104, no. 39, pp. 15436-41, Sep 25 2007, doi: 10.1073/pnas.0704747104.

[312] M. H. Kang *et al.*, "Estrogen-related receptor gamma functions as a tumor suppressor in gastric cancer," *Nature communications,* vol. 9, no. 1, p. 1920, May 15 2018, doi: 10.1038/s41467-018-04244-2.

[313] Y. Ahn *et al.*, "Multiple modes of Lrp4 function in modulation of Wnt/beta-catenin signaling during tooth development," *Development (Cambridge, England),* vol. 144, no. 15, pp. 2824-2836, Aug 1 2017, doi: 10.1242/dev.150680.

[314] K. Tamai *et al.*, "LDL-receptor-related proteins in Wnt signal transduction," (in eng), *Nature,* vol. 407, no. 6803, pp. 530-5, Sep 28 2000, doi: 10.1038/35035117.

[315] P. May, E. Woldt, R. L. Matz, and P. Boucher, "The LDL receptor-related protein (LRP) family: an old family of proteins with new physiological functions," *Ann Med,* vol. 39, no. 3, pp. 219-28, 2007, doi: 10.1080/07853890701214881.

[316] E. I. Christensen *et al.*, "Evidence for an essential role of megalin in transepithelial transport of retinol," (in eng), *J Am Soc Nephrol,* vol. 10, no. 4, pp. 685-95, Apr 1999, doi: 10.1681/asn.V104685.

[317] S. Stefansson, D. A. Chappell, K. M. Argraves, D. K. Strickland, and W. S. Argraves, "Glycoprotein 330/low density lipoprotein receptor-related protein-2 mediates endocytosis of low density lipoproteins via interaction with apolipoprotein B100," *J Biol Chem,* vol. 270, no. 33, pp. 19417-21, Aug 18 1995, doi: 10.1074/jbc.270.33.19417.

[318] A. Niemeier *et al.*, "Identification of megalin/gp330 as a receptor for lipoprotein(a) in vitro," (in eng), *Arterioscler Thromb Vasc Biol,* vol. 19, no. 3, pp. 552-61, Mar 1999, doi: 10.1161/01.atv.19.3.552.

[319] J. Sahi, S. Grepper, and C. Smith, "Hepatocytes as a tool in drug metabolism, transport and safety evaluations in drug discovery," (in eng), *Curr Drug Discov Technol,* vol. 7, no. 3, pp. 188-98, Sep 2010, doi: 10.2174/157016310793180576.

[320] J. Shan *et al.*, "Identification of small molecules for human hepatocyte expansion and iPS differentiation," (in eng), *Nature chemical biology,* vol. 9, no. 8, pp. 514-20, Aug 2013, doi: 10.1038/nchembio.1270.

[321] C. C. Bell *et al.*, "Characterization of primary human hepatocyte spheroids as a model system for drug-induced liver injury, liver function and disease," (in eng), *Scientific reports,* vol. 6, p. 25187, May 4 2016, doi: 10.1038/srep25187.

[322] J. V. Castell, R. Jover, C. P. Martínez-Jiménez, and M. J. Gómez-Lechón, "Hepatocyte cell lines: their use, scope and limitations in drug metabolism studies," (in eng), *Expert Opin Drug Metab Toxicol,* vol. 2, no. 2, pp. 183-212, Apr 2006, doi: 10.1517/17425255.2.2.183.

[323] S. Bartfeld and H. Clevers, "Stem cell-derived organoids and their application for medical research and patient treatment," (in eng), *Journal of molecular medicine (Berlin, Germany),* vol. 95, no. 7, pp. 729-738, Jul 2017, doi: 10.1007/s00109-017-1531-7.

[324] M. D. Davidson, B. R. Ware, and S. R. Khetani, "Stem cell-derived liver cells for drug testing and disease modeling," (in eng), *Discov Med,* vol. 19, no. 106, pp. 349-58, May 2015.

[325] T. Takebe *et al.*, "Generation of a vascularized and functional human liver from an iPSC-derived organ bud transplant," *Nature protocols,* vol. 9, no. 2, pp. 396-409, Feb 2014, doi: 10.1038/nprot.2014.020.

[326] V. Kegel, D. Deharde, E. Pfeiffer, K. Zeilinger, D. Seehofer, and G. Damm, "Protocol for Isolation of Primary Human Hepatocytes and Corresponding Major Populations of Non-parenchymal Liver Cells," *J Vis Exp,* no. 109, p. e53069, Mar 30 2016, doi: 10.3791/53069.

[327] D. M. Keller *et al.*, "A DNA damage-induced p53 serine 392 kinase complex contains CK2, hSpt16, and SSRP1," (in eng), *Mol Cell,* vol. 7, no. 2, pp. 283-92, Feb 2001, doi: 10.1016/s1097-2765(01)00176-9.

