Quantitative mass spectrometry-based proteomics in the study of iPSC-derived neurons and disease modeling

A study in neurodevelopment and neurological disorders

Kwantitatieve massaspectrometrie-gebaseerde proteomics in de studie maar iPSC afkomstige neuronen en ziekte modellen

(met een samenvatting in het Nederlands)

Proefschrift

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

maandag 4 november 2019 des middags te 4.15 uur
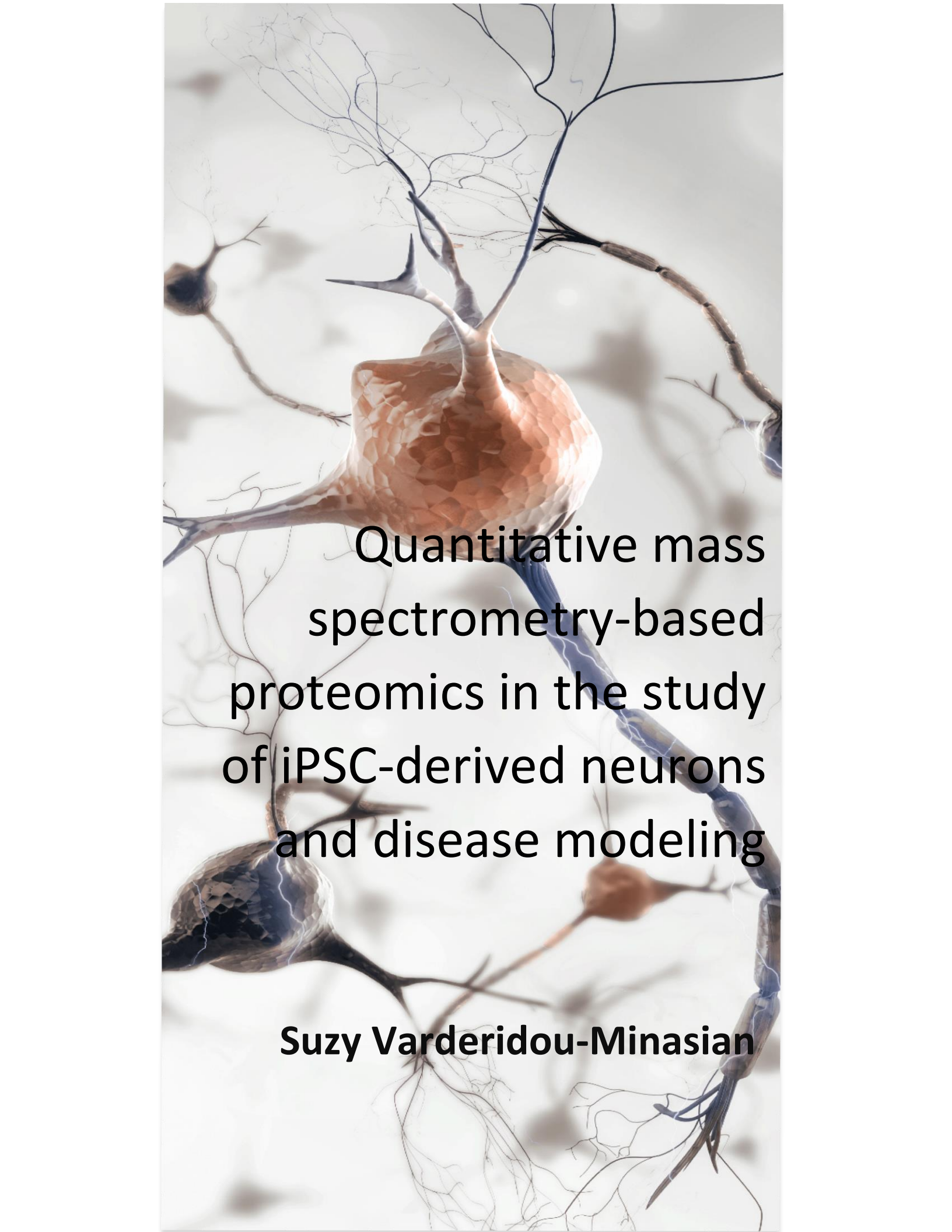
door

Souzanna Varderidou

geboren op 17 december 1986
te Kirovakan, Armenië

Promotoren:
Prof. dr. A.F.M. Altelaar
Prof. dr. A.J.R. Heck

Beoordelingscommissie:
Dr. V.M. Heine
Prof. dr. J.P.H. Burbach
Prof. dr. L.C. Kapitein
Prof. dr. A.D. Kraneveld
Prof. dr. C. Jimenez

Quantitative mass spectrometry-based proteomics in the study of iPSC-derived neurons and disease modeling

Suzy Varderidou-Minasian

Table of contents

Introduction

The human brain

What makes us human? Is it our brain that makes us special in any way? According to several studies humans have a very high brain-to-body weight ratio compared to other animals [1-3]. Humans also have more folds (gyri and sulci) in the outer layer of the brain (cortex). Therefore, humans have more neurons per unit volume than other animals indicating that it makes us more intelligent. Furthermore, unique areas in the brain (ventrolateral frontal cortex) that appear to be present only in humans are associated with higher-level functions such as cognition and language, basically processes that make us particularly humans [4]. In addition, specific types of neurons (Von Economo neurons) are found in humans and large mammals and not in rodents [5], and the list goes on [6]. The human central nervous system (CNS), which consist of the brain together with the spinal cord, exhibits developmental patterns and organizations typical for all mammals; starting from a neural tube that breaks off from the embryonic ectoderm and through several processes it acquires mature organizational features. The human CNS is among the earliest organ systems to begin its development (prenatal) and the last to complete (around agemid-20s) [7]. Gross anatomical features of the human brain comprise cerebral hemispheres, diencephalon, cerebellum and brainstem. The peripheral nervous system (PNS) consist of nerves (cranial and spinal) outside the CNS that connect the CNS to the rest of the body (skeletal muscles and organs). Micro anatomical features of the human brain include neurons, glial cells and neural stem cells. The human brain consists of 100 billion neurons and even more glial cells. Glial cells are quite different from neurons and have supportive functions in maintaining the signaling abilities of neurons. Neurons are specialized in intracellular communication. To do so, obvious morphological specialization arise of extensive branching (dendrites). A typical neuron consists of a soma, multiple dendrites and one axon. Dendrites are the targets for synaptic input from other neurons and axons are the output that transmit information. Key cellular features of CNS maturation are dendrite and axon outgrowth followed by the formation of synapses and myelination of axons.

For millennia, people study how the brain in humans and other animals develop and how they organize and function. About 60 million years ago, humans diverged from rodents and even longer from lower vertebrates. Therefore there are major differences in development and structure between humans and rodent brains [8]. Although animal models have provided fundamental insight into many human neurodevelopmental processes, disease mechanisms and the function of specific genes, it has become evident that they cannot fully model human neurodevelopment or disease, nor fully translate findings to the clinic [9]. Complex neurological disorders such as autism, schizophrenia and depression are difficult

to recapitulate in animal models, since multiple genetic events are involved. Human specific genes can provide novel insight into development and neurological disorders. Human post mortem tissue can provide information on cellular and molecular level, but deriving causal information may be challenging. Building a human CNS to study neurodevelopment and how dysregulations can lead to neurological disorders is the ultimate aim of every neuroscientist.

Generation of iPSCs

The first starting point to recapitulate human embryonic development giving rise to a variety of tissue cell types came in 1998 by Thomson et al [10] using human embryonic stem cell (ESC) lines. ESCs were isolated from the inner cell mass of the blastocyst during embryogenesis and are pluripotent meaning that they can differentiate into the three germ layers: -endoderm, -mesoderm and –ectoderm, and ultimately have the potential to generate any of the 300 cell types in the adult human body. This raised several issues concerning safety in cell-based therapies (such as teratoma formation) as well as discussions on ethical aspects. The discovery of human induced pluripotent stem cells (iPSCs) reprogrammed from a variety of human adult somatic cells in 2007 by Takahashi and Yamanaka et al, opened doors in the study of disease pathogenesis using patient material with genetic disorders [11]. The ectopic expression of only four genes (Oct4, Sox2, c-Myc and Klf4) were enough to reprogram adult cells into iPSCs. This solved most of the issues concerning human ESCs. iPSCs are free from the ethical concerns and they can be derived from patient material carrying the genetic information causing a disease. Especially in development and in psychiatric research, the usage of iPSCs has become increasingly important in advancing our understanding of the cellular mechanisms underlying disease and its development. Another leap forward came with the discovery of genome editing technologies (CRISPR-Cas9) allowing the manipulation (introducing or correcting) of disease-associated mutations to study human (neurological) disorders in vitro [12]. iPSC-based studies suffer from high variability due to differences in genetic background of different patients. Alternative, isogenic controls of patient iPSC lines can be generated reducing the variability linked to different genetic background. Together, these developments provided a significant contribution to our understanding of the developmental origin of several disorders.

Generation of neuronal cell types from human iPSCs

The brain consist of different neuronal subtypes with different neurotransmitter release, expression of several markers and different electrophysiological properties [13]. In many neurological disorders certain cell types and/or brain regions are affected. To be able to study the development of such neurons and

neurological disorders, requires generation of the neuronal cell types involved. In the last decade, several protocols have been developed to generate different neuronal subtypes. Many of these studies use the knowledge derived from animal studies to artificially mimic the signaling pathways that lead to generation of a certain neuronal type. During embryogenesis the development of neuroectoderm is regulated by morphogens along two axes; the rostro-dorsal (by WNT, FGF and retinoic acid (RA)) and dorso-ventral axis (by WNT, BMP and SHH) [14]. The most common method in neuronal differentiation is via monolayer differentiation by the usage of dual SMAD inhibition[15]. This significantly induces a rapid and complete neuronal conversion of iPSCs under adherent culture conditions. To further mimic the CNS development, which is formed from anterior-to-a-posterior fashion, the first neuroepithelial cells are of forebrain identity that mature into cortical glutamatergic neurons [16, 17]. The use of additional morphogens such as sonic hedgehog (SHH) and WNT inhibition is required to mimic the specification of other neuronal subtypes such as GABAergic neurons and cortical interneurons [18, 19]. Furthermore, by exposing the neural precursors to different combinations of WNT agonist with SHH or FGF, dopaminergic neurons can be generated from iPSCS [20-22]. So far only one study managed to generate hippocampal neurons [23]. Also, astrocytes, oligodendrocytes and microglia can be generated from iPSCs. In addition combinatorial actions of WNT, FGF, RA and SHH signals lead to generation of spinal-, cranial motor neurons, sensory neurons and spinal interneurons [24]. An alternative to the use of morphogens is the generation of induced neurons (iNs) where a defined set of transcription factors that are crucial for neuronal fate are induced to generate neurons. Successful protocols were developed to directly convert somatic cells [25-27] or iPSCs [28, 29] towards neurons. Their usage in the study of disease modeling however is less suitable since they bypass the process of neural progenitor specification which is often of importance in the disease pathogenesis. Also, limited number of iNs can be generated from somatic cells compared to iPSCs due to the restricted proliferative capacity of somatic cells. Another breakthrough in the study of brain development was the generation of three-dimensional (3D) tissues called organoids [30]. These brain organoids provide a tool that mimics human brain development and can be used to model neurodevelopmental disorders. They can grow up to 4 millimeters in size and be maintained for several months in a spinning bioreactor, which improves nutrient absorption. They have several characteristics resembling the human brain, such as fluid-filled cavities that resemble ventricles, large continuous brain lobes and they express different brain regional identities. A shortcoming of these organoids compared to the monolayers is the higher variability, both within and between organoids, consisting of different subtypes.

Human neurological disease modeling with iPSCs

The generation of iPSCs provide a valuable tool to study the cellular and molecular mechanisms associated with several neurological disorders. A variety of iPSCs derived from different patients have been used to study neurological disorders, drug discovery and cell therapy development. This has generated the concept of 'disease in a dish' using iPSC-derived relevant cell types in a culture dish [31]. Especially for rare genetic diseases the use of iPSCs has been greatly exploited, for example in using iPSCs from schizophrenia patients to show altered neuronal morphology, neuronal connectivity and gene expression in neurons [32]. Also several other diseases such as Alzheimer and Parkinson have been quite actively studied using such disease-specific iPSC-based research [33]. In **Table 1** a list of neurological diseases is shown that have been studied with iPSCs. In many of these cases, iPSC lines from non-disease-affected individuals were used as controls for the disease. However, this introduced high cell line to cell line variation which complicated the data. The introduction of genome editing technologies allowed the introduction or correction of a mutation leading to the generation of genetically matched isogenic control lines [34-36]. One of the challenges in this field is the fact that the generated neurons are relative immature [37], while cellular aging is of importance in the study several neurodegenerative disorders. Therefore, the use of iPSCs in the study of early onset of neurodevelopmental disorders might have more value. Rett syndrome (RTT) is a neurodevelopmental disorder mainly caused by mutations in the X-linked gene encoding MeCP2 protein. Especially females are affected and suffer from impaired cognitive, social-, motor skills and autistic behavior [38]. RTT patients show deceleration of head growth and on molecular level they have reduced dendrite branching as well as reduced growth of spines and synapses [39, 40]. Mutations in MeCP2 lead to altered regulation of target genes, including those influencing brain development and neuronal maturation [41-44]. Using iPSCs to study the molecular mechanisms involved in early stages of neuronal development might be of great value in the study of RTT syndrome. Another example of a disease that might benefit of this approach is called Spinal Muscular Atrophy (SMA). SMA is caused by the mutation in the survival motor neuron 1 gene (SMN1) that leads to the degeneration of motor neurons and subsequent muscular atrophy [45, 46]. Most humans have one or more copies of SMN2, which is a paralog of SMN1 that differs at a single nucleotide leading to altered splicing events [47, 48]. The SMN2 gene produces low levels of SMN protein and therefore the severity of SMA is inversely correlated with the number of SMN2 copies of the patient [49]. SMN1 is a ubiquitously expressed housekeeping gene with a primary role in the small nuclear ribonucleoprotein (snRNP) assembly process. These are RNA-protein complexes involved in the formation of the spliceosome to remove the introns from pre-mRNA, which is a critical step in post-transcriptional modification. It is unclear why only (lower)

6

motor neurons are affected by the loss of the SMN1 gene. The generation of human motor neurons from iPSCs stimulated the interest in understanding the molecular and cellular mechanisms that lead to MN degeneration.

Disease	Gene	Cell type differentiated from iPSCs	Drug test	References
Adrenoleukodystrophy	ABCD1	Oligodendrocytes and neurons	Lovastatin, 4-phenylbutyrate	[50]
Alzheimer's disease	PS1, PS2, APP, sporadic	Cortical neurons	ϒ, β-secretase inhibitor, docosahexaenoic acid	[50-52]
Amyotrophic lateral sclerosis	SOD1, VAPB, TDP43	Motor neurons and glial cells	Anacardid acid, trichostatin A, spliceostatin A, garcinol	[53-56]
Cockayne's syndrome	ERCC6	iPSCs and neurons	N/A	[57]
Down syndrome	Trisomy 21	Neurons and neural progenitors	N/A	[58-60]
Dravet syndrome	SCNA1A	GABAergic neurons	N/A	[61-64]
Familial dysautonomia	IKBKAP	Neural crest progenitor cells	Kinetin, SKF-86466	[65, 66]
Fragile X-associated tremor/ataxia syndrome	FMR1	Neurons	N/A	[67, 68]
Huntington's disease	HTT	Glutamatergic neurons and GABAergic neurons	N/A	[58, 69-74]
Machado-Joseph disease	ATXN3	Glutamatergic neurons	Calpain	[75]
Parkinson's disease	LRRK2, PINK1, SNCA, PARKIN, sporadic	Dopaminergic neurons	Coenzyme Q10, rapamycin, GW5074, LRRK2-IN1, PD0325901	[76-85]
Rett's syndrome	MeCP2, CDKL5	Neurons and neural progenitor cells	IGF1, gentamicin	[86-90]
Schizophrenia	DISC1, sporadic	Neurons	Loxapine, valproic acid	[91-94]
Spinal muscular atrophy	SMN1	Motor neurons	Valproic acid, tobramycin	[95, 96]
Spinal and bulbar muscular atrophy	CAG repeat in the androgen receptor gene	Motor neurons	17-allylaminogeldanamycin	[97]

Table 1. Neurological diseases modeled with iPSCs. Adopted from Yoichi Imaizumi *et al*, 2014 [33]

Introduction to Mass Spectrometry-Based Proteomics

Proteomics is the large-scale study of proteins in all cellular processes. The term proteomics was merged from "prot̲ein" and "gen̲omics", that was coined by Marc Wilkins in 1994 [98]. As illustrated in **Figure 1**, genomics describes the sequence information of the genes encoded by the DNA. Transcriptomics uses DNA microarray to describe the transcriptome, which is the complete set of RNA transcripts produces by the genome [99]. Proteomics, which is the third downstream "omics" field, describes the total protein complement of cells, including identification, localization, quantification and modification [100]. The proteins are often directly linked to biological function and pathological states, making them particularly relevant to study [101]. Studying proteins in a proteome-wide manner however remains challenging due to the complexity and the dynamic range of the proteome, with several orders of magnitude difference between the most and least abundant proteins in biological systems. Proteomics does not only involve the study of proteins themselves, but also covers post-translational modifications (PTMs) and alternative splicing variants, resulting in approximately one million possible protein forms [102]. PTMs such as phosphorylation, glycosylation or ubiquitination regulate the function of the protein by changing their physical and chemical characteristics that ultimately affects their localization, activation and degradation. On top of that, the abundance distribution of proteomes present in the cell is extremely large resulting in a broad dynamic range. Moreover, a rather large amount of sample material is needed since there are no amplification tools available for proteomics, as in genomics. Therefore, efficient and sensitive methods are required for the study of proteins. Mass spectrometry (MS) is the most comprehensive and versatile tool in large-scale proteomics studies. In the last decade, there has been a rapid development in MS technology (e.g. resolution, mass accuracy, speed and sensitivity) improving proteomic analysis. In this part of the thesis, we will describe several steps of typical proteomic experiments in more detail.
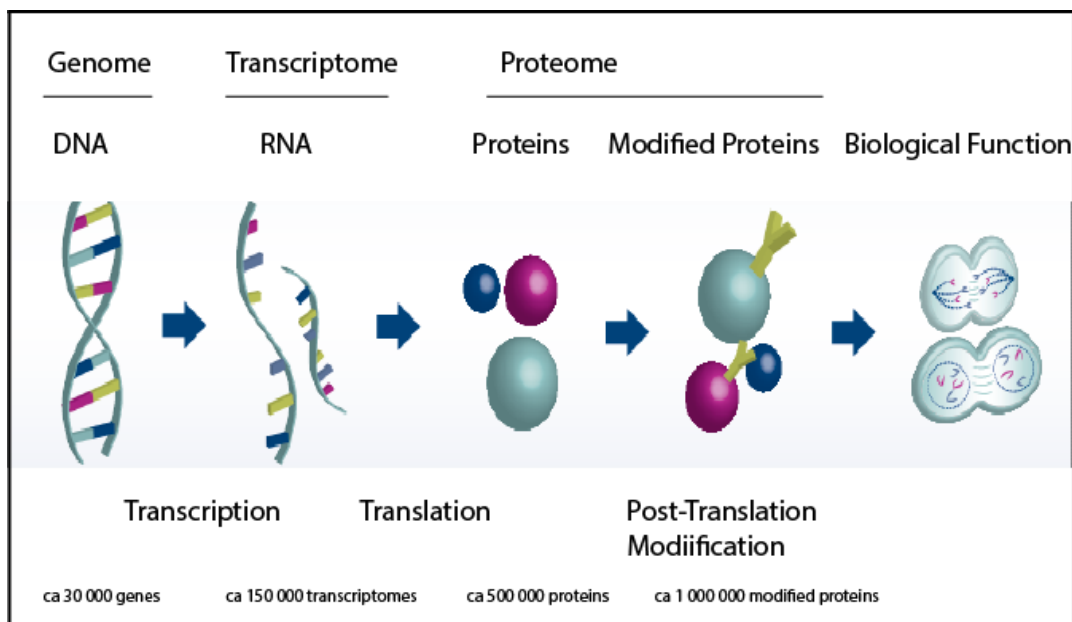
Figure 1. Adopted and slightly modified from pharmaceuticalintelligence.com

General Workflow in Mass Spectrometry-Based Proteomics

Three proteomic strategies can be distinguished, bottom-up, middle-down and top-down proteomics. In bottom-up protein analysis, also referred to as shotgun proteomics, proteins are digested into multiple peptides through proteolytic digestion and then subjected to liquid chromatography (LC)-MS/MS [103-105]. For protein identification, the tandem mass spectra derived through fragmentation are compared to theoretical tandem mass spectra from *in silico* digested proteins derived from curated databases. In top-down protein analysis, intact proteins are analyzed by the MS [106]. The advantage is that different PTMs and protein isoforms can be identified with notable success. The limitation, however, is the number of proteins that can be analyzed and the difficulties in protein ionization/fragmentation compared to peptides. In middle-down protein analysis, a hybrid of both bottom-up and top-down technologies are combined analyzing larger peptide fragments to gain insight into PTMs and minimizing the analytical challenges [107]. However, producing larger peptides and analyzing high number of proteins remains a challenge in this field. To date, shotgun proteomics is still the workhorse used by biologists to study cellular processes. A typical shotgun proteomic workflow is composed of a few steps, starting from sample preparation, separation or enrichment, LC-MS/MS analysis and bioinformatics tools (**Figure 2**).
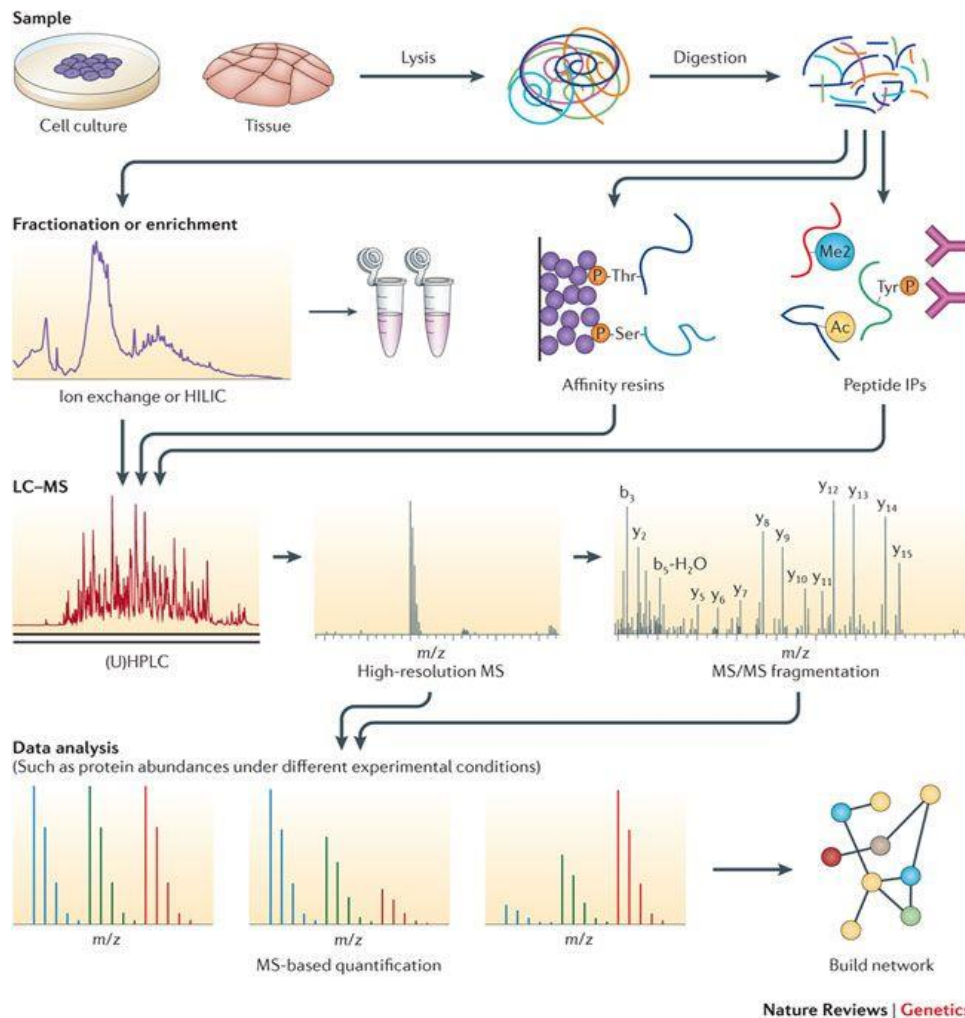
Figure 2. General workflow of shotgun proteomics experiment. Adopted from Altelaar *et al* 2013 [108].