[328] M. Sayed, S. Pelech, C. Wong, A. Marotta, and B. Salh, "Protein kinase CK2 is involved in G2 arrest and apoptosis following spindle damage in epithelial cells," (in eng), *Oncogene,* vol. 20, no. 48, pp. 6994-7005, Oct 25 2001, doi: 10.1038/sj.onc.1204894.

[329] J. L. Brockman, S. D. Gross, M. R. Sussman, and R. A. Anderson, "Cell cycle-dependent localization of casein kinase I to mitotic spindles," (in eng), *Proceedings of the National Academy of Sciences of the United States of America,* vol. 89, no. 20, pp. 9454-8, Oct 15 1992, doi: 10.1073/pnas.89.20.9454.

[330] S. Jiang, M. Zhang, J. Sun, and X. Yang, "Casein kinase 1α: biological mechanisms and theranostic potential," (in eng), *Cell Commun Signal,* vol. 16, no. 1, p. 23, May 24 2018, doi: 10.1186/s12964-018-0236-z.

[331] T. Suga, H. Yamaguchi, T. Sato, M. Maekawa, J. Goto, and N. Mano, "Preference of Conjugated Bile Acids over Unconjugated Bile Acids as Substrates for OATP1B1 and OATP1B3," *PLoS One,* vol. 12, no. 1, p. e0169719, 2017, doi: 10.1371/journal.pone.0169719.

[332] K. Maeda, M. Kambara, Y. Tian, A. F. Hofmann, and Y. Sugiyama, "Uptake of ursodeoxycholate and its conjugates by human hepatocytes: role of Na(+)-taurocholate cotransporting polypeptide (NTCP), organic anion transporting polypeptide (OATP) 1B1 (OATP-C), and oatp1B3 (OATP8)," (in eng), *Mol Pharm,* vol. 3, no. 1, pp. 70-7, Jan-Feb 2006, doi: 10.1021/mp050063u.

[333] H. Huang *et al.*, "Human FABP1 T94A variant enhances cholesterol uptake," *Biochim Biophys Acta,* vol. 1851, no. 7, pp. 946-55, Jul 2015, doi: 10.1016/j.bbalip.2015.02.015.

[334] J. Herbert *et al.*, "Transthyretin: a choroid plexus-specific transport protein in human brain. The 1986 S. Weir Mitchell award," (in eng), *Neurology,* vol. 36, no. 7, pp. 900-11, Jul 1986, doi: 10.1212/wnl.36.7.900.

[335] D. K. Kim *et al.*, "Characterization of the system L amino acid transporter in T24 human bladder carcinoma cells," (in eng), *Biochim Biophys Acta,* vol. 1565, no. 1, pp. 112-21, Sep 20 2002, doi: 10.1016/s0005-2736(02)00516-3.

[336] E. C. Friesema *et al.*, "Thyroid hormone transport by the heterodimeric human system L amino acid transporter," (in eng), *Endocrinology,* vol. 142, no. 10, pp. 4339-48, Oct 2001, doi: 10.1210/endo.142.10.8418.

[337] M. Geiszt, J. Witta, J. Baffi, K. Lekstrom, and T. L. Leto, "Dual oxidases represent novel hydrogen peroxide sources supporting mucosal surface host defense," (in eng), *Faseb j,* vol. 17, no. 11, pp. 1502-4, Aug 2003, doi: 10.1096/fj.02-1104fje.

[338] A. Schroder *et al.*, "Genomics of ADME gene expression: mapping expression quantitative trait loci relevant for absorption, distribution, metabolism and excretion of drugs in human liver," *Pharmacogenomics J,* vol. 13, no. 1, pp. 12-20, Feb 2013, doi: 10.1038/tpj.2011.44.

[339] H. H. Lau, N. H. J. Ng, L. S. W. Loo, J. B. Jasmen, and A. K. K. Teo, "The molecular functions of hepatocyte nuclear factors - In and beyond the liver," *Journal of hepatology,* vol. 68, no. 5, pp. 1033-1048, May 2018, doi: 10.1016/j.jhep.2017.11.026.

[340] E. Currie, A. Schulze, R. Zechner, T. C. Walther, and R. V. Farese, Jr., "Cellular fatty acid metabolism and cancer," *Cell metabolism,* vol. 18, no. 2, pp. 153-61, Aug 6 2013, doi: 10.1016/j.cmet.2013.05.017.

[341] F. Rohrig and A. Schulze, "The multifaceted roles of fatty acid synthesis in cancer," *Nat Rev Cancer,* vol. 16, no. 11, pp. 732-749, Nov 2016, doi: 10.1038/nrc.2016.89.

[342] M. Knobloch *et al.*, "Metabolic control of adult neural stem cell activity by Fasn-dependent lipogenesis," *Nature,* vol. 493, no. 7431, pp. 226-30, Jan 10 2013, doi: 10.1038/nature11689.

[343] J. A. Park, S. Park, W. Y. Park, M. K. Han, and Y. Lee, "Splitomicin, a SIRT1 Inhibitor, Enhances Hematopoietic Differentiation of Mouse Embryonic Stem Cells," *Int J Stem Cells,* vol. 12, no. 1, pp. 21-30, Mar 30 2019, doi: 10.15283/ijsc18040.

[344] S. L. Wild, A. Elghajiji, C. Grimaldos Rodriguez, S. D. Weston, Z. D. Burke, and D. Tosh, "The Canonical Wnt Pathway as a Key Regulator in Liver Development, Differentiation and

Homeostatic Renewal," *Genes (Basel),* vol. 11, no. 10, Sep 30 2020, doi: 10.3390/genes11101163.

[345] B. Boward, T. Wu, and S. Dalton, "Concise Review: Control of Cell Fate Through Cell Cycle and Pluripotency Networks," *Stem cells (Dayton, Ohio),* vol. 34, no. 6, pp. 1427-36, Jun 2016, doi: 10.1002/stem.2345.

[346] I. Neganova *et al.*, "CDK1 plays an important role in the maintenance of pluripotency and genomic stability in human pluripotent stem cells," *Cell Death Dis,* vol. 5, p. e1508, Nov 6 2014, doi: 10.1038/cddis.2014.464.

[347] J. Heo *et al.*, "Phosphorylation of TFCP2L1 by CDK1 is required for stem cell pluripotency and bladder carcinogenesis," *EMBO Mol Med,* vol. 12, no. 1, p. e10880, Jan 9 2020, doi: 10.15252/emmm.201910880.

[348] W. Michowski *et al.*, "Cdk1 Controls Global Epigenetic Landscape in Embryonic Stem Cells," (in eng), *Mol Cell,* vol. 78, no. 3, pp. 459-476.e13, May 7 2020, doi: 10.1016/j.molcel.2020.03.010.

[349] L. Li *et al.*, "Cdk1 interplays with Oct4 to repress differentiation of embryonic stem cells into trophectoderm," (in eng), *FEBS Lett,* vol. 586, no. 23, pp. 4100-7, Nov 30 2012, doi: 10.1016/j.febslet.2012.10.030.

[350] T. Kunath, M. K. Saba-El-Leil, M. Almousailleakh, J. Wray, S. Meloche, and A. Smith, "FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment," *Development (Cambridge, England),* vol. 134, no. 16, pp. 2895-902, Aug 2007, doi: 10.1242/dev.02880.

[351] A. Piliszek, Z. E. Madeja, and B. Plusa, "Suppression of ERK signalling abolishes primitive endoderm formation but does not promote pluripotency in rabbit embryo," *Development (Cambridge, England),* vol. 144, no. 20, pp. 3719-3730, Oct 15 2017, doi: 10.1242/dev.156406.

[352] M. Roode *et al.*, "Human hypoblast formation is not dependent on FGF signalling," *Dev Biol,* vol. 361, no. 2, pp. 358-63, Jan 15 2012, doi: 10.1016/j.ydbio.2011.10.030.

[353] Q. L. Ying *et al.*, "The ground state of embryonic stem cell self-renewal," *Nature,* vol. 453, no. 7194, pp. 519-23, May 22 2008, doi: 10.1038/nature06968.

[354] R. K. Jaiswal, N. Jaiswal, S. P. Bruder, G. Mbalaviele, D. R. Marshak, and M. F. Pittenger, "Adult human mesenchymal stem cell differentiation to the osteogenic or adipogenic lineage is regulated by mitogen-activated protein kinase," *J Biol Chem,* vol. 275, no. 13, pp. 9645-52, Mar 31 2000, doi: 10.1074/jbc.275.13.9645.

[355] Y. Meng *et al.*, "Nicotinamide Promotes Cell Survival and Differentiation as Kinase Inhibitor in Human Pluripotent Stem Cells," *Stem cell reports,* vol. 11, no. 6, pp. 1347-1356, Dec 11 2018, doi: 10.1016/j.stemcr.2018.10.023.

[356] Y. Yu *et al.*, "ERK inhibition promotes neuroectodermal precursor commitment by blocking self-renewal and primitive streak formation of the epiblast," *Stem Cell Res Ther,* vol. 9, no. 1, p. 2, Jan 5 2018, doi: 10.1186/s13287-017-0750-8.

[357] Y. Zhang *et al.*, "Nicotinamide promotes pancreatic differentiation through the dual inhibition of CK1 and ROCK kinases in human embryonic stem cells," *Stem Cell Res Ther,* vol. 12, no. 1, p. 362, Jun 25 2021, doi: 10.1186/s13287-021-02426-2.

[358] S. Yamada, M. Hill, and T. Takakuwa, "Human Embryology," in *New Discoveries in Embryology*, 2015, ch. Chapter 5.