Sample preparation

In a general proteomic experiment, proteins are first extracted from a cell, tissue or body fluid. Normally, this is done by lysing the cells in a certain lysis buffer and disrupting the cell membrane by sonication. In this way mainly cytosolic proteins are recovered, therefore additional methods are developed to extract proteins from specific locations, such as nucleus, cell membrane and even synaptosomes from neurons [109-111]. Proteins extracted from body fluid contain abundant proteins such as serum albumin that further needs to be depleted prior to LC-MS analysis [112]. This allows the identification of low abundant proteins. Since proteins can be modified and degraded by endogenous proteases, certain enzyme inhibitors are added in the lysis buffer [113]. Proteins are then digested into peptides to circumvent some of the MS challenges associated to intact proteins, such as separation, ionization and characterization. In general, Lys-C and Trypsin are the proteolytic enzymes used in standard shotgun proteomics for protein

digestion. Trypsin cleaves at the carboxyl side of arginine and lysine and produces peptides with favorable mass ranges for peptide separation and MS identification [114]. Recently, a number of LC- and MS-compatible surfactants have been developed to avoid additionally clean-up steps prior to MS [115].

Separation

The mixture of peptides generated after digestion change the physio-chemical properties then the proteins they are derived from, allowing the analysis of virtually all proteins, also those that were e.g. poorly soluble before. However, the MS can only capture a certain number of peptides in a certain amount of time. Therefore, the peptides are not introduced to the MS all at once. Instead, the peptides are pre-fractionated using their physio-chemical properties to decrease the complexity of the sample. Properties such as molecular mass, charge or pH gradients are generally employed to add another orthogonal mode prior to LC separation. Often, size exclusion chromatography (SEC), strong cation exchange (SCX), hidrophilic interation liquid chromatography (HILIC) or high-pH reversed-phase liquid chromatography (RPLC) are being used. SEC separates proteins based on their size in a LC-based manner [116]. Alternative to protein fractionations, peptide mixtures are separated in a LC-based manner in HILLIC, SCX and RPLC. HILIC separates peptides based on their hydrophilic interaction with an ionic resin [117]. SCX separates peptides based on their charge but was shown to have limitations to the sample loss due to clean-up steps required prior to MS. Alternative to SCX, RPLC was introduced using pH as a separation buffer and generating cleaner samples [118]. These separation methods prior to LC/MS are optional and are referred to as two dimensional separations. If the protein/peptide mixture is not very complex, then the mixture can be directly loaded into a RPLC (or (ultra)high-performance liquid chromatography ((U) HPLC) column, which is then directly coupled to the MS. Here, peptides are separated based on hydrophobicity, using gradients with increasing organic solvent content. Chromatographic separation is very delicate, therefore the inner diameter of the column, the length, loading particles, flow rate of the solvent are parameters that influence the separation and thus the identification of the proteins. Therefore, considerable expertise is required to operate this part. Later technological developments in the LC such as higher pressure and more sensitive chromatographic columns had great impact in shotgun proteomics [119-121]. The improvement in separation efficiency allows to identify similar protein numbers in a single separation dimension compared to two dimensional separations [122]. However, further improvements in the peptide/protein separation and reduction of sample losses are still needed to improve the identification of protein numbers.

Enrichment

As described before, PTMs increase the complexity of the proteome by covalently adding chemical units such as a phosphate group or glycans. The modification causes a mass shift that is deduced by the MS. Most studied PTMs in MS are phosphorylation, ubiquitination, glycosylation, acetylation and methylation. PTMs have important biological functions but their abundancy is relatively low which requires enrichment methods prior to MS analysis, which are mainly based on affinity chromatography. In phosphoproteomics, IMAC has become very popular to enrich phosphopeptides, but also TiO_2, Fe_3O_4 and ZrO_2 are commonly used [123-125]. They are based on the ionic interaction of the phosphate moiety with a chelating metal ion. Immobilized positively charged metal ions such as Fe^{3+}, Ga^{3+} and ZR^{4+} are used to selectively bind to phosphopeptides. Also, several chromatographic methods such as SCX have been used to separate phosphopeptides from non-phosphopeptides based on charge. Typical tryptic peptides carry a net charge of 2^+ at pH~3 and phophopeptides a net charge of 1^+ that is used in SCX to enrich for phosphopeptides [126]. In the glycoproteomics area, many studies have revealed that glycosylation has many important roles in biology, such as in folding and stability of proteins and in immune responses. Glycans are covalently linked to a protein, and can be enriched by lectin affinity chromatography, titanium dioxide chromatography, hydrazide chemistry based coupling or HILIC [127-130]. Lectin-affinity enrichment is the most used due to its high specificity to a wide range of glycan types. Although enrichment of a certain group of peptides makes the protein/peptide mixture less complex, fractionation afterwards is in some cases required.

Protein-protein interactions

Next to these enrichment methods for a specific PTM, enrichment approaches such as affinity purification is also applied in the study of protein-protein interaction (PPI) [131, 132]. Here, a specific protein of interest is bound to an antibody and immunoprecipitated to identify the interaction partners using MS [133]. This can be used to study the interactions between the proteins of interest. Samples can also be transfected with affinity tags that specifically targets the protein of interest (bait) and after enrichment PPIs are studied. There are different tags available such as GFP-, FLAF-or Strep-tag that can be used for enrichment [133-135]. One drawback of these enrichment methods is the co-enrichment of non-specific interactors. Therefore, multiple negative controls have to be taken into consideration, performing the experiment e.g. without using an antibody (empty beads), using an aspecific antibody, expressing an empty vector or after knock-down of the bait protein.

Ionization

One of the challenges in MS is to transfer the proteins or peptides that are polar and non-volatile into the gas phase. This challenge is largely addressed by two main techniques used in MS analysis to ionize proteins or peptides, matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) [136, 137], which earned the Nobel Prize for chemistry in 2002. In MALDI, the sample is mixed in dry, crystalline matrix [138]. After irradiation with a laser beam, the ionized analyte is transferred into the gas phase, producing predominantly singly charged ions. MALDI has a broad usage range, from relatively simple peptides to high-molecular weight proteins [139] (**Figure 3A**). ESI can be integrated in LC-MS systems, after the LC separation, molecules arrive at the end of the column and go through a needle with a very fine tip (~10uM), which is placed on high voltage (2-6 kV). The solvent containing the molecules of interest are released as highly charged droplets that are further vaporized as ions towards the entrance of the MS (**Figure 3B**). [140].
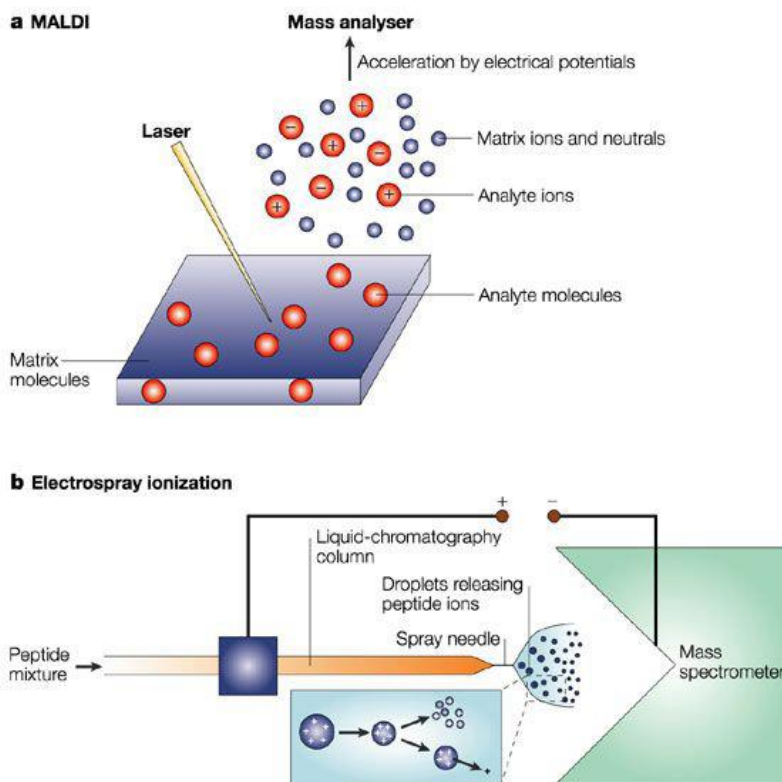

Figure 3. Schematic representation of ionization techniques frequently used in proteomic experiments. A) Matrix Assisted Laser Desorption Ionization (MALDI) and B) Electrospray Ionization (ESI) (Adopted from Hanno Steen and Matthias Mann 2004 [141].

Mass analyzers

Most mass spectrometers consist of the following parts: -an ion source and optics, - an analyzer and – detector. In the last decades advances in MS technology improved the identification rate of proteins. In particular, the development of hybrid MS instruments that combine different mass analyzers improved the accuracy and sensitivity of protein identification. It is the mass analyzer that isolates and the detector measures the peptide mass based on mass-to-charge ratios, allowing the MS to serve its goal. Mass analyzers are categorized into two main types: - the scanning and ion-beam MS analyzers (Time-of-flight, Quadrupole) and the –trapping MS analyzers (Ion trap and Orbitrap). These mass analyzers are very different in their performance and design which will be described below.

Time-of-flight (TOF) analyzers

The introduction of MALDI ionization drove TOF analyzers to the front for the identification of large molecules [142]. A simple TOF analyzer is composed of an acceleration grid and a flight tube (**Figure 4A**). Ionized molecules from the ion source are accelerated to the same potential and drift through the flight tube that operates at a constant electric field to the detector. The speed of each ion is inversely proportional to the square root of its m/z value, therefore ions with different mass will reach the detector at different times.

Quadrupole (Q) analyzer

A quadrupole mass analyzer consists of four (ideally) hyperbolic shaped metal rods placed in parallel. Different radio frequency (RF) voltages and direct currents (DC) are applied to the rods where two opposite rods are electrically connected [143] (**Figure 4B**). When ions travel through the assembly towards the detector, depending on the RF and DC that is applied, the trajectory of the ions is manipulated. Only ions of a certain m/z will stabilize and pass through while other m/z ions are unstable and will collide with the quadrupole rods. This enables the quadrupole to be used as a mass filter to isolate specific m/z ions. This incredible finding by Paul and Steinwedel was awarded with a Nobel Prize [144].

Linear Ion Trap (LIT) analyzer

Linear ion trap, also called 2D-trap, consists of four parallel metal rods. In LIT ions are confined radially by applying an RF field, and axially by applying static electrical potential on the end electrodes of the rods [145]. Ions with different m/z masses are 'trapped' or stored together and ejected one-by-one, scanning from low to high m/z, to the detector to get a mass spectrum. In the trap however, the ions repel each other due to space charge effects and destabilize their trajectories. This can be limited by collisional

cooling of the ions in the trap with inert gas such as Helium. All ions are focused towards the center reducing the kinetic energy and preventing collision between each other. Also the design of the LIT going from 2D-trap to 3D-trap, had a great benefit in reducing the charge to space effect [146]. The 3D-trap, also called Quadrupole ion trap, consist of a circular electrode (ring electrode) and two hyperbolic electrodes (end caps). An oscillating electric field is applied to the circular electrode while the hyperbolic electrodes are kept at ground potential, allowing to trap the ions (**Figure 4C**).

Orbitrap analyzer

The Orbitrap consist of an outer barrel-like electrode and a central spindle-like electrode (**Figure 4D**). Ions are orbitally trapped in a static quadrupole and a logarithmic electric field in which the ions orbit around a central electrode and oscillate in axial direction [147]. A Fast Fourier transform (FFT) algorithm is used to convert the time-domain signal into a mass-to-charge spectrum [148]. The Orbitrap mass analyzer is quite popular in the proteomics due to the high resolution (up to 1,000,000 FMHW at 200 mass-to-charge range) high mass accuracy (up to 1 p.p.m.) and dynamic range (> 10^3) [149]. Due to the combination of high accuracy and resolution at high speed it is possible to detect a wide range of small molecules.

Hybrid/Tribrid analyzer

In modern MS, combinations of multiple mass analyzers are coupled sequentially or in parallel into one mass spectrometer. These so-called 'hybrid' instruments often combine complementary mass spectrometers in one instrument to address the most challenging problems. Among some popular examples of hybrid instruments in proteomics are the triple quadrupole [150], Linear ion trap-Orbitrap (LTQ) [149, 151] and quadrupole-TOF [152]. The new generation (Orbitrap Fusion Lumos) contains tribrid architecture, such as three mass analyzers; a quadrupole, linear ion trap and Orbitrap. This enables the use of different peptide fragmentation methods (will be discussed later) and their combinations as well as the possibility to transfer different ions to the desired mass analyzer. This resulted in improvement in the analysis of PTMs [153], multiplexed relative quantification using isobaric tags (will be discussed later) [154], intact protein characterization [155] as well as the analysis of small molecules [156].
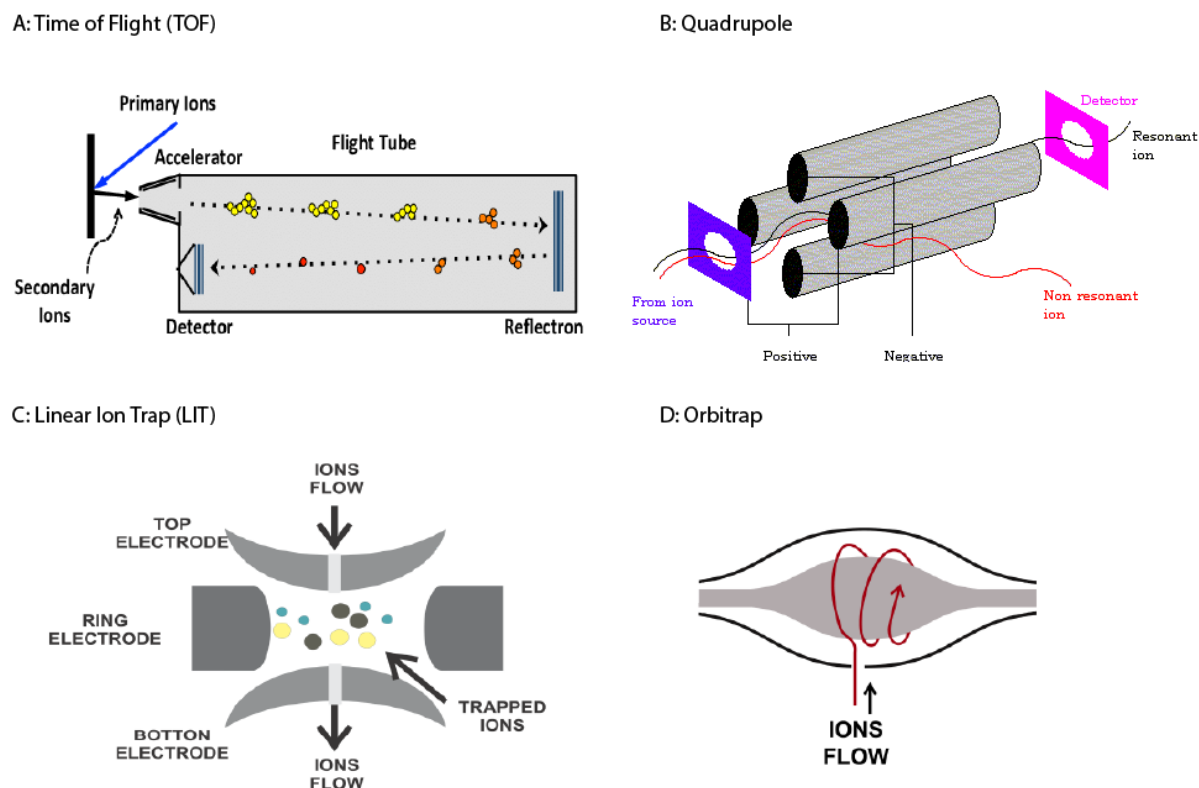
Figure 4. Schematic representation of mass spectrometry analyzers. A) Time of flight (TOF) analyzer B) Quadrupole C) Linear Ion trap (LIT) and D) Orbitrap (Adopted from D.R. Demartini 2013 [157]).

Peptide fragmentation

Ions from the ion source are in the first stage of the analysis selected based on their mass-to charge ratio and then fragmented. In the second stage of the analysis, the charged products are analyzed to get the sequence information about these peptides. This is called tandem MS or MS/MS (MS^2) and MS^3 for even more in-depth characterization, where fragments are additionally isolated and further fragmented. Basically, precursor ions (peptides) are broken down into fragments by applying energy. Depending on how the energy is transferred, the amount of energy and how it is internally distributed, different fragmentation patterns are formed. The most commonly used fragmentation techniques are the collision-induced dissociation (CID) and the electron-driven dissociation (ECD, ETD) techniques. In CID, precursor ions are subjected to collisions with an inert gas (e.g. nitrogen or helium) to transfer the energy to the backbone of the peptide. Upon collision, part of the kinetic energy is converted into vibrational energy leading to preferential cleavage at the weakest bond, producing mainly b- (n-terminal) and y- (c-terminal) ions (**Figure 5**). The fragmentation is performed in an ion trap or quadrupole instrument and is often used for peptides, lipids, small chemical compounds. Most bottom-up applications in MS use CID fragmentation techniques. The fragmentation of the precursor ion can be done in a quadrupole leading to higher

activation energy and shorter time. This is called higher energy collision dissociation (HCD) or beam-type CID which is specific only for LTQ-Orbitrap [158]. HCD generates b- and an predominant y-ions providing more informative ion series [159]. Complementary to CID is ETD where the protonated precursor peptide ion captures an electron, which is provided by a reagent molecule (e.g. fluoranthene or Arzulene) that generates a highly excited odd-electron peptide radical cation. This very unstable ion fragments, by itself or assisted by additional collisions (supplemental activation), mainly producing c- and z- ions via cleavage of the N-Cα backbone bonds. ETD is favorable for highly charged peptide fragments, charge states >3+ and those holding labile PTMs, or fragmentation of intact proteins in top-down proteomics.
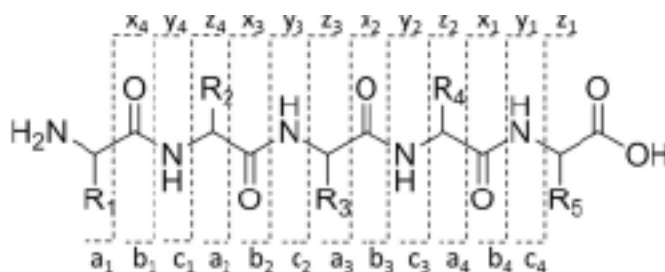

Figure 5. Peptide fragmentation. Nomenclature of a common peptide fragment ion (Adopted from Y. Zhang 2013 [105]).

Peptide sequencing

After a peptide is selected in MS1, it will be fragmented into combinations of amino acids, of which the spectra are referred to as tandem (MS/MS) fragmentation spectra. In principle, it is possible to determine the amino-acid sequence of that peptide by considering the mass difference of each amino acid. There are different approaches for the identification of peptides such as database searching, spectrum library or de novo sequencing. In shotgun proteomics, the most commonly used method is the database searching. Here, a hypothetical proteome (based on whole-genome sequencing) is created and *in silico* digested into hypothetical peptides. Each hypothetical peptide generates hypothetical MS/MS spectra. The actual measured mass of a peptide is matched to the list of the hypothetical peptides and if there is a match, the MS/MS spectra are then compared with the theoretical ones. To decide the best match, a score is assigned reflecting its similarity [160]. Furthermore, to make the software faster and more reliable, different restrictions to the search are taken into account, such as the taxonomy, the enzyme that was used, some fixed and variable peptide modifications, the mass tolerance and the mass spectrometer that was used. There are several database search algorithms developed to search sequence databases with the acquired MS/MS-spectra and the most commonly used are Sequest [161], Mascot [162] and Andromeda [163]. One thing to consider is the fact that when matching thousands of

experimental spectra to even more possible theoretical candidates, there is a chance of random incorrect hits. Therefore, several statistical methods have been developed to limit the number of false positive identifications. In principle, a second search is performed to a nonsense database that is often generated by reversed or scrambled protein sequences. Any match is considered as a false positive and a false discovery rate (FDR) is calculated. In shotgun proteomics, normally an FDR of 1% is acceptable. To increase the true positive matches, several filters or machine learning algorithms are applied known as Percolator [164]. A limitation to this database search strategy however is that this kind of experiments can only be carried out on organisms with sequenced genome. Also, mutations and unknown peptide sequences will not be identified using these database-dependent methods. Therefore, *de novo* sequencing, in which a peptide amino acid sequence is directly determined from the acquired tandem mass spectrometry, is performed. There are different software packages available such as PEAKS [165], Lutefisk [166], PepNono [167]. A drawback of this method however is that the success of *de novo* sequencing depends on the quality of the spectra which are influenced by many factors such as peptide ionization, fragmentation and the resolution of the MS.

Quantitative proteomics

Recent advances in the proteomics workflow in combination with the modern mass spectrometers has a great impact in the study of systems biology. Altered protein levels are directly correlated to disease states and biological processes. However, MS-based proteomics is not directly quantitative. The proteins between different conditions (healthy vs disease) may be changed by introducing variation during sample preparation. Fortunately, several techniques for quantitative measurements have been developed in quantitative proteomics. MS-based quantitative proteomics can be separated in absolute and relative quantification, the latter being obtained by isotope labeled or label-free strategies. In approaches where stable isotopes are introduced into the sample, the MS distinguishes the mass difference between different isotopes and the ratio between isotopes accurately indicates the abundance ratio of a protein between conditions. Two main methods are used to drive information for quantification: MS^1 or MS^2 quantification, depending on which MS scan is used to measure the stable isotope. Moreover, in relative quantification, the isotope labels can be introduced at various steps during the sample preparation (metabolically or chemically) (**Figure 6**). The earlier implementation of labels allows for less variation introduced during sample preparation or LC-Ms runs.
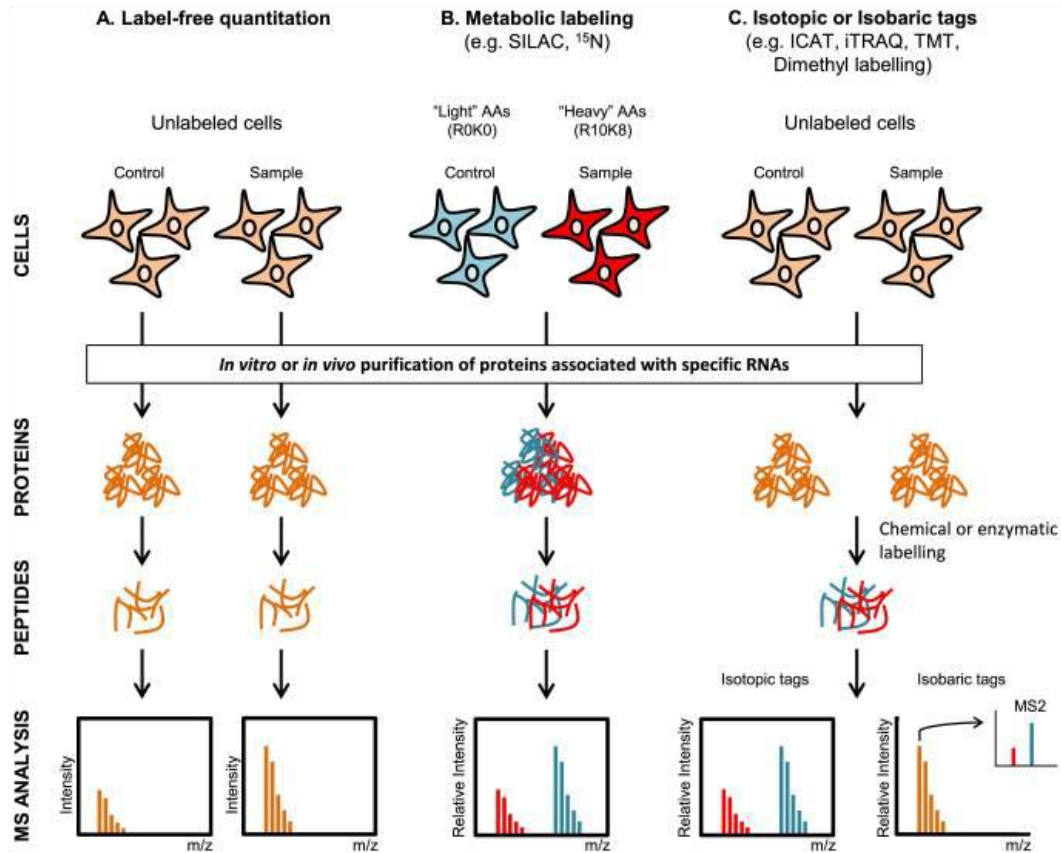
Figure 6. Quantitative mass spectrometry-based workflow. The stages when samples are labeled for quantitative proteomic analysis. Blue and red indicate the labels (in this case light and heavy). (Adopted from M. Jazurek *et al* 2016 [168]

Metabolic labeling

The earliest point to introduce a label is achieved by metabolically label cells, plants or even animals by growing them on a labeled medium or food source. The medium is replaced by specific amino acids that contains ^{15}N, D2 and/or ^{13}C instead of the naturally occurring isotopes ^{14}N, H1 and ^{12}C [169]. The labeled amino acids are incorporated into the proteins and peptides. A very popular metabolic labeling strategy is the stable isotope labeling by amino acids in cell culture (SILAC) [170]. Here, arginine and lysine are labeled with heavy isotopes because in bottom-up proteomics tryptic digest will then contain a label. Relative protein quantification is done using the MS^1 scan by comparing the intensities of the different isotopes. A drawback of this method is that certain cell lines or organisms are not capable to incorporate the label and multiplexing capabilities are limited.

Chemical labeling (MS[1]-based isotopic tags)

In contrast to metabolic labeling, chemical labeling can be applied to all samples because the label is introduced to the proteins or the peptides through a chemical reaction during the sample preparation. In the case of MS[1] quantification, the labelled tag results is a mass shift of the labeled peptide (protein) that can be distinguished in the MS[1] level. An isotope tag is added to the sample that reacts on the side chain of amino acids or to the peptide termini. Isotope-coded affinity tag (ICAT) was the first chemical labeling method in quantitative mass spectrometry using a reaction between cysteine residues and an isotope-coded tag [171]. The ICAT reagent consists of three key parts: - a biotin affinity tag, - a thiol specific reactive group, - and a linker containing either a light or a heavy isotope (^{1}H or ^{2}H atoms). One of the drawbacks however is that only peptides containing cysteine will be labeled eliminating all non-cysteine containing peptides. Another popular chemical approach using MS1 quantification is dimethyl labeling, in which all primary amines (the N-terminus and the amino group of lysine residues) are converted into dimethyl amines [172, 173]. Combination of different isotopic reagents of formaldehyde and cyanoborohydride (^{2}D and ^{13}C), three groups of light, medium and heavy labeled reagents can generate labeled peptides that differ in mass by 4 Da. This mass difference is used to relatively compare the ratios between three conditions. One of the advantages in using dimethyl labeling for quantitative approaches is the very low cost reagents that are used together with its easy and efficient labeling reaction. One of the drawbacks of dimethyl is the fact that deuterium labeling causes chromatographic shift, compromising the quantification. A less common type of MS[1]-based quantification is the proteolytic ^{18}O labeling, where the C-terminal carboxyl group is labeled with two atoms of ^{18}O by digestion in $H_2^{18}O$ [174, 175]. Also here, the labeled peptides differ in mass by 4 Da. In contrast to the previous mentioned isotope labeling methods, ^{18}O labeling has a smaller technical variation because the labeling and the digestion is done in one step. However, a drawback of this approach is the variable incorporation of the label into the peptides leading to variable labeling efficiencies, isotopic peak overlapping and complicated data analysis. Compared to metabolic labeling, chemical labeling in general introduces higher variability because the enzymatic digestion and labeling are performed separately. Furthermore, both for metabolic as well as for isotopic labeling approaches using MS1 quantification, combining different isotopically labeled peptides increases the sample complexity, limiting multiplexing capabilities.

Chemical labeling (MS[2]-based isobaric tags)

An elegant way to tackle the MS[1] sample complexity and to increase the multiplexing is the chemical labeling by isobaric tags. Isobaric tags for relative and absolute quantification (iTRAQ) [176] and tandem

mass tags (TMT) [177] are the most popular approaches in this category. Key advantages in this approach is that different isobaric mass tags have the same total mass but vary in terms of the distribution of heavy isotopes around their structure. Peptides that are differently labeled and represent different biological conditions (time points) have the same mass in MS1 spectra leading to less sample complexity in the MS1 scan. The isobaric tags consist of three key elements: - an amine reactive group, - an isotopic reporter group, - and an isotopic balancer group (**Figure 7A**). Normally, after digestion, peptides from different biological samples are labeled with the different forms of the tag. The reactive amine group of the tag adds the label to the N-terminal and lysine residues of the peptide and is lost after the reaction. The reporter group and the balancer group represent the label containing equal total number of ^{13}C and ^{15}N atoms. Upon fragmentation, the balancer group is lost and the reporter group generates different masses in the MS/MS spectra that is used for quantification. The intensity of the reporter group represents the relative abundance of the original peptide. Recently, TMT 10-plex was developed which is an expansion of the TMT 6-plex by combining the reagents from the TMT 6-plex with another four isotope variants of the tag that have a 6.32 mDa mass difference [178] (**Figure 7B**). Also TMT 11-plex was developed, which is identical to TMT 10-plex tags but is labeled with only ^{13}C isotopes [179] (**Figure 7C**). Isobaric quantification has benefits over other isotope quantification with respect to the multiplexing capabilities increasing the throughput within one experiment, thereby reducing the total LC-MS time and run-to-run variation. Also, when compared to other MS1 quantification strategies or label-free approaches, the MS2-based quantification benefits in reduced missing observations in the MS and increased sensitivity and accuracy [180]. A precursor ion from the MS^1 scan that is selected for fragmentation in the MS^2 scan, may not be selected consistently in another LC-MS/MS run. Therefore, samples that are run separately will result in missing values. In MS^2-based quantification, the same precursor ion is fragmented which will be present in all the labeled samples. One of the drawbacks however is that only high-resolution analyzers can resolve this low mass region. Most used MS instruments are Orbitrap based instruments, or Q-TOFs. Ion trap mass analyzers in CID fragmentation are excluded in this kind of quantification because the low mass reporters are ejected from the ion trap during activation [181]. Another common drawback in isobaric quantification is the so-called co-isolation issue. Here, when a peptide is selected for fragmentation, it can isolate near isobaric peptides as well within the isolation window. This leads to a mixed population of peptides that disturb the ratios of the original reporter ions. Normally this is skewed towards unity since the majority of the proteins don't change between samples [182]. Because the peak of each reporter ion will have some contribution from adjacent reporter ions, prior to data analysis, the values of teach reporter ion must be corrected. Some strategies were developed showing partially

compensation of the co-isolation interference by enhancing fractionation and LC separation, optimizing precursor ion isolation width, computational approaches post data acquisition and delayed fragmentation [183, 184]. Also, MS^3-based quantification showed beneficial outcomes in overcoming the co-isolation issue. Here, peptides are fragmented in ion-trap CID first and then the most abundant fragments are selected and fragmented in the HCD cell, followed by detection of the reporter ion intensities in the MS^3 scan. Another strategy to overcome this co-elution issue is by reducing the charge of the peptides by proton-transfer ion-ion reaction [185]. By reducing all charges by one, the doubly and triply charged peptides of the same m/z will have different m/z avoiding selection of peptides with similar m/z. A drawback of these strategies however is that it comes at the cost of speed and thus peptide identification.
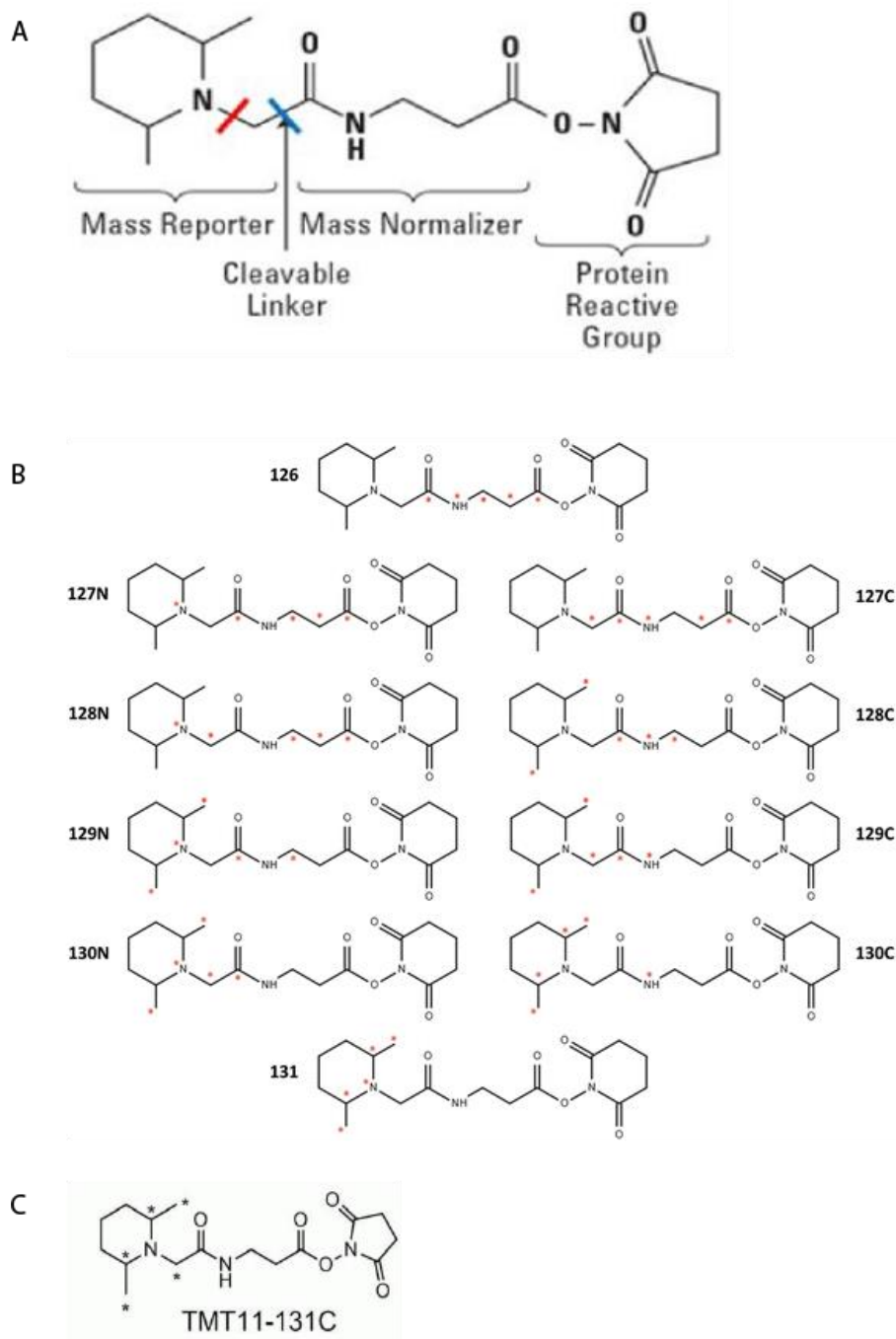
Figure 7. Chemical structure of TMT label reagent. A) General structure of the reagent consisting of a mass reporter group, a mass normalizer group and a reactive group. Arrow indicates the cleavage site, red indicates the cleavage in ETD fragmentation and blue in HCD fragmentation. B) Chemical structure of the TMT 10-plex reagent and isotope positions (* in red). C) Chemical structure of TMT 11-plex with ^{13}C isotope labels indicated (in *).

Label-free approaches

Label-free quantification is a popular alternative to labeling methods in quantitative proteomics. In general, each sample is analyzed separately with the advantage that there is no limit in the number of samples. Furthermore, because no labeling method is applied, less sample handling is needed. Quantification of protein expression is based on the comparison of identified peptides in separate LC-MS runs. There are two main approaches to achieve protein quantification which is based on: - spectral counting or – ion intensity. In spectral counting, the number of peptide spectral matches (PSMs) is correlated with the abundance of a protein [186, 187]. The number of identified peptides is divided by the number of theoretically observed peptides which is represented as protein abundance index (PAI) [188]. This method can be applied to low resolution MS and has simple normalization and statistical analysis. However, one of the drawbacks is that peptides with different physiochemical properties produce variability in quantification accuracy. In ion intensity based quantification instead, the extracted ion chromatogram is calculated for all peptides over the entire LC-MS run [189, 190]. Protein quantification is based on the area under the curve which represents the concentration of the peptides. However various factors can affect the accuracy and reproducibility of the method, such as co-eluting peptides, variations of retention time and chemical interference. Recent developments in MS further improve the quantification accuracy for both spectral counting and ion intensity-based label-free quantification. One of the drawbacks of label-free quantification in general is the reduced accuracy and precision caused due to the fact that sample preparation and LC-MS runs are carried out separately [191].

Absolute quantification

In contrast to the relative quantification methods described above, absolute quantification methods in proteomics is less common. Absolute quantification however provides more precise information of the protein abundance. A commonly used approach in absolute quantification is the use of reference peptides containing heavy isotope labels that are chemically identical to the original peptides and of which the exact (absolute) quantity is known. These labelled standards (e.g. AQUA peptides) are 'spiked in' and used for quantification [192]. This strategy is basically not based on labelling of the sample but introducing labelled standards to the sample. This is followed by determining the ratio between the labelled standards and the naturally occurring peptides. This quantification strategy was more precise compared to label-free quantification [193]. One of the drawbacks however is the limited number of synthetic peptides, the high price and the reliability of protein quantification since factors such as protein degradation and incomplete digestion affect the accuracy of the quantification [105].

References

1.	*Evolution of the Brain and Intelligence.* Current Anthropology, 1975. **16**(3): p. 403-426.
2.	Jerison, H.J., *Brain to body ratios and the evolution of intelligence.* Science, 1955. **121**(3144): p. 447-9.
3.	Herculano-Houzel, S., B. Mota, and R. Lent, *Cellular scaling rules for rodent brains.* Proc Natl Acad Sci U S A, 2006. **103**(32): p. 12138-43.
4.	Neubert, F.-X., et al., *Comparison of Human Ventral Frontal Cortex Areas for Cognitive Control and Language with Areas in Monkey Frontal Cortex.* Neuron, 2014. **81**(3): p. 700-713.
5.	Allman, J.M., et al., *The von Economo neurons in apes and humans.* Am J Hum Biol, 2011. **23**(1): p. 5-21.
6.	Herculano-Houzel, S., *The human brain in numbers: a linearly scaled-up primate brain.* Frontiers in human neuroscience, 2009. **3**: p. 31-31.
7.	Gogtay, N., et al., *Dynamic mapping of human cortical development during childhood through early adulthood.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(21): p. 8174.
8.	Clowry, G., Z. Molnar, and P. Rakic, *Renewed focus on the developing human neocortex.* J Anat, 2010. **217**(4): p. 276-88.
9.	Bae, B.-I., D. Jayaraman, and Christopher A. Walsh, *Genetic Changes Shaping the Human Brain.* Developmental Cell, 2015. **32**(4): p. 423-434.
10.	Thomson, J.A., et al., *Embryonic Stem Cell Lines Derived from Human Blastocysts.* Science, 1998. **282**(5391): p. 1145.
11.	Takahashi, K., et al., *Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors.* Cell, 2007. **131**(5): p. 861-872.
12.	Doudna, J.A. and E. Charpentier, *The new frontier of genome engineering with CRISPR-Cas9.* Science, 2014. **346**(6213): p. 1258096.
13.	Nelson, S.B., K. Sugino, and C.M. Hempel, *The problem of neuronal cell types: a physiological genomics approach.* Trends Neurosci, 2006. **29**(6): p. 339-45.
14.	Liu, H. and S.C. Zhang, *Specification of neuronal and glial subtypes from human pluripotent stem cells.* Cell Mol Life Sci, 2011. **68**(24): p. 3995-4008.
15.	Chambers, S.M., et al., *Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling.* Nature biotechnology, 2009. **27**(3): p. 275-280.
16.	Boissart, C., et al., *Differentiation from human pluripotent stem cells of cortical neurons of the superficial layers amenable to psychiatric disease modeling and high-throughput drug screening.* Translational psychiatry, 2013. **3**(8): p. e294-e294.
17.	Espuny-Camacho, I., et al., *Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits in vivo.* Neuron, 2013. **77**(3): p. 440-56.
18.	Liu, Y., et al., *Directed differentiation of forebrain GABA interneurons from human pluripotent stem cells.* Nature protocols, 2013. **8**(9): p. 1670-1679.
19.	Nicholas, Cory R., et al., *Functional Maturation of hPSC-Derived Forebrain Interneurons Requires an Extended Timeline and Mimics Human Neural Development.* Cell Stem Cell, 2013. **12**(5): p. 573-586.
20.	Denham, M., et al., *Glycogen synthase kinase 3beta and activin/nodal inhibition in human embryonic stem cells induces a pre-neuroepithelial state that is required for specification to a floor plate cell lineage.* Stem Cells, 2012. **30**(11): p. 2400-11.
21.	Kirkeby, A., et al., *Generation of Regionally Specified Neural Progenitors and Functional Neurons from Human Embryonic Stem Cells under Defined Conditions.* Cell Reports, 2012. **1**(6): p. 703-714.