[359] D. Jiang, H. W. Jarrett, and W. E. Haskins, "Methods for proteomic analysis of transcription factors," (in eng), *J Chromatogr A,* vol. 1216, no. 41, pp. 6881-9, Oct 9 2009, doi: 10.1016/j.chroma.2009.08.044.

[360] D. Jiang, Y. Zhou, R. A. Moxley, and H. W. Jarrett, "Purification and identification of positive regulators binding to a novel element in the c-Jun promoter," (in eng), *Biochemistry,* vol. 47, no. 35, pp. 9318-34, Sep 2 2008, doi: 10.1021/bi800285q.

[361] B. Schwanhausser *et al.*, "Global quantification of mammalian gene expression control," (in eng), *Nature,* vol. 473, no. 7347, pp. 337-42, May 19 2011, doi: 10.1038/nature10098.

[362] G. Kustatscher, P. Grabowski, and J. Rappsilber, "Pervasive coexpression of spatially proximal genes is buffered at the protein level," *Molecular systems biology,* vol. 13, no. 8, p. 937, Aug 23 2017, doi: 10.15252/msb.20177548.

[363] S. Doll and A. L. Burlingame, "Mass spectrometry-based detection and assignment of protein posttranslational modifications," (in eng), *ACS Chem Biol,* vol. 10, no. 1, pp. 63-71, Jan 16 2015, doi: 10.1021/cb500904b.

[364] P. V. Hornbeck, B. Zhang, B. Murray, J. M. Kornhauser, V. Latham, and E. Skrzypek, "PhosphoSitePlus, 2014: mutations, PTMs and recalibrations," (in eng), *Nucleic Acids Res,* vol. 43, no. Database issue, pp. D512-20, Jan 2015, doi: 10.1093/nar/gku1267.

[365] E. J. Needham, B. L. Parker, T. Burykin, D. E. James, and S. J. Humphrey, "Illuminating the dark phosphoproteome," (in eng), *Science signaling,* vol. 12, no. 565, Jan 22 2019, doi: 10.1126/scisignal.aau8645.

[366] C. Schölz *et al.*, "Acetylation site specificities of lysine deacetylase inhibitors in human cells," (in eng), *Nat Biotechnol,* vol. 33, no. 4, pp. 415-23, Apr 2015, doi: 10.1038/nbt.3130.

[367] A. J. Hughes, D. P. Spelke, Z. Xu, C. C. Kang, D. V. Schaffer, and A. E. Herr, "Single-cell western blotting," *Nature methods,* vol. 11, no. 7, pp. 749-55, Jul 2014, doi: 10.1038/nmeth.2992.

[368] O. I. Ornatsky *et al.*, "Study of cell antigens and intracellular DNA by identification of element-containing labels and metallointercalators using inductively coupled plasma mass spectrometry," (in eng), *Analytical chemistry,* vol. 80, no. 7, pp. 2539-47, Apr 1 2008, doi: 10.1021/ac702128m.

[369] L. J. Tracey, Y. An, and M. J. Justice, "CyTOF: An Emerging Technology for Single-Cell Proteomics in the Mouse," *Curr Protoc,* vol. 1, no. 4, p. e118, Apr 2021, doi: 10.1002/cpz1.118.

[370] R. Gadalla *et al.*, "Validation of CyTOF Against Flow Cytometry for Immunological Studies and Monitoring of Human Cancer Clinical Trials," (in eng), *Front Oncol,* vol. 9, p. 415, 2019, doi: 10.3389/fonc.2019.00415.

[371] B. Budnik, E. Levy, G. Harmange, and N. Slavov, "SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation," (in eng), *Genome biology,* vol. 19, no. 1, p. 161, Oct 22 2018, doi: 10.1186/s13059-018-1547-5.

[372] A. A. Petelski *et al.*, "Multiplexed single-cell proteomics using SCoPE2," (in eng), *Nature protocols,* vol. 16, no. 12, pp. 5398-5425, Dec 2021, doi: 10.1038/s41596-021-00616-z.

[373] T. K. Cheung, C. Y. Lee, F. P. Bayer, A. McCoy, B. Kuster, and C. M. Rose, "Defining the carrier proteome limit for single-cell proteomics," (in eng), *Nature methods,* vol. 18, no. 1, pp. 76-83, Jan 2021, doi: 10.1038/s41592-020-01002-5.

[374] C. Ctortecka, K. Stejskal, G. Krššáková, S. Mendjan, and K. Mechtler, "Quantitative Accuracy and Precision in Multiplexed Single-Cell Proteomics," (in eng), *Analytical chemistry,* Dec 30 2021, doi: 10.1021/acs.analchem.1c04174.

[375] Y. Zhu *et al.*, "Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells," (in eng), *Nature communications,* vol. 9, no. 1, p. 882, Feb 28 2018, doi: 10.1038/s41467-018-03367-w.