22. Xi, J., et al., *Specification of midbrain dopamine neurons from primate pluripotent stem cells.* Stem Cells, 2012. **30**(8): p. 1655-63.

23. Yu, Diana X., et al., *Modeling Hippocampal Neurogenesis Using Human Pluripotent Stem Cells.* Stem Cell Reports, 2014. **2**(3): p. 295-310.

24. Maury, Y., et al., *Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes.* Nat Biotechnol, 2015. **33**(1): p. 89-96.

25. Ambasudhan, R., et al., *Direct Reprogramming of Adult Human Fibroblasts to Functional Neurons under Defined Conditions.* Cell Stem Cell, 2011. **9**(2): p. 113-118.

26. Hu, W., et al., *Direct Conversion of Normal and Alzheimer's Disease Human Fibroblasts into Neuronal Cells by Small Molecules.* Cell Stem Cell, 2015. **17**(2): p. 204-212.

27. Ladewig, J., et al., *Small molecules enable highly efficient neuronal conversion of human fibroblasts.* Nat Methods, 2012. **9**(6): p. 575-8.

28. Pang, Z.P., et al., *Induction of human neuronal cells by defined transcription factors.* Nature, 2011. **476**(7359): p. 220-3.

29. Zhang, Y., et al., *Rapid single-step induction of functional neurons from human pluripotent stem cells.* Neuron, 2013. **78**(5): p. 785-98.

30. Lancaster, M.A. and J.A. Knoblich, *Generation of cerebral organoids from human pluripotent stem cells.* Nature protocols, 2014. **9**(10): p. 2329-2340.

31. Shi, Y., et al., *Induced pluripotent stem cell technology: a decade of progress.* Nat Rev Drug Discov, 2017. **16**(2): p. 115-130.

32. Brennand, K.J., et al., *Modelling schizophrenia using human induced pluripotent stem cells.* Nature, 2011. **473**(7346): p. 221-225.

33. Imaizumi, Y. and H. Okano, *Modeling human neurological disorders with induced pluripotent stem cells.* J Neurochem, 2014. **129**(3): p. 388-99.

34. Hockemeyer, D., et al., *Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases.* Nat Biotechnol, 2009. **27**(9): p. 851-7.

35. Christian, M., et al., *Targeting DNA double-strand breaks with TAL effector nucleases.* Genetics, 2010. **186**(2): p. 757-61.

36. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

37. Studer, L., E. Vera, and D. Cornacchia, *Programming and Reprogramming Cellular Age in the Era of Induced Pluripotency.* Cell Stem Cell, 2015. **16**(6): p. 591-600.

38. Amir, R.E., et al., *Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2.* Nat Genet, 1999. **23**(2): p. 185-8.

39. Qiu, Z., et al., *The Rett syndrome protein MeCP2 regulates synaptic scaling.* J Neurosci, 2012. **32**(3): p. 989-94.

40. Smrt, R.D., et al., *Mecp2 deficiency leads to delayed maturation and altered gene expression in hippocampal neurons.* Neurobiol Dis, 2007. **27**(1): p. 77-89.

41. Lewis, J.D., et al., *Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA.* Cell, 1992. **69**(6): p. 905-14.

42. Nan, X., et al., *Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex.* Nature, 1998. **393**(6683): p. 386-9.

43. Chahrour, M., et al., *MeCP2, a key contributor to neurological disease, activates and represses transcription.* Science, 2008. **320**(5880): p. 1224-9.

44. Gabel, H.W., et al., *Disruption of DNA-methylation-dependent long gene repression in Rett syndrome.* Nature, 2015. **522**(7554): p. 89-93.

45. Lefebvre, S., et al., *Identification and characterization of a spinal muscular atrophy-determining gene.* Cell, 1995. **80**(1): p. 155-65.

46.   Lefebvre, S., et al., *Correlation between severity and SMN protein level in spinal muscular atrophy.* Nat Genet, 1997. **16**(3): p. 265-9.

47.   Lorson, C.L., et al., *A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy.* Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6307-11.

48.   Monani, U.R., et al., *A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2.* Hum Mol Genet, 1999. **8**(7): p. 1177-83.

49.   Harada, Y., et al., *Correlation between SMN2 copy number and clinical phenotype of spinal muscular atrophy: three SMN2 copies fail to rescue some patients from the disease severity.* J Neurol, 2002. **249**(9): p. 1211-9.

50.   Jang, J., et al., *Induced pluripotent stem cell models from X-linked adrenoleukodystrophy patients.* Ann Neurol, 2011. **70**(3): p. 402-9.

51.   Israel, M.A., et al., *Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells.* Nature, 2012. **482**(7384): p. 216-20.

52.   Kondo, T., et al., *Modeling Alzheimer's disease with iPSCs reveals stress phenotypes associated with intracellular Abeta and differential drug responsiveness.* Cell Stem Cell, 2013. **12**(4): p. 487-96.

53.   Dimos, J.T., et al., *Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons.* Science, 2008. **321**(5893): p. 1218-21.

54.   Mitne-Neto, M., et al., *Downregulation of VAPB expression in motor neurons derived from induced pluripotent stem cells of ALS8 patients.* Hum Mol Genet, 2011. **20**(18): p. 3642-52.

55.   Bilican, B., et al., *Mutant induced pluripotent stem cell lines recapitulate aspects of TDP-43 proteinopathies and reveal cell-specific vulnerability.* Proc Natl Acad Sci U S A, 2012. **109**(15): p. 5803-8.

56.   Egawa, N., et al., *Drug screening for ALS using patient-specific induced pluripotent stem cells.* Sci Transl Med, 2012. **4**(145): p. 145ra104.

57.   Andrade, L.N.d.S., et al., *Evidence for premature aging due to oxidative stress in iPSCs from Cockayne syndrome.* Human Molecular Genetics, 2012. **21**(17): p. 3825-3834.

58.   Park, I.H., et al., *Disease-specific induced pluripotent stem cells.* Cell, 2008. **134**(5): p. 877-86.

59.   Li, L.B., et al., *Trisomy correction in Down syndrome induced pluripotent stem cells.* Cell Stem Cell, 2012. **11**(5): p. 615-9.

60.   Weick, J.P., et al., *Deficits in human trisomy 21 iPSCs and neurons.* Proc Natl Acad Sci U S A, 2013. **110**(24): p. 9962-7.

61.   Higurashi, N., et al., *A human Dravet syndrome model from patient induced pluripotent stem cells.* Mol Brain, 2013. **6**: p. 19.

62.   Jiao, J., et al., *Modeling Dravet syndrome using induced pluripotent stem cells (iPSCs) and directly converted neurons.* Hum Mol Genet, 2013. **22**(21): p. 4241-52.

63.   Liu, Y., et al., *Directed differentiation of forebrain GABA interneurons from human pluripotent stem cells.* Nat Protoc, 2013. **8**(9): p. 1670-9.

64.   Liu, Y., et al., *Dravet syndrome patient-derived neurons suggest a novel epilepsy mechanism.* Ann Neurol, 2013. **74**(1): p. 128-39.

65.   Lee, G., et al., *Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs.* Nature, 2009. **461**(7262): p. 402-6.

66.   Lee, G., et al., *Large-scale screening using familial dysautonomia induced pluripotent stem cells identifies compounds that rescue IKBKAP expression.* Nat Biotechnol, 2012. **30**(12): p. 1244-8.

67.   Urbach, A., et al., *Differential modeling of fragile X syndrome by human embryonic stem cells and induced pluripotent stem cells.* Cell Stem Cell, 2010. **6**(5): p. 407-11.

68.   Liu, L., et al., *31P and 1h NMR as probes of domain alignment in a rigid crystalline surfactant mesophase.* Langmuir, 2005. **21**(9): p. 3795-801.

69. Zhang, N., et al., *Characterization of Human Huntington's Disease Cell Model from Induced Pluripotent Stem Cells.* PLoS Curr, 2010. **2**: p. Rrn1193.
70. An, Mahru C., et al., *Genetic Correction of Huntington's Disease Phenotypes in Induced Pluripotent Stem Cells.* Cell Stem Cell, 2012. **11**(2): p. 253-263.
71. Camnasio, S., et al., *The first reported generation of several induced pluripotent stem cell lines from homozygous and heterozygous Huntington's disease patients demonstrates mutation related enhanced lysosomal activity.* Neurobiology of Disease, 2012. **46**(1): p. 41-51.
72. The Hd iPsc Consortium, *Induced Pluripotent Stem Cells from Patients with Huntington's Disease Show CAG-Repeat-Expansion-Associated Phenotypes.* Cell Stem Cell, 2012. **11**(2): p. 264-278.
73. Jeon, I., et al., *Neuronal properties, in vivo effects, and pathology of a Huntington's disease patient-derived induced pluripotent stem cells.* Stem Cells, 2012. **30**(9): p. 2054-62.
74. Juopperi, T.A., et al., *Astrocytes generated from patient induced pluripotent stem cells recapitulate features of Huntington's disease patient cells.* Molecular Brain, 2012. **5**(1): p. 17.
75. Koch, P., et al., *Excitation-induced ataxin-3 aggregation in neurons from patients with Machado–Joseph disease.* Nature, 2011. **480**: p. 543.
76. Devine, M.J., et al., *Parkinson's disease induced pluripotent stem cells with triplication of the α-synuclein locus.* Nature Communications, 2011. **2**: p. 440.
77. Nguyen, H.N., et al., *LRRK2 mutant iPSC-derived DA neurons demonstrate increased susceptibility to oxidative stress.* Cell Stem Cell, 2011. **8**(3): p. 267-80.
78. Seibler, P., et al., *Mitochondrial Parkin recruitment is impaired in neurons derived from mutant PINK1 induced pluripotent stem cells.* J Neurosci, 2011. **31**(16): p. 5970-6.
79. Cooper, O., et al., *Pharmacological Rescue of Mitochondrial Deficits in iPSC-Derived Neural Cells from Patients with Familial Parkinson's Disease.* Science Translational Medicine, 2012. **4**(141): p. 141ra90.
80. Imaizumi, Y., et al., *Mitochondrial dysfunction associated with increased oxidative stress and alpha-synuclein accumulation in PARK2 iPSC-derived neurons and postmortem brain tissue.* Mol Brain, 2012. **5**: p. 35.
81. Jiang, H., et al., *Parkin controls dopamine utilization in human midbrain dopaminergic neurons derived from induced pluripotent stem cells.* Nat Commun, 2012. **3**: p. 668.
82. Liu, G.H., et al., *Progressive degeneration of human neural stem cells caused by pathogenic LRRK2.* Nature, 2012. **491**(7425): p. 603-7.
83. Rakovic, A., et al., *PTEN-induced putative kinase 1 (PINK1)-dependent ubiquitination of endogenous Parkin attenuates mitophagy: study in human primary fibroblasts and induced pluripotent stem (iPS) cell-derived neurons.* Journal of Biological Chemistry, 2012.
84. Sanchez-Danes, A., et al., *Disease-specific phenotypes in dopamine neurons from human iPS-based models of genetic and sporadic Parkinson's disease.* EMBO Mol Med, 2012. **4**(5): p. 380-95.
85. Reinhardt, P., et al., *Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression.* Cell Stem Cell, 2013. **12**(3): p. 354-67.
86. Marchetto, M.C., et al., *A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells.* Cell, 2010. **143**(4): p. 527-39.
87. Muotri, A.R., et al., *L1 retrotransposition in neurons is modulated by MeCP2.* Nature, 2010. **468**(7322): p. 443-6.
88. Ananiev, G., et al., *Isogenic pairs of wild type and mutant induced pluripotent stem cell (iPSC) lines from Rett syndrome patients as in vitro disease model.* PLoS One, 2011. **6**(9): p. e25255.
89. Cheung, A.Y., et al., *Isolation of MECP2-null Rett Syndrome patient hiPS cells and isogenic controls through X-chromosome inactivation.* Hum Mol Genet, 2011. **20**(11): p. 2103-15.

90.   Ricciardi, S., et al., *CDKL5 ensures excitatory synapse stability by reinforcing NGL-1-PSD95 interaction in the postsynaptic compartment and is impaired in patient iPSC-derived neurons.* Nat Cell Biol, 2012. **14**(9): p. 911-23.

91.   Brennand, K.J., et al., *Modelling schizophrenia using human induced pluripotent stem cells.* Nature, 2011. **473**(7346): p. 221-5.

92.   Chiang, C.H., et al., *Integration-free induced pluripotent stem cells derived from schizophrenia patients with a DISC1 mutation.* Mol Psychiatry, 2011. **16**(4): p. 358-60.

93.   Pedrosa, E., et al., *Development of patient-specific neurons in schizophrenia using induced pluripotent stem cells.* J Neurogenet, 2011. **25**(3): p. 88-103.

94.   Paulsen Bda, S., et al., *Altered oxygen metabolism associated to neurogenesis of induced pluripotent stem cells derived from a schizophrenic patient.* Cell Transplant, 2012. **21**(7): p. 1547-59.

95.   Ebert, A.D., et al., *Induced pluripotent stem cells from a spinal muscular atrophy patient.* Nature, 2009. **457**(7227): p. 277-80.

96.   Chang, T., et al., *Brief report: phenotypic rescue of induced pluripotent stem cell-derived motoneurons of a spinal muscular atrophy patient.* Stem Cells, 2011. **29**(12): p. 2090-3.

97.   Nihei, Y., et al., *Enhanced Aggregation of Androgen Receptor in Induced Pluripotent Stem Cell-derived Neurons from Spinal and Bulbar Muscular Atrophy.* Journal of Biological Chemistry, 2013. **288**(12): p. 8043-8052.

98.   Wasinger, V.C., et al., *Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium.* Electrophoresis, 1995. **16**(7): p. 1090-4.

99.   Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray.* Science, 1995. **270**(5235): p. 467-70.

100.   Yarmush, M.L. and A. Jayaraman, *Advances in proteomic technologies.* Annu Rev Biomed Eng, 2002. **4**: p. 349-73.

101.   Mann, M. and O.N. Jensen, *Proteomic analysis of post-translational modifications.* Nat Biotechnol, 2003. **21**(3): p. 255-61.

102.   Gstaiger, M. and R. Aebersold, *Applying mass spectrometry-based proteomics to genetics, genomics and network biology.* Nature Reviews Genetics, 2009. **10**: p. 617.

103.   Wolters, D.A., M.P. Washburn, and J.R. Yates, 3rd, *An automated multidimensional protein identification technology for shotgun proteomics.* Anal Chem, 2001. **73**(23): p. 5683-90.

104.   Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.

105.   Zhang, Y., et al., *Protein analysis by shotgun/bottom-up proteomics.* Chem Rev, 2013. **113**(4): p. 2343-94.

106.   Toby, T.K., L. Fornelli, and N.L. Kelleher, *Progress in Top-Down Proteomics and the Analysis of Proteoforms.* Annu Rev Anal Chem (Palo Alto Calif), 2016. **9**(1): p. 499-519.

107.   Wu, C., et al., *A protease for 'middle-down' proteomics.* Nat Methods, 2012. **9**(8): p. 822-4.

108.   Altelaar, A.F.M., J. Munoz, and A.J.R. Heck, *Next-generation proteomics: towards an integrative view of proteome dynamics.* Nature Reviews Genetics, 2012. **14**: p. 35.

109.   Mayne, J., et al., *Bottom-Up Proteomics (2013-2015): Keeping up in the Era of Systems Biology.* Anal Chem, 2016. **88**(1): p. 95-121.

110.   Andrews, N.C. and D.V. Faller, *A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells.* Nucleic Acids Res, 1991. **19**(9): p. 2499.

111.   Babu, M., et al., *Sequential peptide affinity purification system for the systematic isolation and identification of protein complexes from Escherichia coli.* Methods Mol Biol, 2009. **564**: p. 373-400.

112. Colantonio, D.A., et al., *Effective removal of albumin from serum.* Proteomics, 2005. **5**(15): p. 3831-5.
113. Muller, T. and D. Winter, *Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents.* Mol Cell Proteomics, 2017. **16**(7): p. 1173-1187.
114. Olsen, J.V., S.E. Ong, and M. Mann, *Trypsin cleaves exclusively C-terminal to arginine and lysine residues.* Mol Cell Proteomics, 2004. **3**(6): p. 608-14.
115. Chen, E.I., et al., *Optimization of mass spectrometry-compatible surfactants for shotgun proteomics.* J Proteome Res, 2007. **6**(7): p. 2529-38.
116. Andersen, J.S., et al., *Proteomic characterization of the human centrosome by protein correlation profiling.* Nature, 2003. **426**(6966): p. 570-4.
117. Boersema, P.J., S. Mohammed, and A.J.R. Heck, *Hydrophilic interaction liquid chromatography (HILIC) in proteomics.* Analytical and Bioanalytical Chemistry, 2008. **391**(1): p. 151-159.
118. Yang, F., et al., *High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis.* Expert Rev Proteomics, 2012. **9**(2): p. 129-34.
119. MacNair, J.E., K.C. Lewis, and J.W. Jorgenson, *Ultrahigh-pressure reversed-phase liquid chromatography in packed capillary columns.* Anal Chem, 1997. **69**(6): p. 983-9.
120. Iwasaki, M., et al., *One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the Escherichia coli proteome on a microarray scale.* Anal Chem, 2010. **82**(7): p. 2616-20.
121. Iwasaki, M., et al., *Human proteome analysis by using reversed phase monolithic silica capillary columns with enhanced sensitivity.* J Chromatogr A, 2012. **1228**: p. 292-7.
122. Yamana, R., et al., *Rapid and deep profiling of human induced pluripotent stem cell proteome by one-shot NanoLC-MS/MS analysis with meter-scale monolithic silica columns.* J Proteome Res, 2013. **12**(1): p. 214-21.
123. Andersson, L. and J. Porath, *Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography.* Anal Biochem, 1986. **154**(1): p. 250-4.
124. Posewitz, M.C. and P. Tempst, *Immobilized gallium(III) affinity chromatography of phosphopeptides.* Anal Chem, 1999. **71**(14): p. 2883-92.
125. Michel, H., et al., *Tandem mass spectrometry reveals that three photosystem II proteins of spinach chloroplasts contain N-acetyl-O-phosphothreonine at their NH2 termini.* J Biol Chem, 1988. **263**(3): p. 1123-30.
126. Beausoleil, S.A., et al., *A probability-based approach for high-throughput protein phosphorylation analysis and site localization.* Nat Biotechnol, 2006. **24**(10): p. 1285-92.
127. Gonzalez-Begne, M., et al., *Characterization of the human submandibular/sublingual saliva glycoproteome using lectin affinity chromatography coupled to multidimensional protein identification technology.* J Proteome Res, 2011. **10**(11): p. 5031-46.
128. Zhang, H., et al., *Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry.* Nat Biotechnol, 2003. **21**(6): p. 660-6.
129. Hagglund, P., et al., *A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation.* J Proteome Res, 2004. **3**(3): p. 556-66.
130. Monzo, A., G.K. Bonn, and A. Guttman, *Boronic acid-lectin affinity chromatography. 1. Simultaneous glycoprotein binding with selective or combined elution.* Anal Bioanal Chem, 2007. **389**(7-8): p. 2097-102.
131. Gingras, A.C., et al., *Analysis of protein complexes using mass spectrometry.* Nat Rev Mol Cell Biol, 2007. **8**(8): p. 645-54.

132.   von Mering, C., et al., *Comparative assessment of large-scale data sets of protein–protein interactions.* Nature, 2002. **417**: p. 399.

133.   Malovannaya, A., et al., *Analysis of the human endogenous coregulator complexome.* Cell, 2011. **145**(5): p. 787-99.

134.   Einhauer, A. and A. Jungbauer, *The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins.* J Biochem Biophys Methods, 2001. **49**(1-3): p. 455-65.

135.   Galan, J.A., et al., *Proteomic studies of Syk-interacting proteins using a novel amine-specific isotope tag and GFP nanotrap.* J Am Soc Mass Spectrom, 2011. **22**(2): p. 319-28.

136.   Hillenkamp, F., et al., *Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers.* Anal Chem, 1991. **63**(24): p. 1193a-1203a.

137.   Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

138.   Tanaka, K., *The origin of macromolecule ionization by laser irradiation (Nobel lecture).* Angew Chem Int Ed Engl, 2003. **42**(33): p. 3860-70.

139.   Amstalden van Hove, E.R., D.F. Smith, and R.M. Heeren, *A concise review of mass spectrometry imaging.* J Chromatogr A, 2010. **1217**(25): p. 3946-54.

140.   Felitsyn, N., M. Peschke, and P. Kebarle, *Origin and number of charges observed on multiply-protonated native proteins produced by ESI.* International Journal of Mass Spectrometry, 2002. **219**(1): p. 39-62.

141.   Steen, H. and M. Mann, *The abc's (and xyz's) of peptide sequencing.* Nature Reviews Molecular Cell Biology, 2004. **5**: p. 699.

142.   Karas, M., et al., *Matrix-assisted ultraviolet laser desorption of non-volatile compounds.* International Journal of Mass Spectrometry and Ion Processes, 1987. **78**: p. 53-68.

143.   Paul, W. and H. Steinwedel, *Ein Neues Massenspektrometer Ohne Magnetfeld*. Vol. 8. 1953. 448.

144.   The Nobel Prize in Physics 1989. *NobelPrize.org. Nobel Media AB 2018*. Wed. 5 Dec 2018; Available from: https://www.nobelprize.org/prizes/physics/1989/summary/.

145.   Douglas, D.J., A.J. Frank, and D. Mao, *Linear ion traps in mass spectrometry.* Mass Spectrom Rev, 2005. **24**(1): p. 1-29.

146.   Hollerbach, A., P.W. Fedick, and R.G. Cooks, *Ion Mobility-Mass Spectrometry Using a Dual-Gated 3D Printed Ion Mobility Spectrometer.* Anal Chem, 2018. **90**(22): p. 13265-13272.

147.   Hardman, M. and A.A. Makarov, *Interfacing the orbitrap mass analyzer to an electrospray ion source.* Anal Chem, 2003. **75**(7): p. 1699-705.

148.   Senko, M.W., et al., *A high-performance modular data system for Fourier transform ion cyclotron resonance mass spectrometry.* Rapid Commun Mass Spectrom, 1996. **10**(14): p. 1839-44.