[376] Y. Zhu *et al.*, "Proteomic analysis of single mammalian cells enabled by microfluidic nanodroplet sample preparation and ultrasensitive nanoLC-MS," (in eng), *Angewandte Chemie (International ed. in English),* May 24 2018, doi: 10.1002/anie.201802843.

[377] S. M. Williams *et al.*, "Automated Coupling of Nanodroplet Sample Preparation with Liquid Chromatography-Mass Spectrometry for High-Throughput Single-Cell Proteomics," (in eng), *Analytical chemistry,* vol. 92, no. 15, pp. 10588-10596, Aug 4 2020, doi: 10.1021/acs.analchem.0c01551.

[378] J. Woo *et al.*, "High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip," *Nature communications,* vol. 12, no. 1, p. 6246, Oct 29 2021, doi: 10.1038/s41467-021-26514-2.

[379] K. Stejskal, J. Op de Beeck, G. Dürnberger, P. Jacobs, and K. Mechtler, "Ultrasensitive NanoLC-MS of Subnanogram Protein Samples Using Second Generation Micropillar Array LC Technology with Orbitrap Exploris 480 and FAIMS PRO," (in eng), *Analytical chemistry,* vol. 93, no. 25, pp. 8704-8710, Jun 29 2021, doi: 10.1021/acs.analchem.1c00990.

[380] J. Woo *et al.*, "Robust, sensitive, and quantitative single-cell proteomics based on ion mobility filtering," *bioRxiv,* p. 2021.01.30.428333, 2021, doi: 10.1101/2021.01.30.428333.

[381] Y. Zhu *et al.*, "Subnanogram proteomics: impact of LC column selection, MS instrumentation and data analysis strategy on proteome coverage for trace samples," (in eng), *Int J Mass Spectrom,* vol. 427, pp. 4-10, Apr 2018, doi: 10.1016/j.ijms.2017.08.016.

[382] Y. Cong *et al.*, "Improved Single-Cell Proteome Coverage Using Narrow-Bore Packed NanoLC Columns and Ultrasensitive Mass Spectrometry," (in eng), *Analytical chemistry,* vol. 92, no. 3, pp. 2665-2671, Feb 4 2020, doi: 10.1021/acs.analchem.9b04631.

[383] H. Ouldali *et al.*, "Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore," *Nat Biotechnol,* vol. 38, no. 2, pp. 176-181, Feb 2020, doi: 10.1038/s41587-019-0345-2.

List of Abbreviations

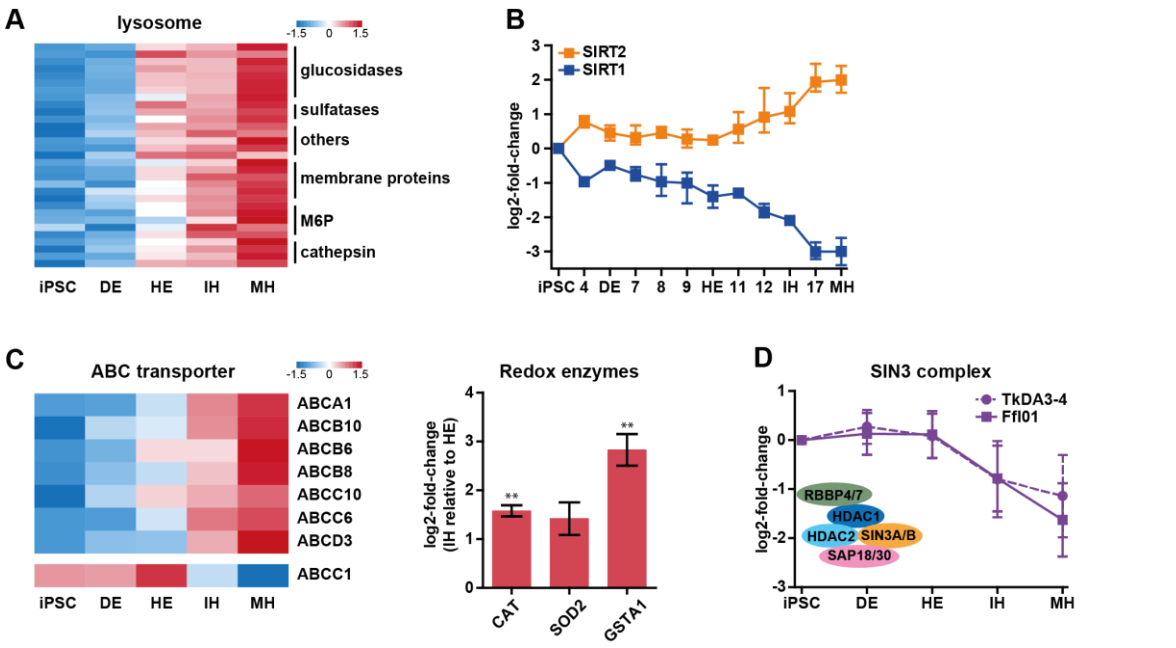
2D	Two-dimensional
3D	Three-dimensional
ACN	Acetonitrile
Ac-site	Acetylation site
AGC	Automatic gain control
ANOVA	Analysis of variance
BH	Benjamini-Hochberg
CID	Collision-induced dissociation
CV	Coefficient of variation
DDA	Data-dependent acquisition
DE	Definitive endoderm
EMT	Epithelial-mesenchymal transition
ESC	Embryonic stem cell
FA	Formic acid
FDR	false discovery rate
HCD	Higher-energy collisional dissociation
HDAC	Histone deacetylase
HE	Hepatic endoderm
HPLC	High-performance liquid chromatography
ICM	Inner cell mass
IH	Immature hepatocyte
IMAC	Immobilized metal ion affinity chromatography
iPSC	induced pluripotent stem cell
iST	in-StageTip
KDAC	Lysine deacetylase
KEGG	Kyoto Encyclopedia of Gene and Genomes
LB	Liver bud
LC	Liquid chromatography
m/z	Mass-to-charge ratio
maxIT	Maximum injection time
MH	Mature hepatocyte
MS	Mass spectrometry
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PHH	Primary human hepatocyte
PSC	Pluripotent stem cell
P-site	Phosphorylation site
PSM	Peptide spectrum match
PTM	Post-translational modification

R	Pearson´s correlation coefficient
R^2	Coefficient of determination
ROS	Reactive oxygen species
rpm	Revolutions per minute
RT	Room temperature
SDS	Sodium dodecyl sulphate
SP3	Single-pot, solid-phase-enhanced sample preparation
STAGETip	Stop-and-go-extraction tip
TFA	Trifluoroacetic acid
TMT	Tandem mass tag

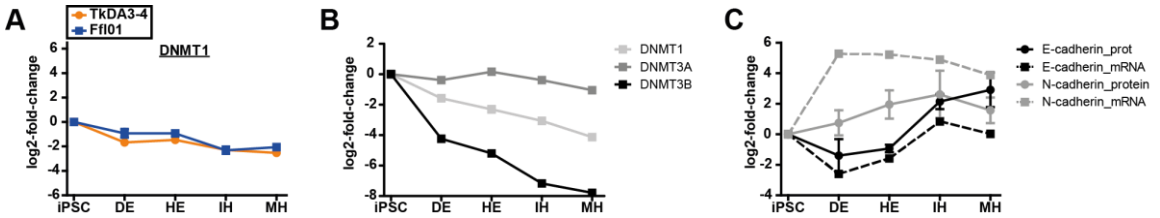
List of Figures

Appendix

1 Supplementary Figures



0-Figure 1: Temporal characterization of selected protein classes. (A) Z-scored expression of lysosomal proteins from the Ff-I01 differentiation. (B) High-temporal resolution expression of SIRT1 and SIRT2 as log2-fold-change relative to iPSC. X-axis showing days or specific differentiation stage (iPSC=d0, DE=d6, HE=d10, IH=d13, MH=d21). Data represent the average and error bars show range. (C) Left panel shows the z-scored expression of ABC transporter proteins. Right panel depicts average expression of ROS reducing enzymes. Data is from the Ff-I01 differentiation, error bars depict the average, and asterisk denote significance (ANOVA, BH corrected: *FDR<0.05; **FDR<0.01; ***FDR<0.001). (D) Dynamic expression of members from the SIN3 complex. Datapoints denoting the median of identified SIN3 complex members relative to iPSC. Error bars represent the standard deviation.



0-Figure 2: Expression profiles related to novel stage-specific marker proteins. (A) Dynamic protein expression relative to iPSC. Data represents the average and error bars the range. (B) Relative mRNA expression of all three DNA methyltransferases as derived from Camp *et al*.[100]. (C) Protein and mRNA expression of the epithelial marker E-cadherin and the mesenchymal marker N-cadherin, as quantified via MS or scRNA-seq (from Camp *et al*.[100]), respectively.

2 Supplementary Tables

0-Table 1: Primer list. List of forward and reverse primers with their respective Roche Universal ProbeLibrary number.