149.   Makarov, A., et al., *Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer.* J Am Soc Mass Spectrom, 2006. **17**(7): p. 977-982.

150.   Lange, V., et al., *Selected reaction monitoring for quantitative proteomics: a tutorial.* Mol Syst Biol, 2008. **4**: p. 222.

151.   Senko, M.W., et al., *Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates.* Anal Chem, 2013. **85**(24): p. 11710-4.

152.   Ens, W. and K.G. Standing, *Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules.* Methods Enzymol, 2005. **402**: p. 49-78.

153.   Navarrete-Perea, J., et al., *Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3.* J Proteome Res, 2018. **17**(6): p. 2226-2236.

154. Paulo, J.A., J.D. O'Connell, and S.P. Gygi, *A Triple Knockout (TKO) Proteomics Standard for Diagnosing Ion Interference in Isobaric Labeling Experiments.* J Am Soc Mass Spectrom, 2016. **27**(10): p. 1620-5.

155. Li, S., et al., *Selective fragmentation of the N-glycan moiety and protein backbone of ribonuclease B on an Orbitrap Fusion Lumos Tribrid mass spectrometer.* Rapid Commun Mass Spectrom, 2018. **32**(23): p. 2031-2039.

156. Hao, L., et al., *Mass Defect-Based N,N-Dimethyl Leucine Labels for Quantitative Proteomics and Amine Metabolomics of Pancreatic Cancer Cells.* Anal Chem, 2017. **89**(2): p. 1138-1146.

157. Demartini, D., *A Short Overview of the Components in Mass Spectrometry Instrumentation for Proteomics Analyses*. 2013. p. 39-57.

158. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis.* Nat Methods, 2007. **4**(9): p. 709-12.

159. Chi, H., et al., *pNovo: de novo peptide sequencing and identification using HCD spectra.* J Proteome Res, 2010. **9**(5): p. 2713-24.

160. Sadygov, R.G., D. Cociorva, and J.R. Yates, 3rd, *Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book.* Nat Methods, 2004. **1**(3): p. 195-202.

161. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-89.

162. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.

163. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.

164. Kall, L., et al., *Semi-supervised learning for peptide identification from shotgun proteomics datasets.* Nat Methods, 2007. **4**(11): p. 923-5.

165. Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry.* Rapid Commun Mass Spectrom, 2003. **17**(20): p. 2337-42.

166. Johnson, R.S. and J.A. Taylor, *Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry.* Mol Biotechnol, 2002. **22**(3): p. 301-15.

167. Frank, A. and P. Pevzner, *PepNovo: de novo peptide sequencing via probabilistic network modeling.* Anal Chem, 2005. **77**(4): p. 964-73.

168. Jazurek, M., et al., *Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases.* Nucleic acids research, 2016. **44**(19): p. 9050-9070.

169. Paša-Tolić, L., et al., *High Throughput Proteome-Wide Precision Measurements of Protein Expression Using Mass Spectrometry.* Journal of the American Chemical Society, 1999. **121**(34): p. 7949-7950.

170. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.

171. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.

172. Hsu, J.L., et al., *Stable-isotope dimethyl labeling for quantitative proteomics.* Anal Chem, 2003. **75**(24): p. 6843-52.

173. Boersema, P.J., et al., *Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics.* Nat Protoc, 2009. **4**(4): p. 484-94.

174. Reynolds, K.J., X. Yao, and C. Fenselau, *Proteolytic 18O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent.* J Proteome Res, 2002. **1**(1): p. 27-33.

175. Yao, X., et al., *Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus.* Anal Chem, 2001. **73**(13): p. 2836-42.

176. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.

177. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. **75**(8): p. 1895-904.

178. McAlister, G.C., et al., *Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses.* Anal Chem, 2012. **84**(17): p. 7469-78.

179. Rauniyar, N. and J.R. Yates, 3rd, *Isobaric labeling-based relative quantification in shotgun proteomics.* Journal of proteome research, 2014. **13**(12): p. 5293-5309.

180. Li, Z., et al., *Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos.* J Proteome Res, 2012. **11**(3): p. 1582-90.

181. Louris, J.N., et al., *Instrumentation, applications, and energy deposition in quadrupole ion-trap tandem mass spectrometry.* Analytical Chemistry, 1987. **59**(13): p. 1677-1685.

182. Ow, S.Y., et al., *iTRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly".* Journal of Proteome Research, 2009. **8**(11): p. 5347-5355.

183. Ow, S.Y., et al., *Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation.* Proteomics, 2011. **11**(11): p. 2341-6.

184. Savitski, M.M., 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.* Anal Chem, 2011. **83**(23): p. 8959-67.

185. Wenger, C.D., et al., *Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging.* Nat Methods, 2011. **8**(11): p. 933-5.

186. Washburn, M.P., D. Wolters, and J.R. Yates Iii, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nature Biotechnology, 2001. **19**: p. 242.

187. Liu, H., R.G. Sadygov, and J.R. Yates, 3rd, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics.* Anal Chem, 2004. **76**(14): p. 4193-201.

188. Cutillas, P.R. and B. Vanhaesebroeck, *Quantitative profile of five murine core proteomes using label-free functional proteomics.* Mol Cell Proteomics, 2007. **6**(9): p. 1560-73.

189. Bondarenko, P.V., D. Chelius, and T.A. Shaler, *Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry.* Anal Chem, 2002. **74**(18): p. 4741-9.

190. Chelius, D., et al., *Analysis of the adenovirus type 5 proteome by liquid chromatography and tandem mass spectrometry methods.* J Proteome Res, 2002. **1**(6): p. 501-13.

191. Old, W.M., et al., *Comparison of label-free methods for quantifying human proteins by shotgun proteomics.* Mol Cell Proteomics, 2005. **4**(10): p. 1487-502.

192. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.* Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.

193. Christoforou, A.L. and K.S. Lilley, *Isobaric tagging approaches in quantitative proteomics: the ups and downs.* Anal Bioanal Chem, 2012. **404**(4): p. 1029-37.

Deciphering the protein dynamics and molecular determinants of iPSC-derived neurons.

Suzy Varderidou-Minasian[1,2], Philipp Schätzle[3], Casper.C.Hoogenraad[3], R. Jeroen Pasterkamp[4], Maarten Altelaar[1,2]

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

[2] Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, the Netherlands.

[3] Cell Biology, Department of Biology, Faculty of Science, Utrecht University, 3584 CH Utrecht, the Netherlands.

[4] Department of Translational Neuroscience, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, 3584 CG Utrecht, the Netherlands.

SUMMARY

Neuronal development is a multistep process with different regulatory programs that shapes neurons to form dendrites, axons and synapses. To date, knowledge on neuronal development is largely based on murine data and largely restricted to the genomic and transcriptomic level. Advances in stem cell differentiation now enable the study of human neuronal development, and here we provide a mass spectrometry-based quantitative proteomic signature, at high temporal resolution, of human stem cell-derived neurons. To reveal proteomic changes during neuronal development we make use of two different differentiation approaches, leading to glutamatergic induced neurons (iN) or small molecule-derived patterned motor neurons. Our analysis revealed key proteins that show significant expression changes (FDR <0.001) during neuronal differentiation. We overlay our proteomics data with available transcriptomic data during neuronal differentiation and show distinct, datatype-specific, signatures. Overall, we provide a rich resource of information on proteins associated with human neuronal development, and moreover, highlight several signaling pathways involved, such as Wnt and Notch.

KEYWORDS

INTRODUCTION

The human brain is a complex system with different regions and cell types, having billions of cells and trillions of synapses [1, 2]. This diversity presents a great challenge in understanding the molecular and cellular function of this organ. As recent studies have revealed, perturbation of brain development underlies many neurological disorders such as autism and schizophrenia, however, much of our current knowledge is derived from the rodent brain [3-5]. Human neural development remains difficult to study given the ethical constraints in primary human brain tissues, together with the paucity of high-quality post-mortem tissue. Moreover, the degree of cell and tissue heterogeneity, in combination with complex developmental and environmental factors, further complicate (human) neuronal research [6, 7]. One approach with great promise to study neurological disorders is the use of induced pluripotent stem cells (iPSC) from e.g. human fibroblasts [8]. Ever since the first report on iPSC, major efforts have been directed towards developing differentiation protocols to induce neurons [9, 10]. Given the rapid developments in the field of iPSC-derived neurons, a comprehensive understanding of the mechanism underlying iPSC differentiation towards neurons is required. Neuronal development is coordinated by morphogens and neurogenic factors that can be captured *in vitro* using iPSC [11, 12]. Over the last years, major advances

in iPSC differentiation improved the generation of homogeneous population of neurons and has been used to study various neurological disorders [13]. Although many of these regulatory pathways involved in neuronal development have been studied on genomic and transcriptomic studies, their mechanisms at protein levels have not [14]. Since proteins are the final molecular effectors of cellular processes and their perturbation is linked to pathological states, their investigation is essential.

Multiple protocols exist for generating neurons from iPSC. Here, to monitor the differentiation process of iPSC-derived neurons by high-resolution proteomics, we adapted two different approaches, often used to model neuronal development and neurological disorders [15-18]. Forced expression of a single neurogenic transcription factor (Ngn2) causes rapid differentiation of human iPSC into functional excitatory cortical neurons (iN cells) [19]. This approach shows, within 10 days, rapid and reproducible production of a homogeneous population of glutamatergic neurons. In addition, extrinsic-factor-based strategies of different morphogens such as Wnt, fibroblast growth factor (FGF), retinoic acid (RA) and sonic hedgehog (SHH) can be used to generate neuronal subtypes [20]. Here, the course of differentiation is a three-step process, with neural crest cell activation by dual SMAD inhibition, caudalization by RA signaling and ventralization by SHH signaling. We will refer to these neurons as patterned motor neurons (MNs). Both approaches can be used as model systems to study the molecular mechanisms during neuronal development.

The research presented here quantitatively probes proteome changes during differentiation of iN cells and patterned MNs at 10 different time points **(Figure 1)**. We observe a two-step resetting of the global proteome, showing abundant proteins in iPSC decreasing and neuronal proteins increasing over time. We highlight both well-established and novel proteins up- and downregulated during differentiation. Additionally, we compare our proteomic data with comparable transcriptomic data during neuronal differentiation showing distinct (gene/protein) signatures, and we illustrate the relative fold change of proteins associated with signaling pathways such as Wnt, Notch and hedgehog signaling. Finally, we illustrate which proteins are specifically changing during differentiation of iPSC into either iN cells or patterned MNs.

RESULTS

Differentiation towards iN cells and patterned MNs

Differentiation into iN cells was performed by doxycycline-induced expression of neurogenin-2 (Ngn2) [19]. At day 7, the iPSC produced an apparently mature neuronal morphology and were positive for the neuronal marker TuJ1 and the cortical marker FOXG1 (**Figure 2A**). In addition, these cells were negative for MN marker ISLET1. Patterned MNs were generated by the combined actions of small molecules for neural induction and cell fate determination [20] **(Figure 2B, supplementary video)**. At day 20, cells were positive for βIII-tubulin (>80%) and ISL1 (~40 %) markers. For mass spectrometry-based quantitative proteomics, we performed TMT-10plex labeling of the tryptic peptides originating from 10 distinct time points within the differentiation timeline of either iN cells or patterned MNs, resulting in a total of four replicates **(Figure 1)**. After separation by high-pH fractionation, the labeled peptides were analyzed with high-resolution tandem mass spectrometry (LC-MS/MS).
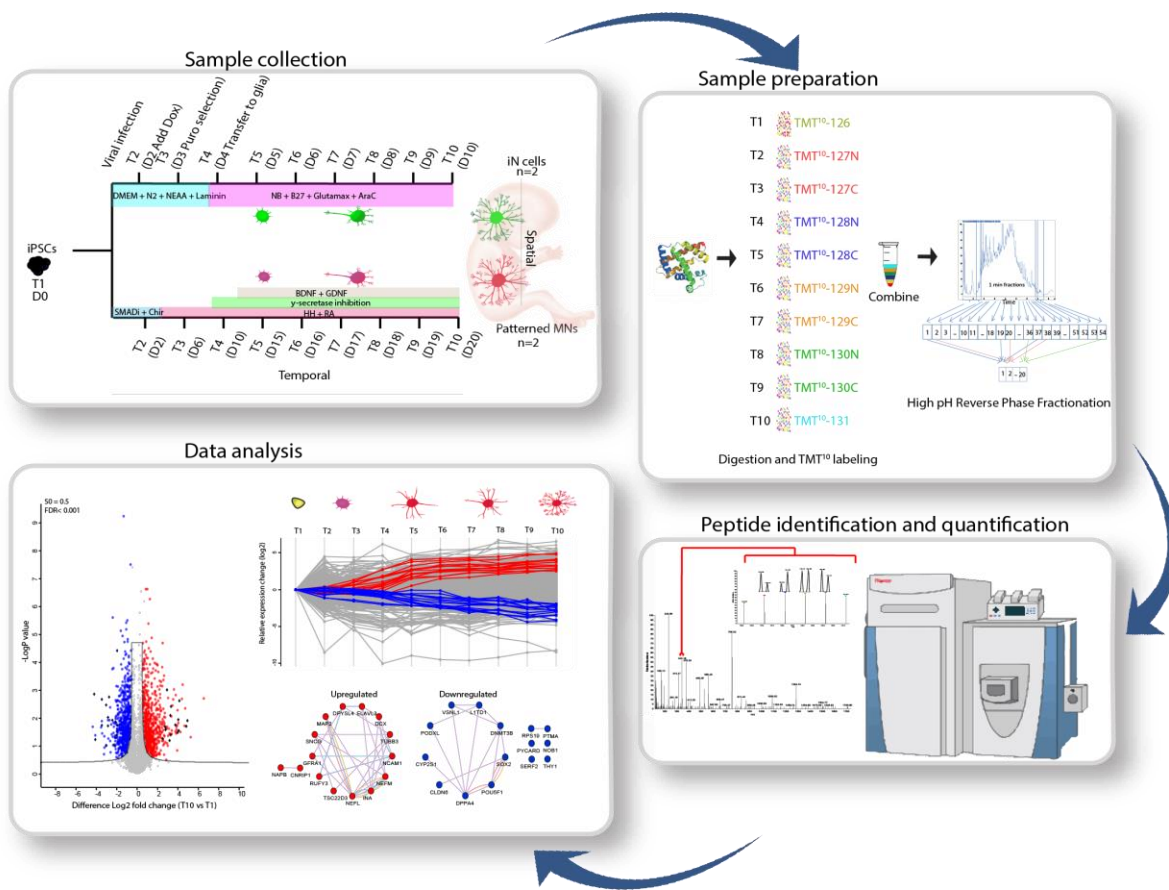
Figure 1. Workflow of MS-based quantitative proteomics during neuronal differentiation. Differentiation of iPSC towards iN cells was performed using doxycycline-induced expression of Ngn2. Differentiation of patterned MNs was performed using the action of small molecules for neural induction and cell fate determination. Proteins extracted at ten time points from two biological replicates for each cell type were digested and TMT 10-plex labeled. Peptides were combined and fractionated using high pH fractionation. The resulting fractions were analyzed by high-resolution nano-LC-MS/MS and quantification was achieved using TMT 10-plex isobaric labeling.
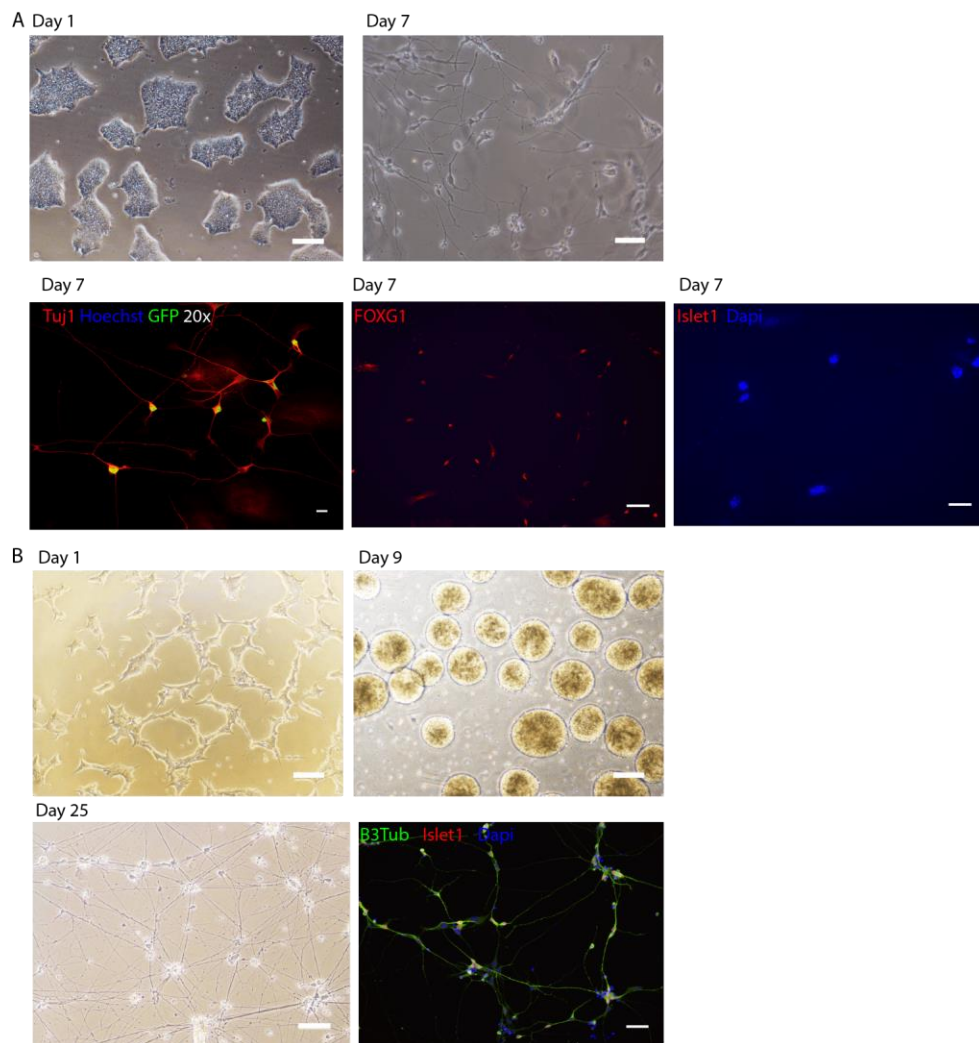

Figure 2. Bright field images and immunocytochemistry of iN cells and patterned MNs at indicated time points. (A). Bright field images of iPSC at day 1, iN cells at day 7 and immunostainings for neuronal marker (TuJ1), cortical marker (FOXG1) and motor neuron marker (Islet1). (B). Bright field images of patterned MNs and immunocytochemistry staining for neuronal marker (Tubulin beta 3), motor neuron marker (Islet1) and nuclear marker (Dapi). Scale bars = 100 µm