Gene name	Forward primer	Reverse primer	UPL #
OCT4	CTTCGCAAGCCCTCATTTC	GAGAAGGCGAAATCCGAAG	60
GATA4	GAAAACGGAAGCCCAAGAAC	CATCTCCTCGCTGCTGCT	11
SOX17	ACGCCGAGTTGAGCAAGA	TCTGCCTCCTCCACGAAG	61
HNF1B	CCTCTCACCTGATGGTAAAATGA	GGATATTCGTCAAGGTGCTGA	63
HNF4A	TCAGACCCTGAGCCACCT	AGCAACGGACAGATGTGTGA	27
FOXA2	TGTCTGAGGAGTCGGAGAGC	ACCGCTCCCAGCATACTTT	7
RBP4	CCAGAAGCGCAGAAGATTG	TTTCTTTCTGATCTGCCATCG	17
AFP	TCCTTGTAAGTGGCTTCTTGAAC	TGTACTGCAGAGATAAGTTTAGCTGAC	61
ALB	AATGTTGCCAAGCTGCTGA	CTTCCCTTCATCCCGAAGTT	27
KRT7	CAGGCTGAGATCGACAACATC	CTTGGCACGAGCATCCTT	24

0-Table 2: Cell surface marker proteins. List showing cell surface proteins that are differentially expressed (ANOVA, BH corrected FDR<0.05) and with a fold-change >2 at one or more stages. The 'CD' column shows the assigned cluster of differentiation name and the 'Stage' column depicts the respective developmental stage in which the marker protein is highly expressed.

Gene name	CD	Stage
ABCC4	no	DE
SIL1	no	DE
TMEM63A	no	DE
AGRN	no	HE
ASPH	no	HE
ATP1B1	no	HE
ATP6V0A1	no	HE
B4GALT5	no	HE
BST1	CD157	HE
CLU	no	HE
ERBB2	CD340	HE
FGFR1	CD331	HE
FGFR4	CD334	HE
FLRT2	no	HE
GGT7	no	HE
GOLM1	no	HE
HLA-DRB1	no	HE
IL6ST	CD130	HE
LAMA1	no	HE
LAMB1	no	HE
LAMC1	no	HE

LRP2	no	HE
PLD3	no	HE
PTPRA	no	HE
SCARB2	no	HE
SLC22A5	no	HE
SLC44A2	no	HE
ST3GAL1	no	HE
SULF2	no	HE
TMEM132A	no	HE
TMEM2	no	HE
TSPAN9	no	HE
TTYH3	no	HE
VCAN	no	HE
WNT11	no	HE
A2M	no	IH
ACAA1	no	IH
ANPEP	CD13	IH
APOB	no	IH
ASPH	no	IH
BST1	CD157	IH
C3	no	IH
CDH1	CD324	IH
CPD	no	IH
EGFR	no	IH
ENPP1	no	IH
FGA	no	IH
FGB	no	IH
FLRT2	no	IH
FN1	no	IH
HPX	no	IH
HS2ST1	no	IH
IGDCC4	no	IH
IGSF1	no	IH
IL1RAP	no	IH
ITGA3	CD49c	IH
LDLR	no	IH
LMF2	no	IH
LRP1	CD91	IH
LRP4	no	IH
MERTK	no	IH
PLOD1	no	IH
PON2	no	IH
PTPRF	no	IH
SCARB1	no	IH

SEMA4G	no	IH
SERPINA1	no	IH
SERPING1	no	IH
SERPINH1	no	IH
SIDT2	no	IH
SLC12A7	no	IH
SLC22A5	no	IH
SLC46A1	no	IH
SLIT3	no	IH
TFPI	no	IH
VTN	no	IH
A2M	no	MH
ANPEP	CD13	MH
C3	no	MH
CASP7	no	MH
CTSA	no	MH
CTSD	no	MH
DPP7	no	MH
ERO1LB	no	MH
GGT1	CD224	MH
HLA-C	no	MH
ICAM1	CD54	MH
ITGA2	CD49b	MH
LAMP2	CD107b	MH
LMNA	no	MH
NAGLU	no	MH
P2RX4	no	MH
SMPDL3B	no	MH
SQRDL	no	MH
ST14	no	MH
TFRC	CD71	MH
TMEM2	no	MH

Danksagungen

Geschafft! Nach etwas über 4 Jahren bin ich nun am Ende meiner Doktorarbeit und möchte mich bei einigen Menschen bedanken ohne die ich nicht so weit gekommen wäre.

Vielen Dank, Bernhard! Danke für deine Betreuung in den letzten Jahren. Für die vielen Meetings und die zahlreichen Tips, die du mir als Doktorvater mit auf den Weg gegeben hast! Danke, dass nicht nur deine Tür immer offen ist, sondern auch dein Ohr, und das trotz deines vollen Terminkalenders.

Ein weiterer Dank geht an Barbara. Durch dich ist das ganze Projekt erst entstanden und ich konnte einige interessante Einblicke in euer Labor werfen. Bei den Aufenthalten in Leipzig habe ich viel gelernt und wurde herzlich in die Arbeitsgruppe aufgenommen. Ein besonderer Dank geht an Agnieska für die Einblicke, die du mir in die Zelldifferenzierung gegeben hat.

Desweiteren möchte ich mich bei Markus Breunig bedanken. Mit dir habe ich nicht nur während der letzten Jahre erfolgreich wissenschaftlich zusammengearbeitet, sondern du bist auch einer meiner engsten Freunde und stehst mir immer mit Rat und Tat zur Seite. Danke auch für das Korrektur lesen meiner Arbeit.

Als nächstes möchte ich mich bei meinem Studenten Karl Kristian für die gute Zusammenarbeit während des Forschungspraktikums und der Masterarbeit bedanken. Du warst sehr fleißig und zuverlässig und es hat Spaß gemacht, mir dir zusammenzuarbeiten.

Ein dickes „Dankeschön“ geht an den kompletten Lehrstuhl für Proteomik und Bioanalytik aka TTT. Ich hätte mir in den vergangenen Jahren keinen schöneren Arbeitsplatz vorstellen können. Die tolle Arbeitsathmosphäre und das hohe Knowhow sucht wahrscheinlich seinesgleichen und ich werde noch oft an diese Zeit mit euch zurückdenken. Besonders möchte ich mich bei Andrea, Micha, Andi, Martina, Silvia und Gabi bedanken, ihr haltet den Lehrstuhl am Laufen und seid immer hilfsbereit zur Stelle. Ich möchte mich auch explizit bei Severin bedanken, mit dir habe ich nicht nur den ein oder anderen Kaffee getrunken, sondern währendessen viel über Wissenschaft aber auch Privates reden können. Danke auch für deinen Input für diese Doktorarbeit!

Ich möchte mich auch bei meinem Prüfungskomitee für die Bereitschaft bedanken, meine Doktorarbeit zu begutachten! Außerdem möchte ich mich auch bei meinen Kollaborationspartnern in den letzten Jahren für die Zusammenarbeit in vielen interessanten Projekten bedanken.

Mama, Papa, ich bin froh, dass ich euch habe! Vielen Dank für die Unterstützung und Freiheit, die ihr mir nicht nur während des Studiums oder der Doktorarbeit gegeben habt, sondern die ich schon mein ganzes Leben genießen darf. Eure positive Einstellung und euer Zuspruch, dass ich das schon alles schaffe und ich meinen Weg finde, haben die Weichen für diese Arbeit gestellt.

Liebe Kiki, was wäre ich nur ohne dich? Wie hätte ich das alles nur ohne dich schaffen können? Ich kann meinen Dank dir gegenüber nicht in Worte fassen und bin einfach nur glücklich, dass ich dich habe! Ohne deine alltägliche Unterstützung würde ich ganz schön alt aussehen!

List of publications

Published articles:

- Krumm J., Sekine K., Samaras P., Brazovskaja A., Breunig M., Yasui R., Kleger A., Taniguchi H., Wilhelm M., Treutlein B., Camp J. G., Kuster B. **"High temporal resolution proteome and phosphoproteome profiling of stem cell-derived hepatocyte development"**. Cell Reports 2022; 38(13):110604; doi: 10.1016/j.celrep.2022.110604

- Giansanti P, Samaras P, Bian Y, Meng C, Coluccio A, Frejno M, Jakubowsky H, Dobiasch S, Hazarika RR, Rechenberger J, Calzada-Wack J, Krumm J, Mueller S, Lee CY, Wimberger N, Lautenbacher L, Hassan Z, Chang YC, Falcomatà C, Bayer FP, Bärthel S, Schmidt T, Rad R, Combs SE, The M, Johannes F, Saur D, de Angelis MH, Wilhelm M, Schneider G, Kuster B. **"Mass spectrometry-based draft of the mouse proteome."**. Nature Methods 2022; 19(7):803-811; doi: 10.1038/s41592-022-01526-y

- Breunig M., Merkle J., Wagner M., Melzer M. K., Barth T. F. E., Engleitner T., Krumm J., Wiedenmann S., Cohrs C. M., Perkhofer L., Jain G., Krüger J., Hermann P. C., Schmid M., Madácsy T., Varga Á., Griger J., Azoitei N., Müller M., Wessely O., Robey P. G., Heller S., Dantes Z., Reichert M., Günes C., Bolenz C., Kuhn F., Maléth J., Speier S., Liebau S., Sipos B., Kuster B., Seufferlein T., Rad R., Meier M., Hohwieler M., and Kleger A. **"Modelling Plasticity and Dysplasia of Pancreatic Ductal Organoids Derived from Human Pluripotent Stem Cells"**. Cell Stem Cell 2021; 28(6):1105-1124; doi: 10.1016/j.stem.2021.03.005

- Philippi A., Heller S., Costa I., Senée V., Breunig M., Kwon G., Zhijian L., Illing A., Lin Q., Hohwieler M., Degavre A., Kassai B., Liebau S., Schuster M., Krumm J., Zhang N., Geusz R., Russell R., Besse C., Kuster B., Hebrok M., Seufferlein T., Boehm B. O., Oswald F., Sander M., Nicolino M., Julier C., and Kleger A. **"Mutations and variants of *ONECUT1* in diabetes"**. Nature Medicine 2021; 27:1928-1940; doi: 10.1038/s41591-021-01502-7