MS-based quantitative proteomics

We identified 4070 and 3517 protein groups with a false discovery of 1%, in iN cells and in patterned MNs, respectively **(Figure S1A)**. The relative abundances of the quantified proteins span more than four orders of magnitude, indicating a broad dynamic range in our quantitative measurement. Tissue enrichment

analysis of the 25% most abundant proteins against the whole human proteome background, revealed enrichment for 'Cajal-Retzius cell', 'fetal brain cortex' and 'epithelium' **(Figure S1B)**. The expression of Cajal-Retzius cells is not surprising since these cells are involved in the organization of brain development [21]. The enrichment associated with epithelium may be caused by origin of the fibroblast. Finally, annotation of protein class for all identified proteins revealed a wide coverage, including typically low abundant protein classes such as transcription factors and storage proteins **(Figure S1C)**. Furthermore, high run-to-run quantitative reproducibility was observed between biological replicates both for iN cells and patterned MN differentiation **(Figure S1D)**. The TMT reporter intensity values of proteins identified in each time point were normalized to the reference intensity value of these proteins before initiation of differentiation (T1). As a consequence, all data throughout the study reflects a ratio change relative to T1. After global analysis of our proteomics data, we set out to confirm the quality of the neuronal differentiation. We investigated known stem cell markers (e.g. SOX2 and OCT4) that indeed decreased during differentiation while, as expected, several neuronal markers (e.g. NEFL, GAP43 and MAP2) increased in both neuron types **(Figure 3A)** [22-28]. Progenitor markers (e.g. BIRC5, SPARC and TOP2A) that are required for fetal development and regulation of cortex development show, in our data, higher expression in iN cells compared to patterned MNs [29-32]. Furthermore, the rostral marker OTX1 was only identified in iN cells and the caudal marker HOXB5 only in patterned MNs. Finally, NEUROG2 expression was only observed in iN cells, where it is overexpressed, and previous studies showed its involvement in cortical specification [33]. Taken together, these data confirm the validity of our approach, showing neuronal subtype-specific development.
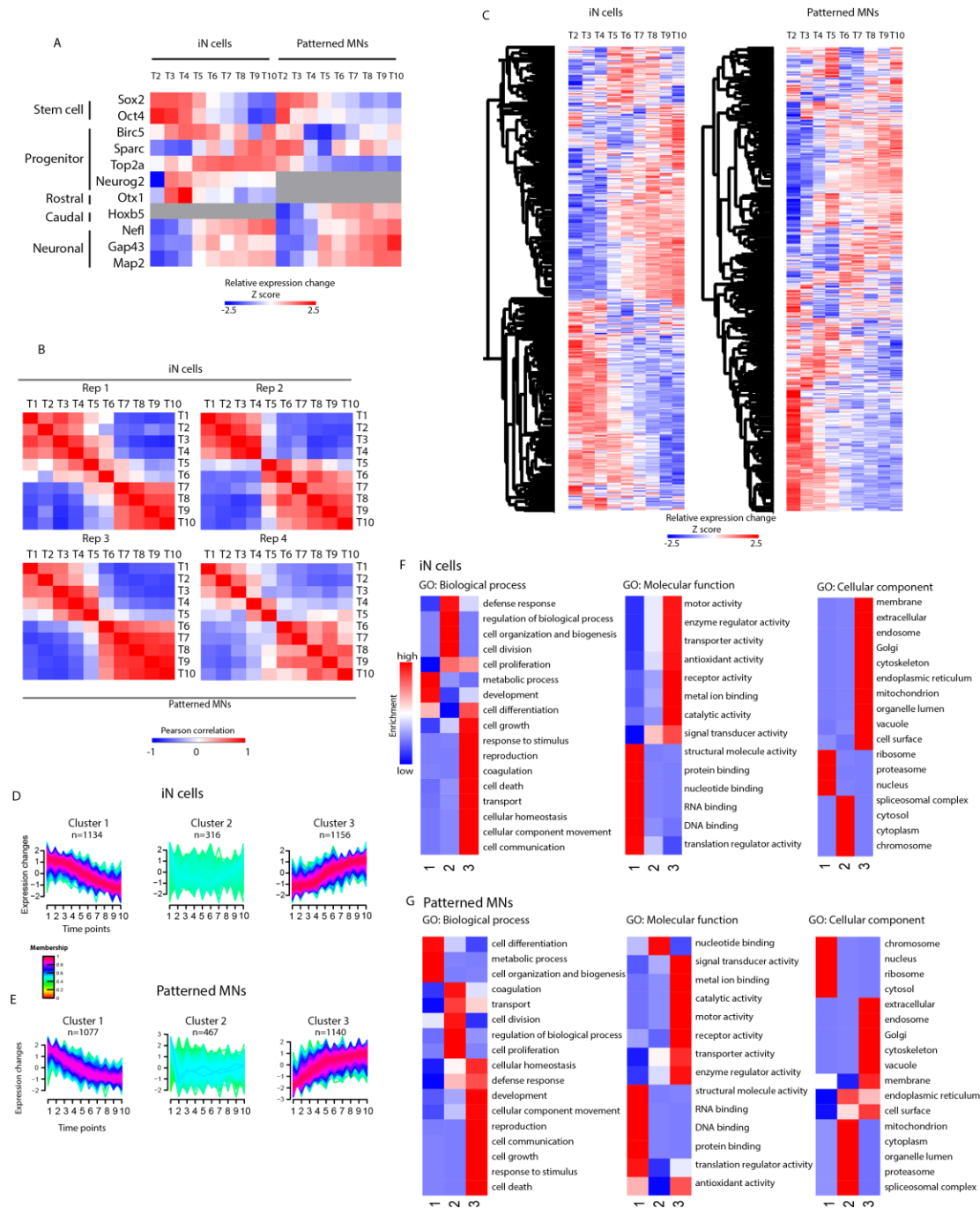
Figure 3. Proteome dynamics during neuronal differentiation. (A). Differential expression of general iPSC, progenitor and neuronal markers represented in a heat map. Each row represents average protein log2 ratios relative to iPSC (T1) in both replicates for each approach. (B). Pearson correlation within each biological replicate. Note for each biological replicate the high correlation between early time point (T1-5) and late (T6-T10). (C). Heat map of all proteins identified in iN cells and patterned MNs showing the expression changes along the course of differentiation for each approach relative to T1. (D-E). Clusters of the protein dynamics during differentiation towards iN cells (E) and patterned MNs (E). Three clusters with increasing, decreasing and not changing profiles were revealed in both differentiation approaches. Upper and lower ratio limit of log2 (0.5) and log2 (-0.5) was selected for inclusion into a cluster. The number of proteins is represented by n while the membership value indicates how

well each protein fits to the average cluster profile. (F-G). Gene ontology enrichment analysis of each cluster tested for overrepresented biological processes (BP), molecular function (MF) and cellular component (CC) compared to unregulated proteins.

To obtain a global proteome view, changes of proteins along the course of differentiation for the two approaches were evaluated **(Figure 3B, C)**. We observed a 'two-step resetting' of the global proteome, with proteins highly expressed in iPSC showing down-regulation over time, and vice versa, proteins with low expression in iPSC increasing over time. This distinct separation of the proteome, which happens around the intermediate stage (T5), independent of the differentiation approach, can be further emphasized by performing Pearson correlation within each biological replicate (**Figure 3B**). Here, an anti-correlation (R = between -0.58 and -0.83) was found between T1 and T10, which was similarly observed during reprogramming of somatic cells to pluripotency [34]. Next, we performed an unsupervised clustering of the protein expression across the time points for both differentiation approaches **(Figure 3D, E)**. Gene Ontology (GO) enrichment analysis of these clusters (**Figure 3F, G)** revealed biological processes such as cell communication, growth and movement being upregulated (cluster 3), while cell differentiation and biogenesis is slowly downregulated (cluster 1). Closer inspection of the corresponding proteins in cluster 3 revealed pan-neuronal proteins such as tubulin-beta-3, neurofilament and MAP2. Cluster 3 also reveals increased expression of proteins localized in the membrane, cytoskeleton and several organelles, as well as molecular functions related to transport, enzymatic activity and signaling. Cluster 1 contains nuclear, chromosomal and ribosomal proteins that gradually decrease over time having a function in RNA and DNA binding. Downregulation of RNA processing during differentiation was previously shown in mouse by proteomics analysis and in human by transcriptomic analysis [14, 35]. Cluster 2 includes a group of proteins with general terms such as regulation of biological processes or cell division.

Proteomic changes across time points during neuronal differentiation

To identify differentially expressed proteins during neuronal differentiation in the two approaches, we compared T10 relative to T1 and performed a t-test for both neuronal subtypes separately. **Figure 4A, C** show the volcano plots with the most extreme comparison T10 (mature neurons) against T1 (iPSC) for both approaches, depicting proteins with a FDR below 0.1%. This resulted in 696 significantly downregulated proteins and 686 upregulated proteins in iN cells and 706 significantly downregulated proteins and 620 upregulated proteins in patterned MNs **(Supplementary Table 1, 2)**. Close inspection of the proteins significantly enriched in T10 reveals general microtubule-associated proteins necessary for

neuronal stabilization, outgrowth and migration, whereas specific iPSC enriched proteins consisted of transcription factors and proteins having a role in embryonic development. Furthermore, we highlight the top 15 proteins showing the largest fold change over time in both neuronal differentiation approaches **(Figure 4B, D)**. As expected, proteins such as SOX2, POU5F1 (OCT4) (iPSC markers) and MAP2 and DCX (neuronal markers) were enriched. Moreover, several proteins (e.g. INA, DPYSL4, and RPS19) were identified, that have less established roles in the context of pluripotency and differentiation [36-38]. The analysis of these regulated proteins shows an interaction network around the proteins INA, NCAM, NEFM and NEFL, being upregulated, and POU5F1, SOX2, DNMT3B and DPPA4, being downregulated **(Figure 4E)**.

Next, we compared the significantly altered proteins during differentiation towards iN cells or patterned MNs, and observed only moderate overlap (23%) (**Figure 5A**). The differences in protein expression might suggest specificity towards a neuronal subtype or differentiation approach. We selected proteins with the largest fold change between the two neuronal subtypes (fold change cut-off ≥ 1.5 and p value ≤ 0.1). Proteins that increase most in iN cells (compared to patterned MNs), are S100A11, S100A13 and S100A6 **(Figure 5B)**. The S100 family of calcium binding proteins was first identified in the brain, regulating processes such as cell cycle progression and differentiation [39]. They have been found in sub-cortical structures and in a subpopulation of astrocytes [40-43]. Proteins highly enriched in patterned MNs are PPARD and MDN1. They act downstream of the Wnt-β-catenin and NOTCH signaling pathway [44, 45]. We then selected proteins exclusively identified either in iN cells or patterned MNs that show a minimum twofold change in at least one time point during differentiation **(Figure S5)**. The highly expressed protein SULF2 in iN cells is involved in brain development and neurite outgrowth in mice [46, 47] but its involvement in human neuronal development is unclear. Furthermore, highly expressed proteins in patterned MNs (e.g. PBX3 and HOXB5) regulate dorsal spinal cord development which is in line with the neuronal origin [48].
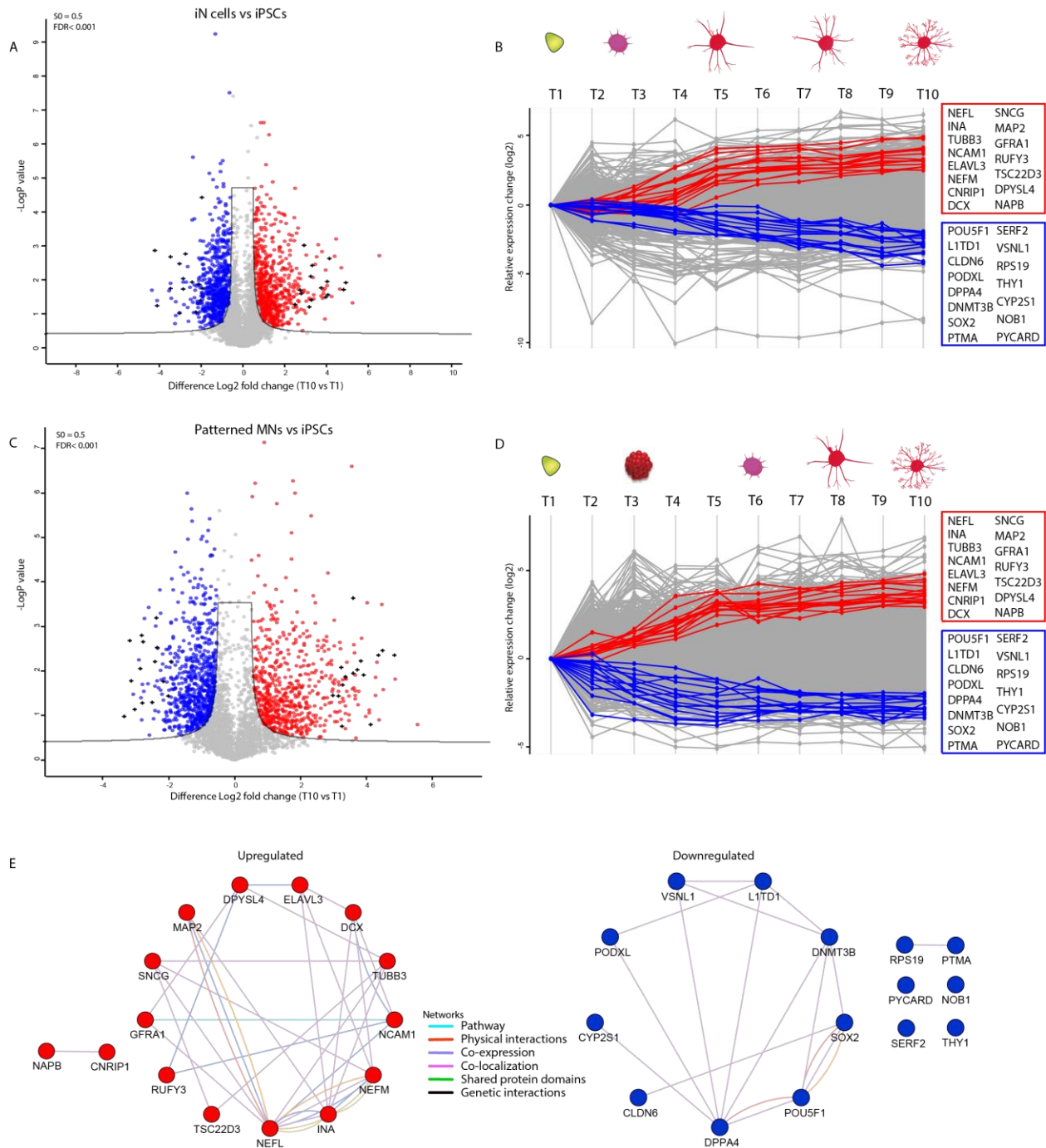
Figure 4. Significance analysis of protein expression changes during neuronal differentiation. (A) Volcano plot illustrating differentially expressed proteins during iN cell differentiation. The −log10 P value is plotted against the log2 fold change of T1 (iPSC) and T10 (mature iN cells) with significant threshold t-test with FDR < 0.001 and S = 0.5. Red represents proteins up regulated in T10 and blue represents proteins downregulated in T10. Black cross represents the top 15 proteins with the highest fold change. (B) Profile plot highlighting the top 15 significant up- and downregulated proteins in iN cells. (C) Volcano plot illustrating differentially expressed proteins during patterned MN differentiation. The −log10 P value is plotted against the log2 fold change of T1 and T10 with significant

threshold t-test with FDR < 0.001 and S = 0.5. Red represents proteins up regulated in T10 and blue represents proteins downregulated in T10. Black cross represents the top 15 proteins with the highest fold change. (D) Profile plot highlighting the top 15 significant up- and downregulated proteins in patterned MN differentiation. (E) Protein network analysis showing the top 15 upregulated or downregulated proteins that interact with each other.
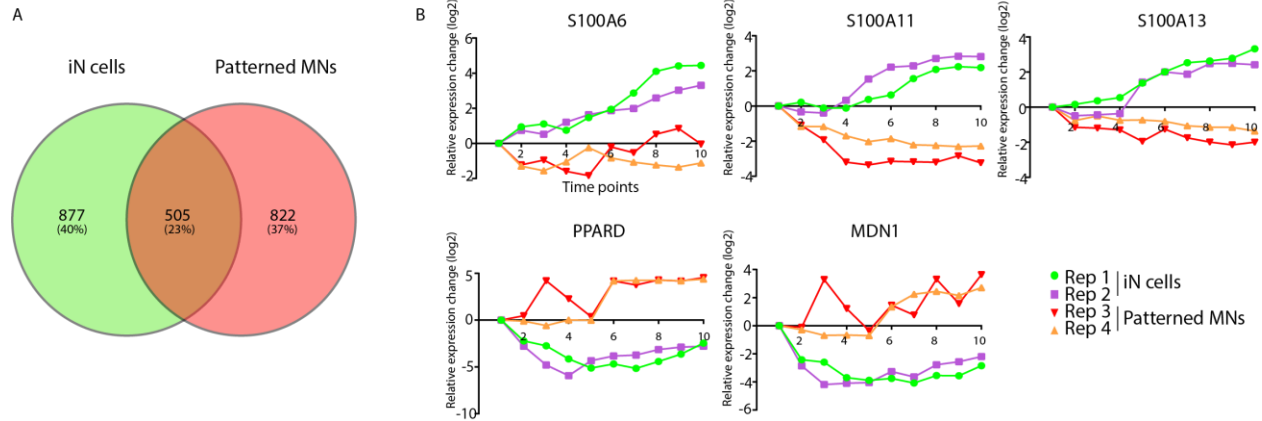

Figure 5. Differential expressed proteins in iN cells and patterned MNs. (A). Venn diagram indicating the overlap between the differentially expressed proteins in iN cells or patterned MNs (FDR 0.1%). (B) Example of proteins showing the largest differences between iN cells and patterned MNs during differentiation.

Neurogenesis associated proteins

To further investigate proteins involved in specific processes related to neuronal differentiation we next looked for enrichment of proteins associated with neurogenesis using the DAVID database (https://david.ncifcrf.gov/summary.jsp). This analysis revealed 129 proteins for which we had a closer look in their differentiation profile, plotting their expression over time in a heatmap **(Figure S2)**. These neurogenesis-associated proteins for the largest part increase their expression over time, such as DCX, CPLX2 and ELAVL3. Upregulated microtubule associated protein DCX has been increasingly used as a marker for neurogenesis and ELAVL3 is a neuron-specific protein [49, 50]. Several proteins also show downregulation, such as SLC7A5 and the transcription factor ZIC2, known to play a role in neuronal cell proliferation (neurogenesis) and early stages of central nervous system development [51, 52]. Additionally, ARF6 is known to regulate neuronal development in the brain via regulation of actin dynamics and synaptic plasticity. Its relative lower expression here however, might indicate that additional factors are needed to fully recapitulate the neuronal development.

Transcription factors and cytoskeletal proteins

During neuronal differentiation, we showed that most of the up- and downregulated proteins are cytoskeletal and transcription factors (TFs), respectively. To better categorize the involvement of these proteins in neuronal development, we aimed to capture all TF's and cytoskeletal proteins identified in our proteomic data **(Figure S3)**. TFs play essential roles in both reprogramming and neuronal development and have been used to induce neuronal differentiation from multiple cell lineages such as human fibroblasts, neuronal progenitors and stem cells [53-55]. We identified 259 regulated TFs (with a fold change cut-off ≥ 2 in at least in one time point for both approaches separately) of which PHC2, CRTC1, NCOA3 and SMAD3 are the most upregulated in both neuronal differentiation approaches **(Figure S3A)**. Interestingly, PHC2 has been reported to have an early developmental role in Drosophila, while its specific function in human cells has not been determined. CRTC1 is ubiquitously expressed in fetal brain and acts as a coactivator of the CREM (cAMP-responsive element binding)-dependent gene transcription pathway [56]. Furthermore, CRTC1 has been associated to different neuronal functions, such as synaptic plasticity and dendritic growth in cortical neurons and is also down-regulated in Alzheimer's disease [57, 58]. NCOA3 has been identified as a novel microRNA regulator of dendritogenesis in mouse cortical neurons [59]. In line with this, we show that NCOA3 increases over the course of human neuronal differentiation. SMAD3 which was shown to promote neuronal differentiation in spinal cord of zebrafish [60] was found to be higher expressed in patterned MNs compared to iN cells. In addition, several TFs are upregulated at early time points of neuronal differentiation. NFXL1 was upregulated at the early time points of both neuronal subtypes while CCNT1 and LITAF for iN cells and ZIC5, NFATC4 and ZBTB40 for patterned MNs. Both ZIC5 and NFATC4 are essential in neural development and survival, however their association in early stages of caudalization, as seen here, has not been studied yet [61, 62].

In addition, we captured 256 cytoskeletal proteins as shown in **Figure S3B** with TUBB3, TAGLN3 and MAP1LC3A being most upregulated in both differentiation approaches. TUBB3 and MAP1LC3A are highly expressed in neurons and function in neurite formation and stabilization [63, 64]. TAGLN3, is an actin-binding protein involved in cytoskeletal organization [65]. Interestingly, the transgelin family members TAGLN, TAGLN2 and TAGLN3 show differential expression between iN cells and patterned MN differentiation. **(Figure S3C)**. TAGLN decreases during patterned MN differentiation while in iN cells there is an increase only in the early time points. TAGLN2 has a moderate increase in iN cells and decreases in the early time points of patterned MN differentiation. Only the third member of the transgelin family (TAGLN3) shows strong upregulation during neuronal differentiation in both approaches.

Signaling pathways

In the last 30 years, several studies have demonstrated the importance of signaling pathways for regulating the developmental program, specify cell fate and patterning [66, 67]. Some of these signaling pathways will be discussed here, displaying the expression of their associated proteins in heatmaps for both iN cells and patterned MNs **(Figure S4)**. The TGFβ/ BMP signaling is known to inhibit neuronal development by blocking the proliferation of precursor cells in the adult brain. The dual SMAD inhibitor LDN-193189 and SB-431542 (together SMADi) induce the specification of cells with neural plate identity by selectively inhibiting the TGFβ/ BMP signaling pathway [68]. We applied this to our patterned MN differentiation protocol (at T2) and identified expression changes for all SMAD members, as well as SMAD interacting protein 1 (SNIP1) over time **(Figure S4A)**. SMAD3 rapidly increases at the late time points of patterned MN differentiation whereas the other members remain more constant. Interestingly, we observed in our iN cell differentiation protocol, that did not use of SMADi a similar expression pattern over time for most of the SMAD family members. Wnt signaling regulates cell migration, cell polarity and neuronal development during embryonic development and has an essential role in body axis formation, particularly in the formation of anteroposterior and dorsoventral axes [69-71]. Furthermore, the combinatorial effects of Wnt, retinoic acid (RA) and hedgehog (HH) signaling specify neuronal subtypes [72, 73]. Upon activation of Wnt signaling in patterned MNs (by Chir-99021 at T2 for four days), elevated levels of WNT5A, WLS, ZBTB16 and NFATC4 are observed in the early time points of patterned MNs while not identified or very low expressed in iN cells **(Figure S4B)**. These proteins are involved in cell fate pattering and development [74-77] . PLCB1 and PLCB4 are highly expressed in iN cells, compared to patterned MNs. A previous study showed their high expression and differential distribution in several regions of the brain including cortex, suggesting a specific role in different regions of the brain [78]. Proteins associated with hedgehog (HH) and retinoic acid (RA) signaling act as posteriorizing agent for spinal cord development [79, 80]. Both HH and RA are induced at T3 in the patterned MN differentiation protocol and kept till the end of the experiment. We looked for proteins associated with HH and RA and observed that the majority of these proteins are slightly elevated in patterned MNs compared to iN cells **(Figure S4C, D)**. CSNK1E and GSK3B together with CCND2 protein are higher in expression in the patterned MNs compared to iN cells. GSK3B has a critical role in axonal regulation and CCND2 has a function in neuronal development [81, 82].

The expression of cellular retinoic acid-binding protein 1 and 2 (CRABP1 and CRABP2) is increased in the patterned MNs after addition of RA. Interestingly, CRABP1 expression was restricted to certain neuronal subtypes in the hypothalamus suggesting for a cell-type specific function in the brain [83].

Homeobox proteins, which promote the expression of posterior neural genes, are all identified in the patterned MNs but not in the iN cells **(Figure S4E)** [84].

Inhibition of Notch signaling is induced (at T4) for the patterned MN differentiation, which is necessary to induce proneural genes [85, 86]. We identified 81 proteins involved in notch signaling, of which THBS1 showed a direct response after induction in the patterned MNs **(Figure S4F)**. THBS1 has been shown to play roles in neuronal development, such as in neurite outgrowth and cell migration [87, 88]. We further explored our data for enrichment of axon guidance proteins. They are important in regeneration along the anterior-posterior axis [89]. We identified 36 proteins associated with axon guidance of which the majority increases along the course of patterned MN differentiation **(Figure S4G)**. As described previously, axonal guidance cues are often categorized as 'attractive' or 'repulsive' [90]. NTN1 and its receptor (DCC), having attractive roles, are both increased in expression only in patterned MNs whereas ENAH and VASP, which function downstream of the repulsive axon guidance receptor Robo, are downregulated along the course of differentiation for both approaches [91].
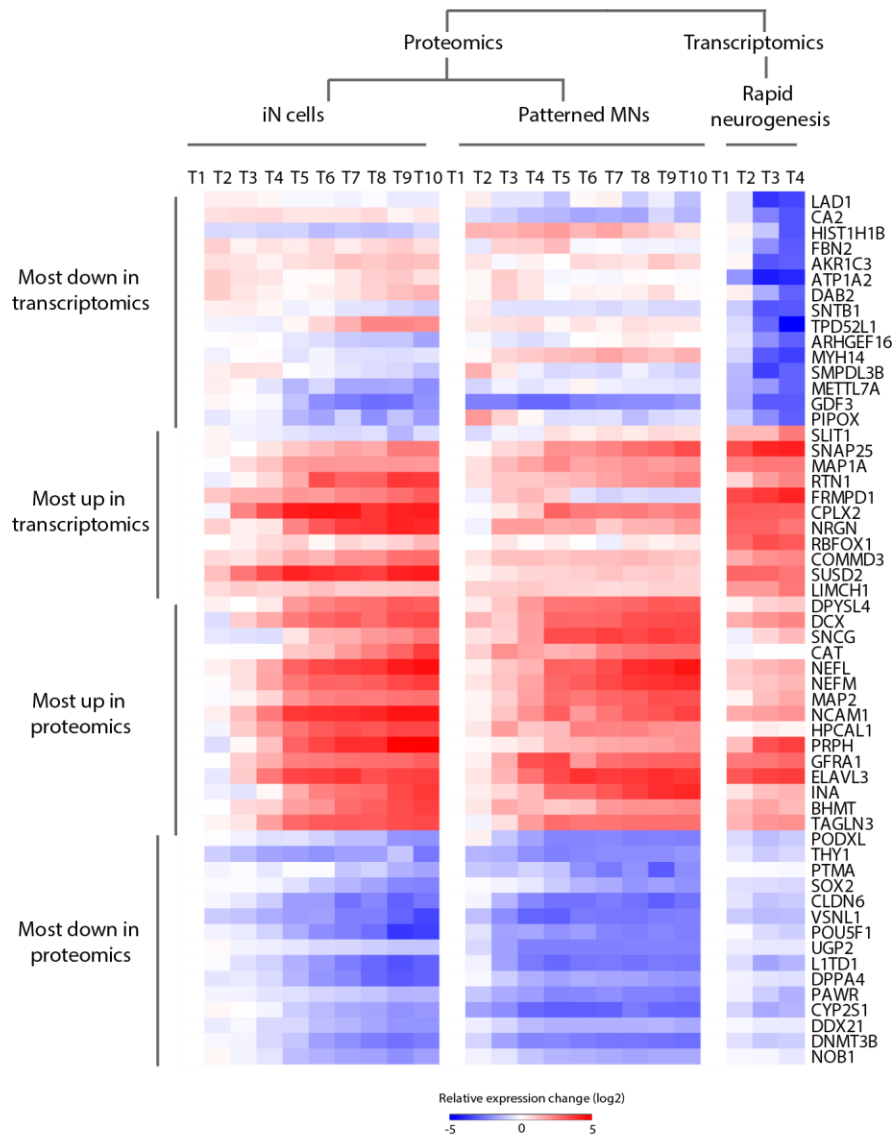
Figure 6. Comparison proteomic and transcriptomic profile. Heatmap of the top 15 most up- and downregulated proteins and top 15 most up- and downregulated genes along the course of differentiation towards neurons.

Comparison of neuronal development expression profile by quantitative proteomics and transcriptomics

We compared our proteomic measurements with RNA-seq data derived from a comparable transcriptome analysis during neuronal differentiation by overexpression of two transcription factors (Neurogenin-1 and Neurogenin-2) that resulted in homogeneous population of neurons within 4 days [14]. The RNA-seq data covers four time points with T1, which resembles the iPSC, and T4, which resembles mature neurons

across the full course of iPSC differentiation. Out of 16589 transcripts that were identified, 5010 proteins were identified in our proteomics data. We selected the top 15 up- or downregulated transcripts and proteins during the differentiation and compared their expression levels (**Figure 6**). Interestingly, the majority of the transcripts that were downregulated showed a relative high protein expression level, as for instance TPD52L1, AKR1C3 and DAB2. Recent data have shown that these proteins are regulators of cell differentiation [92-94]. The majority of these most upregulated transcripts were upregulated in proteomics data for both neuronal subtypes and, vice versa, the majority of these upregulated proteins were upregulated in transcriptomic data. Interestingly, DCX, ELAVL3, GFRA1 and PRPH were identified as among the top 15 most upregulated (transcripts/proteins) in both transcriptomic and proteomic analysis separately, further emphasizing their crucial role in neuronal differentiation. They are all expressed in mature neurons and adult brain, however, only DCX is used as a neuronal marker [95-98]. The majority of the most downregulated proteins decrease to a lesser extent in transcriptomic data. PTMA for example, which is only downregulated on proteome level, was previously shown to be highly expressed in iPSC and is one of the validated targets of the c-MYC reprogramming transcription activator [99]. Together, these variations, between transcriptome and proteome, suggest that additional mechanisms such as posttranscriptional regulatory mechanisms might control the expression level of several proteins involved in neuronal differentiation.

DISCUSSION

Our study provide insight into the remodeling of the proteome during human neuronal development. To our knowledge, this is the first comprehensive proteomic profiling of human iPSC differentiation towards neurons. The rapid neurogenesis through transcriptional activation towards iN cells and the introduction of small molecules towards patterned MNs allowed us to identify key regulatory proteins involved in differentiation. Quantitative proteomics was applied to profile the dynamic changes of proteins during neuronal differentiation by using TMT 10-plex labeling coupled to high resolution LC-MS/MS, which resulted in the identification of 7230 proteins. We found a proteome-wide change in expression, which occurs in a two-step fashion, revealing a clear switch in protein expression levels, halfway our time window, leading to an anti-correlation between iPSC (T1) and neurons (T10). A similar but opposite trend was observed previously in a study monitoring proteome changes during the course of fibroblast reprogramming to iPSC [34]. According to GO classification, most of the proteins that downregulate during differentiation are located in the nucleus and are involved in RNA and DNA binding. Proteins that upregulate are mainly cytoskeletal proteins involved in cell communication and transport that could play

a role in dendritic/axon outgrowth and branching. These findings are in line with our previous study, characterizing the proteome changes of rat hippocampal neuron development [35]. The majority of the proteins are steadily upregulated or downregulated along the course of differentiation. We highlighted the most extreme comparison of iPSC (T1) against mature neurons (T10) and showed the top 15 most up- or downregulated proteins during neuronal differentiation, consisting of proteins such as MAP2, a neuron-specific cytoskeletal protein having an established role in neurodevelopment [100]. Moreover, we detect several proteins (e.g. INA, DPYSL4, and RPS19) being highly upregulated with less established association with neuronal development suggesting that they may be considered as novel pan neuronal markers. In conjunction, we captured all TFs and cytoskeletal proteins in our data and illustrated their expression changes during neuronal differentiation towards the two subtypes. Several of these TF's (PHC2, CRTC1, NCOA3 and SMAD3) have been reported to promote neuronal differentiation in zebrafish and mice, however their association to human neuronal differentiation was not detected. In addition, we illustrated several subunits of a protein family having differential expression profiles in both cell types, such as strong upregulation of TAGLN3 in both cell types and the upregulation of TAGLN and TAGLN2 only in iN cells. TAGLN was previously shown to regulate cytoskeleton organization, [101] and was differentially expressed in neuronal subpopulations of the rat central nervous system [102], highlighting this protein as an interesting target for further investigation. Furthermore, we provide a rich source of information of proteins associated with several signaling pathways, such as Wnt and Notch, involved in neuronal development. We identified several proteins associated to these signaling pathways being upregulated towards one of the two neuronal subtypes and being involved in cell fate pattering and development, further emphasizing their critical role in neuronal (subtype) differentiation. Also, we combined our proteomics data with comparable transcriptomic data during neuronal differentiation showing similarities as well as differences for several transcripts/proteins. These variations might be regulated by additional mechanisms such as posttranscriptional modifications that control the expression level of several proteins. In summary, we provide a quantitative view of key proteins that promote the loss of pluripotency and rapid neuronal development along the course of differentiation. This data constitute a rich resource that can be used in the future to better understand specific molecular mechanisms involved in neurodevelopment.

MATERIAL AND METHODS

Cell culture

iPSC generation

Generation of iPSC was performed using a previously established protocol [103]. Briefly, skin biopsies from healthy individuals were taken and fibroblast were maintained in mouse embryonic fibroblast (MEF) medium containing DMEM GlutaMAX (Life Technologies), 10% fetal bovine serum (Sigma Aldrich) and 1% penicillin/streptomycin (Life Technologies). The iPSC were generated by lentiviral transduction expressing OCT4, KLF4, SOX2 and c-MYC in MEF medium containing 4 mg/mL hexadimethrine bromide (Sigma). After 24 h of incubation, cells were cultured in MEF medium for another 5 days. Subsequently, cells were detached with Trypsin-EDTA (Life Technologies) and cultured in a 10 cm dish containing irradiated MEFs in human embryonic stem cell (huES) medium containing DMEM-F12 (Life Technologies), knockout 10% serum replacement (Life Technologies), 1% penicillin/streptomucin (Life Technologies), 2% L-glutamine (Life Technologies), 0.1% β-mercaptoethanol (Meck Millipore) and 20 ng/mL recombinant human fibroblast growth factor-basic (Life Technologies). After 3 to 6 weeks, colonies were picked manually and maintained in huES medium on irradiated MEFs for another 3-6 weeks. The iPSC were passaged using Accutase (Innovative Cell Technologies) and cultured feeder-free on Getrex (Life Technologies) coated dishes in mTeSR1 medium (STEMCELL technologies).

Glial cells

Newborn mouse forebrains homogenates were incubated with trypsin for 5 min at 37 °C. Following harsh trituration, cells were plated in MEM Alpha medium containing diglucose 20%, FBS 10%, Pen-strep 1%. After three passages, glial cells were used for co-culture purposes.

Virus generation

The two lentivirus were produced as described [104]. HEK293T cells plated on a 500 cm^2 dish were co-transfected using MAXPEI solution with 35 µg pMD2.G, 65 µg psPAX and 100 µg pSIN- FUW-TeTO-Ngn2-P2A-EGFP-T2A-Puromycin or pSIN- FUW-M2rtTA in OPTI-MEM (Thermo Fisher Scientific). Six hours after transfection, the medium was replaced with OPTI-MEM supplemented with Penicillin-Streptomycin 1% (Thermo Fisher Scientific). The lentiviral particles were harvested 48 h after transfection, concentrated by tangential flow filtration using a 100 kDa cut-off spin filter (Millipore) and resuspended in phosphate buffered saline (PBS).

Differentiation

Generation of iN cells was performed using a slightly modified version of the protocol described in Yingsha Zhang *et al,* 2013 [19]. In short, on day 0, iPSC were treated with Accutase and plated at a density of 50 x 10^3 cells/well of a 24-wells plate on top of a Matrigel (BD Biosciences) coated coverslip in mTeSR containing Y27632 10 µM (Miltenyibiotec). On day 1, medium was changed to mTeSR supplemented with 2,5 µl lentivirus and polybrene 8 µg/µL (Sigma-Aldrich). On day 2, medium was replaced with N2/DMEM/F12/NEAA (Invitrogen) supplemented with BDNF 10 µg/L (R&D systems), human NT-3 10 µg/L (ReproTech), mouse laminin 0.2 mg/L (Invitrogen) and doxycycline 2g/L (Clontech). On day 3, puromycin 1mg/L (Sigma-Aldrich) was added to the cells. On day 4, coverslips with neurons was transferred on top of a 12-well plate cultured glial cells in Neurobasal medium containing B27 (ThermoFisher), Glutamax (Invitrogen), BDNF, NT-3, doxycycline, Ara-C 2 g/L (Sigma-Aldrich). Every other day thereafter, 50% of the medium was changed. At predetermined time points, cells were fixed for immunohistochemistry or for proteomics approaches.

Patterned MN differentiation was performed using a slightly modified version of previously described protocol in Yves Maury *et al,* Nature biotechnology,2015 [20]. Briefly, on day 0, iPSC were dissociated with Accutase and resuspended in differentiation medium containing DMEM F-12, Neurobasal *(v/v)*, N2 supplement (Life Technologies), B27 without vitamin A (Life Technologies), Pen-strep 1%, ascorbic acid (0.5 µM, Sigma-Aldrich) and 5 µM Y27632 (STemGent). Embryoid body (EB) formation was accompanied by a standardized microwell assay [105]. IPSC were seeded at a density of 150 cells/microwell in differentiation medium. Chir-99021 (Tocris), LDN193189, SB-431542, Smoothened agonist (SAG, Calbiochem), retinoic acid (RA, Sigma-Aldrich), DAPT (Tocris), BDNF (Peprotech) and GDNF (Peprotech) were added at indicated time points and concentrations. Medium was changed every other day. After 2-3 days, EBs were flushed out of the microwell and transferred to a non-adherent 10 cm petri dish (Greiner Bio-one). On day 15, EBs were dissociated into single cells using Papain (Worthington Biochemical Corporation) and DNAse (Worthington Biochemical Corporation). Cells were plated on PDL (20 µg/mL, Sigma-Aldrich) and laminin (5 µg/mL, Invitrogen) coated coverslips at 60-70 % confluency.

Immunohistochemistry

At the time points of interest, cells were fixed in 4% paraformaldehyde in PBS for 20 min at room temperature and washed with PBS. Cells were blocked for 45 min at room temperature in blocking solution (0.1% Tween in PBS containing 2% bovine serum albumin (BSA) and 20 % goat serum). Cells were washed three times in PBS and primary antibody was incubated for 1 h at room temperature. After

washing the cells three times, secondary antibody was applied and incubated for 1 h at room temperature in darkness. Primary and secondary antibodies were mixed in staining solution (0.1% Tween in PBS containing 5% goat serum). Cells were washed again and fixed with prolong gold and mounted on glass slides. The following commercial antibodies were used: rabbit anti-Tubbulin-β3 (Sigma), mouse anti-Isl-1 (DSHB), Hoechst or DAPI (invitrogen) rabbit-anti-FOXG1 (Abcam), mouse anti-TuJ1 (Covance). Alexa-488, and Alexa 546 conjugated secondary antibodies were obtained from Invitrogen.

Sample preparation for MS analysis

Samples were collected at indicated time points (10 time points) for MS analysis as follows. This sample size was sufficient to have statistical power. Cells were lysed in lysis buffer containing 8 M urea, 50 mM ammonium bicarbonate and one complete mini protease inhibitor (Roche). The lysate was sonicated on ice using Bioruptor-Diagenode and centrifuged at 2500 g for 10 min at 4 °C. Supernatant with the proteins was reduced with 4 mM dithiothreitol at 56 °C for 30 min and alkylated with 8 mM iodoacetamide at room temperature for 30 min in darkness. The lysate was enzymatically pre-digested with Lys-C (1:75; Wako) incubation for 4 h at 37 °C. The mixture was fourfold diluted with ammonium bicarbonate and digested with trypsin (1:100; Promega) at 37 °C. The sample was quenched by acidification with formic acid (FA, final concentration 10%) and peptides were desalted using a Sep-Pak C18 column (Waters). Peptides were dried in vacuum and resuspended in 50 mM triethyl ammonium bicarbonate at a final concentration of 5 mg/mL.

Tandem Mass Tag (TMT) 10-plex labelling

Aliquots of ~ 100 μg of each sample were chemically labeled with TMT reagents (Thermo Fisher) according to **Table 1**. Peptides were resuspended in 80 μl resuspension buffer containing 50 mM HEPES buffer and 12.5 % acetonitrile (ACN, pH 8.5). TMT reagents (0.8 mg) were dissolved in 80 μl anhydrous ACN of which 20 μl was added to the peptides. Following incubation at room temperature for 1 hour, the reaction was quenched using 5% hydroxylamine in HEPES buffer for 15 min at room temperature. The TMT-labeled samples were pooled at equal protein ratios followed by vacuum centrifuge to near dryness and desalting using Sep-Pak C18 cartridges.

Off-line basic pH fractionation

We fractionated and pooled samples using basic pH Reverse Phase HPLC. Samples were solubilized in buffer A (5% ACN, 10 mM ammonium bicarbonate, pH 8.0) and subjected to a 50 min linear gradient from 18% to 45% ACN in 10 mM ammonium bicarbonate pH 8 at flow rate of 0.8 ml/min. We used an Agilent 1100 pump equipped with a degasser and a photodiode array (PDA) detector and Agilent 300 Extend C18

column (5 µm particles, 4.6 mm i.d., and 20 cm in length). The peptide mixture was fractionated into 96 fractions and consolidated into 24. Samples were acidified with 10% formic acid and vacuum-dried followed by redissolving with 5% formic acid/5% ACN for LC-MS/MS processing.

Tabel 1:TMT-10plex label reagents each corresponding to a time point of differentiation.

Sample	Label Reagent
T1	TMT^{10}-126
T2	TMT^{10}-127N
T3	TMT^{10}-127C
T4	TMT^{10}-128N
T5	TMT^{10}-128C
T6	TMT^{10}-129N
T7	TMT^{10}-129C
T8	TMT^{10}-130N
T9	TMT^{10}-130C
T10	TMT^{10}-131

Mass spectrometry analysis

Mass spectrometry was performed on an Orbitrap Fusion (Thermo Fisher Scientific) and Orbitrap Fusion Lumos (Thermo Fisher Scientific) coupled to an Agilent 1290 HPLC system (Agilent Technologies). Peptides were separated on a double frit trap column of 20 mm x 100 µm inner diameter (ReproSil C18, Dr Maisch GmbH, Ammerbuch, Germany). This was followed by a 40 cm x 50 µm inner diameter analytical column (ReproSil Pur C18-AQ (Dr Maisch GmbH, Ammerbuch, Germany). Both columns were packed in-house. Trapping was done at 5 µl/min in 0.1 M acetic acid in H2O for 10 min and the analytical separation was done at 100 nl/min for 2 h by increasing the concentration of 0.1 M acetic acid in 80% acetonitrile *(v/v)*. The instrument was operated in a data-dependent mode to automatically switch between MS and MS^2 for patterned MNs or MS and MS^3 for iN cells. Full-scan MS spectra were acquired in the Orbitrap from m/z 350-1500 with a resolution of 60,000 FHMW, automatic gain control (AGC) target of 200,000 and maximum injection time of 50 ms. For the MS^2 analysis, the ten most intense precursors at a threshold above 5,000 were selected with an isolation window of 1.2 Th after accumulation to a target value of 30,000 (maximum injection time was 115 ms). Fragmentation was carried out using higher-energy collisional dissociation (HCD) with collision energy of 38% and activation time of 0.1 ms. Fragment ion analysis was performed on Orbitrap with resolution of 60,000 FHMW and a low mass cut-off setting of 120 Th. For the MS^3 analysis, first MS^2 analysis was performed with CID fragmentation on top 10 most intense ions with AGC target of 10 000 and isolation window of 0.7 Th, followed by MS^3 scan for each MS^2 scan with HCD fragmentation with 35% collision energy. The MS isolation window was set to 2 m/z and AGC target to 100,000 and maximum injection time was 120 ms.

Data processing

Mass spectra were processed using Proteome Discover (version 2.1, Thermo Scientific). Peak list was searched using Swissprot database (version 2014_08) with the search engine Mascot (version 2.3, Matrix Science). The following parameters were used. The enzyme was specified as trypsin and allowed up to two missed cleavages. Taxonomy was chosen for Homo sapiens and precursor mass tolerance was set to 50 ppm with 0.05 Da fragment mass tolerance for MS^2 analysis or 0.6 Da for MS^3 analysis. TMT tags on lysine residues and peptide N termini and oxidation of methionine residues were set as dynamic modifications, while carbamidomethylation on cysteine residues was set as static modification. For the reporter ion quantification, integration tolerance was set to 20 ppm with the most confident centroid method. Results were filtered with a Mascot score of at least 20 and Percolator was used to adjust the peptide-spectrum matches (PSMs) to a false discovery rate (FDR) below 1%. Finally, peptides lower than 6 amino-acid residues were discarded.

Data visualization

The software Perseus was used to generate the plots, heatmaps and to calculate the Pearson correlation. For quantitative analysis, the TMT reporter intensity values of proteins in each time point was normalized to the reference intensity value of T1 (reference), within each biological replicate. The ratios of the biological replicates were then log_2 transformed. All the peptide ratios were normalized against the median. Volcano plots for each time point was generated and up- or downregulated proteins were considered significant with FDR < 0.05. Z scores were used to generate heatmaps. Functional analysis to enrich to GO, MF and CO terms were done using Gprox open source software package [106]. Clustering parameters such as Fuzzification, regulation threshold and number of clusters were set to 2.00, upper limit =-0.58 lower limit = 0.58 and 3, respectively. For the identification of time specific proteins between iN and patterned MNs, volcano plots was generated comparing the two groups for each time point and proteins were considered to be significant with a fold change cut-off ≥ 1.5 and a p value ≤ 0.1. Furthermore, protein classification was performed using PANTHER [107] classification system and GO analysis and classification of transcription factors, cytoskeletal proteins were done using DAVID database. In addition, Reactome pathway database analysis was used to identify pathways enriched in several clusters.

Data availability

All mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD013399

ACKNOWLEDGEMENTS

REFERENCES

1. Herculano-Houzel, S., *The human brain in numbers: a linearly scaled-up primate brain.* Front Hum Neurosci, 2009. **3**: p. 31.
2. Rockland, K.S., *Non-uniformity of extrinsic connections and columnar organization.* J Neurocytol, 2002. **31**(3-5): p. 247-53.
3. Toro, R., et al., *Key role for gene dosage and synaptic homeostasis in autism spectrum disorders.* Trends Genet, 2010. **26**(8): p. 363-72.
4. Hall, J., et al., *Associative learning and the genetics of schizophrenia.* Trends Neurosci, 2009. **32**(6): p. 359-65.
5. Harvey, M., P. Belleau, and N. Barden, *Gene interactions in depression: pathways out of darkness.* Trends Genet, 2007. **23**(11): p. 547-56.
6. Willsey, A.J., et al., *Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism.* Cell, 2013. **155**(5): p. 997-1007.
7. Parikshak, N.N., et al., *Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism.* Cell, 2013. **155**(5): p. 1008-21.
8. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
9. Han, S.S., L.A. Williams, and K.C. Eggan, *Constructing and deconstructing stem cell models of neurological disease.* Neuron, 2011. **70**(4): p. 626-44.
10. Marchetto, M.C. and F.H. Gage, *Modeling brain disease in a dish: really?* Cell Stem Cell, 2012. **10**(6): p. 642-5.
11. Tao, Y. and S.C. Zhang, *Neural Subtype Specification from Human Pluripotent Stem Cells.* Cell Stem Cell, 2016. **19**(5): p. 573-586.
12. Bertrand, N., D.S. Castro, and F. Guillemot, *Proneural genes and the specification of neural cell types.* Nat Rev Neurosci, 2002. **3**(7): p. 517-30.
13. Kelava, I. and M.A. Lancaster, *Stem Cell Models of Human Brain Development.* Cell Stem Cell, 2016. **18**(6): p. 736-48.
14. Busskamp, V., et al., *Rapid neurogenesis through transcriptional activation in human stem cells.* Mol Syst Biol, 2014. **10**: p. 760.
15. Hoffmann, A., M. Ziller, and D. Spengler, *Childhood-Onset Schizophrenia: Insights from Induced Pluripotent Stem Cells.* Int J Mol Sci, 2018. **19**(12).
16. Taoufik, E., et al., *Synaptic dysfunction in neurodegenerative and neurodevelopmental diseases: an overview of induced pluripotent stem-cell-based disease models.* Open Biol, 2018. **8**(9).
17. Guo, W., et al., *Current Advances and Limitations in Modeling ALS/FTD in a Dish Using Induced Pluripotent Stem Cells.* Front Neurosci, 2017. **11**: p. 671.
18. Sances, S., et al., *Modeling ALS with motor neurons derived from human induced pluripotent stem cells.* Nat Neurosci, 2016. **19**(4): p. 542-53.

19.   Zhang, Y., et al., *Rapid single-step induction of functional neurons from human pluripotent stem cells.* Neuron, 2013. **78**(5): p. 785-98.

20.   Maury, Y., et al., *Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes.* Nat Biotechnol, 2015. **33**(1): p. 89-96.

21.   Bielle, F., et al., *Multiple origins of Cajal-Retzius cells at the borders of the developing pallium.* Nat Neurosci, 2005. **8**(8): p. 1002-12.

22.   Rosskothen-Kuhl, N. and R.B. Illing, *Gap43 transcription modulation in the adult brain depends on sensory activity and synaptic cooperation.* PLoS One, 2014. **9**(3): p. e92624.

23.   Hoffman, P.N. and R.J. Lasek, *The slow component of axonal transport. Identification of major structural polypeptides of the axon and their generality among mammalian neurons.* J Cell Biol, 1975. **66**(2): p. 351-66.

24.   Lee, H.K., et al., *Dynamic Ca2+-dependent stimulation of vesicle fusion by membrane-anchored synaptotagmin 1.* Science, 2010. **328**(5979): p. 760-3.

25.   Graham, V., et al., *SOX2 functions to maintain neural progenitor identity.* Neuron, 2003. **39**(5): p. 749-65.

26.   Zeineddine, D., et al., *The Oct4 protein: more than a magic stemness marker.* Am J Stem Cells, 2014. **3**(2): p. 74-82.

27.   Mazzoni, E.O., et al., *Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity.* Nat Neurosci, 2013. **16**(9): p. 1219-27.

28.   Li, X., et al., *Small-Molecule-Driven Direct Reprogramming of Mouse Fibroblasts into Functional Neurons.* Cell Stem Cell, 2015. **17**(2): p. 195-203.

29.   Vincent, A.J., P.W. Lau, and A.J. Roskams, *SPARC is expressed by macroglia and microglia in the developing and mature nervous system.* Dev Dyn, 2008. **237**(5): p. 1449-62.

30.   Zhou, S., et al., *Survivin Improves Reprogramming Efficiency of Human Neural Progenitors by Single Molecule OCT4.* Stem cells international, 2016. **2016**: p. 4729535-4729535.

31.   Gongidi, V., et al., *SPARC-like 1 regulates the terminal phase of radial glia-guided migration in the cerebral cortex.* Neuron, 2004. **41**(1): p. 57-69.

32.   Harkin, L.F., et al., *Distinct expression patterns for type II topoisomerases IIA and IIB in the early foetal human telencephalon.* J Anat, 2016. **228**(3): p. 452-63.

33.   Schuurmans, C., et al., *Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways.* Embo j, 2004. **23**(14): p. 2892-902.

34.   Hansson, J., et al., *Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency.* Cell Rep, 2012. **2**(6): p. 1579-92.

35.   Frese, C.K., et al., *Quantitative Map of Proteome Dynamics during Neuronal Differentiation.* Cell Rep, 2017. **18**(6): p. 1527-1542.

36.   Liao, M.-L., et al., *Distribution patterns of the zebrafish neuronal intermediate filaments inaa and inab.* Journal of Neuroscience Research, 2019. **97**(2): p. 202-214.

37.   Quach, T.T., et al., *Collapsin response mediator protein 3 increases the dendritic arborization of hippocampal neurons.* Mol Psychiatry, 2015. **20**(9): p. 1027.

38.   da Rocha Boeira, T., et al., *Polymorphism Located in the Upstream Region of the RPS19 Gene (rs2305809) Is Associated With Cervical Cancer: A Case-control Study.* J Cancer Prev, 2018. **23**(3): p. 147-152.

39.   Moore, B.W., *A soluble protein characteristic of the nervous system.* Biochem Biophys Res Commun, 1965. **19**(6): p. 739-44.

40.   Filipek, A., et al., *Calcyclin--Ca(2+)-binding protein homologous to glial S-100 beta is present in neurones.* Neuroreport, 1993. **4**(4): p. 383-6.

41.   Yamashita, N., et al., *Distribution of a specific calcium-binding protein of the S100 protein family, S100A6 (calcyclin), in subpopulations of neurons and glial cells of the adult rat nervous system.* J Comp Neurol, 1999. **404**(2): p. 235-57.

42.   Chan, W.Y., et al., *Differential expression of S100 proteins in the developing human hippocampus and temporal cortex.* Microsc Res Tech, 2003. **60**(6): p. 600-13.

43.   Girard, F., et al., *The EF-hand Ca(2+)-binding protein super-family: a genome-wide analysis of gene expression patterns in the adult mouse brain.* Neuroscience, 2015. **294**: p. 116-55.

44.   Hupe, M., et al., *Gene expression profiles of brain endothelial cells during embryonic development at bulk and single-cell levels.* Sci Signal, 2017. **10**(487).

45.   Chantha, S.C., et al., *The MIDASIN and NOTCHLESS genes are essential for female gametophyte development in Arabidopsis thaliana.* Physiol Mol Biol Plants, 2010. **16**(1): p. 3-18.

46.   Kalus, I., et al., *Sulf1 and Sulf2 Differentially Modulate Heparan Sulfate Proteoglycan Sulfation during Postnatal Cerebellum Development: Evidence for Neuroprotective and Neurite Outgrowth Promoting Functions.* PLoS One, 2015. **10**(10): p. e0139853.

47.   Kalus, I., et al., *Differential involvement of the extracellular 6-O-endosulfatases Sulf1 and Sulf2 in brain development and neuronal and behavioural plasticity.* J Cell Mol Med, 2009. **13**(11-12): p. 4505-21.

48.   Briscoe, J. and J. Ericson, *Specification of neuronal fates in the ventral neural tube.* Current Opinion in Neurobiology, 2001. **11**(1): p. 43-49.

49.   Couillard-Despres, S., et al., *Doublecortin expression levels in adult brain reflect neurogenesis.* Eur J Neurosci, 2005. **21**(1): p. 1-14.

50.   Kim, D.H., et al., *Pan-neuronal calcium imaging with cellular resolution in freely swimming zebrafish.* Nature Methods, 2017. **14**: p. 1107.

51.   Abbott, N.J., et al., *Structure and function of the blood–brain barrier.* Neurobiology of Disease, 2010. **37**(1): p. 13-25.

52.   Zhang, Y. and L. Niswander, *Zic2 is required for enteric nervous system development and neurite outgrowth: a mouse model of enteric hyperplasia and dysplasia.* Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society, 2013. **25**(6): p. 538-541.

53.   Morrison, S.J., *Neuronal differentiation: proneural genes inhibit gliogenesis.* Curr Biol, 2001. **11**(9): p. R349-51.

54.   Serre, A., et al., *Overexpression of basic helix-loop-helix transcription factors enhances neuronal differentiation of fetal human neural progenitor cells in various ways.* Stem Cells Dev, 2012. **21**(4): p. 539-53.

55.   Ladewig, J., et al., *Small molecules enable highly efficient neuronal conversion of human fibroblasts.* Nat Methods, 2012. **9**(6): p. 575-8.

56.   Conkright, M.D., et al., *TORCs: transducers of regulated CREB activity.* Mol Cell, 2003. **12**(2): p. 413-23.

57.   Li, S., et al., *TORC1 regulates activity-dependent CREB-target gene transcription and dendritic growth of developing cortical neurons.* J Neurosci, 2009. **29**(8): p. 2334-43.

58.   Mendioroz, M., et al., *CRTC1 gene is differentially methylated in the human hippocampus in Alzheimer's disease.* Alzheimers Res Ther, 2016. **8**(1): p. 15.

59.   Storchel, P.H., et al., *A large-scale functional screen identifies Nova1 and Ncoa3 as regulators of neuronal miRNA function.* Embo j, 2015. **34**(17): p. 2237-54.

60.   Casari, A., et al., *A Smad3 transgenic reporter reveals TGF-beta control of zebrafish spinal cord development.* Developmental Biology, 2014. **396**(1): p. 81-93.

61.   Inoue, T., et al., *Mouse Zic5 deficiency results in neural tube defects and hypoplasia of cephalic neural crest derivatives.* Dev Biol, 2004. **270**(1): p. 146-62.

62.   Quadrato, G., et al., *Nuclear factor of activated T cells (NFATc4) is required for BDNF-dependent survival of adult-born neurons and spatial memory formation in the hippocampus.* Proceedings of the National Academy of Sciences, 2012. **109**(23): p. E1499-E1508.

63.   Mann, S.S. and J.A. Hammarback, *Gene localization and developmental expression of light chain 3: a common subunit of microtubule-associated protein 1A(MAP1A) and MAP1B.* J Neurosci Res, 1996. **43**(5): p. 535-44.

64.   Sullivan, K.F. and D.W. Cleveland, *Identification of conserved isotype-defining variable region sequences for four vertebrate beta tubulin polypeptide classes.* Proc Natl Acad Sci U S A, 1986. **83**(12): p. 4327-31.

65.   Mori, K., et al., *Neuronal protein NP25 interacts with F-actin.* Neurosci Res, 2004. **48**(4): p. 439-46.

66.   Stiles, J. and T.L. Jernigan, *The basics of brain development.* Neuropsychology review, 2010. **20**(4): p. 327-348.

67.   Tao, Y. and S.-C. Zhang, *Neural Subtype Specification from Human Pluripotent Stem Cells.* Cell stem cell, 2016. **19**(5): p. 573-586.

68.   Chambers, S.M., et al., *Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling.* Nat Biotechnol, 2009. **27**(3): p. 275-80.

69.   Martin, B.L. and D. Kimelman, *Wnt signaling and the evolution of embryonic posterior development.* Curr Biol, 2009. **19**(5): p. R215-9.

70.   Ciani, L. and P.C. Salinas, *WNTs in the vertebrate nervous system: from patterning to neuronal connectivity.* Nat Rev Neurosci, 2005. **6**(5): p. 351-62.

71.   Budnik, V. and P.C. Salinas, *Wnt signaling during synaptic development and plasticity.* Curr Opin Neurobiol, 2011. **21**(1): p. 151-9.

72.   Hirabayashi, Y., et al., *The Wnt/beta-catenin pathway directs neuronal differentiation of cortical neural precursor cells.* Development, 2004. **131**(12): p. 2791-801.

73.   Nordstrom, U., et al., *An early role for WNT signaling in specifying neural patterns of Cdx and Hox gene expression and motor neuron subtype identity.* PLoS Biol, 2006. **4**(8): p. e252.

74.   Bhatt, P.M. and R. Malgor, *Wnt5a: a player in the pathogenesis of atherosclerosis and other inflammatory disorders.* Atherosclerosis, 2014. **237**(1): p. 155-62.

75.   Zhu, X., et al., *Wls-mediated Wnts differentially regulate distal limb patterning and tissue morphogenesis.* Dev Biol, 2012. **365**(2): p. 328-38.

76.   Barna, M., P.P. Pandolfi, and L. Niswander, *Gli3 and Plzf cooperate in proximal limb patterning at early stages of limb development.* Nature, 2005. **436**(7048): p. 277-81.

77.   Serfling, E., et al., *NFATc1 autoregulation: A crucial step for cell-fate determination*. Vol. 27. 2006. 461-9.

78.   Yang, Y.R., et al., *Primary phospholipase C and brain disorders.* Adv Biol Regul, 2016. **61**: p. 80-5.

79.   Lara-Ramirez, R., E. Zieger, and M. Schubert, *Retinoic acid signaling in spinal cord development.* Int J Biochem Cell Biol, 2013. **45**(7): p. 1302-13.

80.   Cayuso, J., et al., *The Sonic hedgehog pathway independently controls the patterning, proliferation and survival of neuroepithelial cells by regulating Gli activity.* Development, 2006. **133**(3): p. 517-28.

81.   Numata-Uematsu, Y., et al., *Inhibition of collapsin response mediator protein-2 phosphorylation ameliorates motor phenotype of ALS model mice expressing SOD1G93A.* Neuroscience Research, 2018.

82.   Kowalczyk, A., et al., *The critical role of cyclin D2 in adult neurogenesis.* J Cell Biol, 2004. **167**(2): p. 209-13.

83.   Chen, R., et al., *Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity.* Cell Reports, 2017. **18**(13): p. 3227-3241.

84.   Papalopulu, N. and C. Kintner, *A posteriorising factor, retinoic acid, reveals that anteroposterior patterning controls the timing of neuronal differentiation in Xenopus neuroectoderm.* Development, 1996. **122**(11): p. 3409-18.

85.   Elkabetz, Y., et al., *Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage.* Genes Dev, 2008. **22**(2): p. 152-65.

86.   Borghese, L., et al., *Inhibition of notch signaling in human embryonic stem cell-derived neural stem cells delays G1/S phase transition and accelerates neuronal differentiation in vitro and in vivo.* Stem Cells, 2010. **28**(5): p. 955-64.

87.   Adams, J.C. and F.M. Watt, *Regulation of development and differentiation by the extracellular matrix.* Development, 1993. **117**(4): p. 1183-98.

88.   Christopherson, K.S., et al., *Thrombospondins Are Astrocyte-Secreted Proteins that Promote CNS Synaptogenesis.* Cell, 2005. **120**(3): p. 421-433.

89.   Rasmussen, J.P. and A. Sagasti, *Learning to swim, again: Axon regeneration in fish.* Exp Neurol, 2017. **287**(Pt 3): p. 318-330.

90.   Dickson, B.J., *Molecular Mechanisms of Axon Guidance.* Science, 2002. **298**(5600): p. 1959-1964.

91.   Bashaw, G.J., et al., *Repulsive axon guidance: Abelson and Enabled play opposing roles downstream of the roundabout receptor.* Cell, 2000. **101**(7): p. 703-15.

92.   Desmond, J.C., et al., *The Aldo-Keto Reductase AKR1C3 Is a Novel Suppressor of Cell Differentiation That Provides a Plausible Target for the Non-Cyclooxygenase-dependent Antineoplastic Actions of Nonsteroidal Anti-Inflammatory Drugs.* Cancer Research, 2003. **63**(2): p. 505-512.

93.   Capo-Chichi, C.D., et al., *Alteration of Differentiation Potentials by Modulating GATA Transcription Factors in Murine Embryonic Stem Cells.* Stem cells international, 2010. **2010**: p. 602068-602068.

94.   Hoerder-Suabedissen, A., et al., *Expression profiling of mouse subplate reveals a dynamic gene network and disease association with autism and schizophrenia.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(9): p. 3555-3560.

95.   Fu, X., et al., *Doublecortin (Dcx) family proteins regulate filamentous actin structure in developing neurons.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 2013. **33**(2): p. 709-721.

96.   Ince-Dunn, G., et al., *Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability.* Neuron, 2012. **75**(6): p. 1067-1080.

97.   Fleming, M.S., et al., *Cis and trans RET signaling control the survival and central projection growth of rapidly adapting mechanoreceptors.* eLife, 2015. **4**: p. e06828-e06828.

98.   Jones, I., et al., *Development and validation of an in vitro model system to study peripheral sensory neuron development and injury.* Scientific reports, 2018. **8**(1): p. 15961-15961.

99.   Lin, M., et al., *RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders.* PLoS One, 2011. **6**(9): p. e23356.

100.   Neve, R.L., et al., *Identification of cDNA clones for the human microtubule-associated protein tau and chromosomal localization of the genes for tau and microtubule-associated protein 2.* Brain Res, 1986. **387**(3): p. 271-80.

101.   Elsafadi, M., et al., *Transgelin is a TGFß-inducible gene that regulates osteoblastic and adipogenic differentiation of human skeletal stem cells through actin cytoskeleston organization.* Cell death & disease, 2016. **7**(8): p. e2321-e2321.

102.   Ren, W.-Z., et al., *The identification of NP25: a novel protein that is differentially expressed by neuronal subpopulations.* Molecular Brain Research, 1994. **22**(1): p. 173-185.

103.   Harschnitz, O., et al., *Autoantibody pathogenicity in a multifocal motor neuropathy induced pluripotent stem cell-derived model.* Ann Neurol, 2016. **80**(1): p. 71-88.

104.   Naldini, L., et al., *In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector.* Science, 1996. **272**(5259): p. 263-7.

105.   Rivron, N.C., et al., *Tissue deformation spatially modulates VEGF signaling and angiogenesis.* Proc Natl Acad Sci U S A, 2012. **109**(18): p. 6886-91.

106.   Rigbolt, K.T., J.T. Vanselow, and B. Blagoev, *GProX, a user-friendly platform for bioinformatics analysis and visualization of quantitative proteomics data.* Mol Cell Proteomics, 2011. **10**(8): p. O110.007450.

107.   Mi, H., et al., *PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways.* Nucleic Acids Res, 2007. **35**(Database issue): p. D247-52.
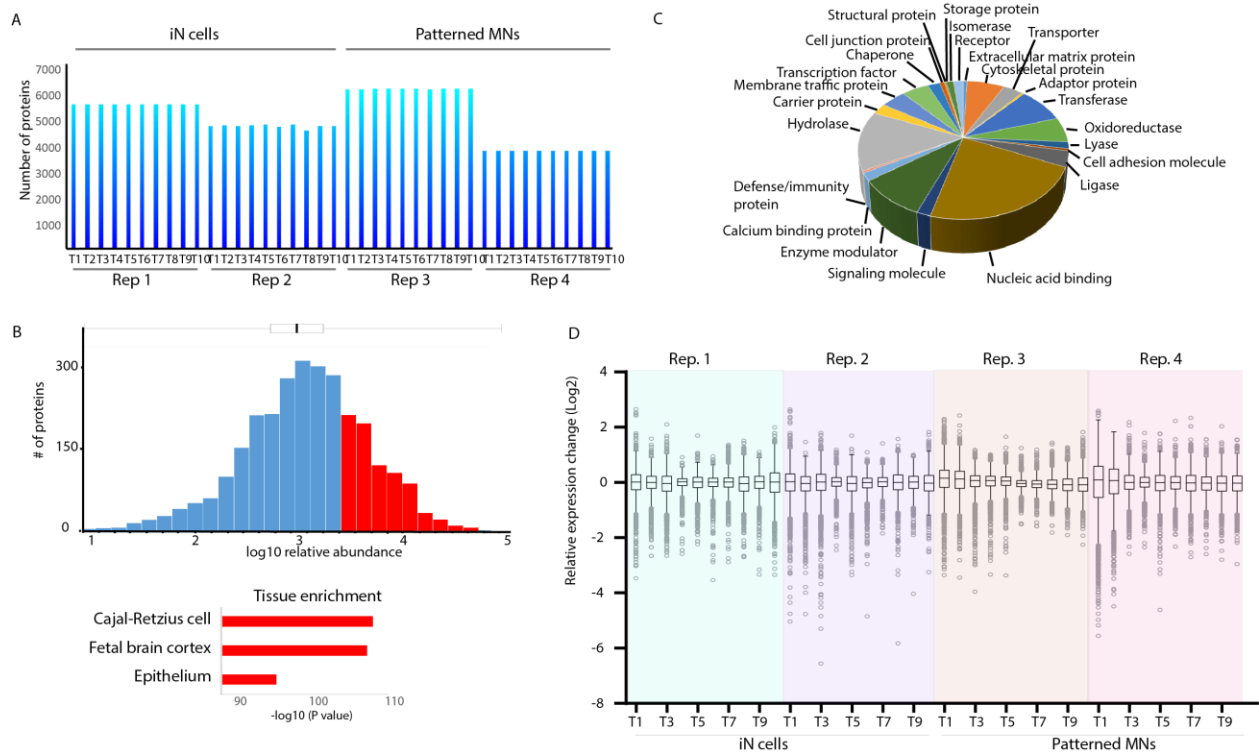
Supplementary data

Figure S1. Global proteome analysis. (A) Bar graph showing the number of proteins identified in each biological replicate and in each time point. (B) Distribution of quantified protein abundances in all biological replicas and all time points, spanning four orders of magnitude. In red tissue enrichment analysis using DAVID, of the 25% most abundant proteins. (C). Pie charts display the distribution of protein classes, based on PANTHER classification from all quantified proteins. (D) Box plots of log2 transformed data normalized on protein peak areas of each individual biological replicate
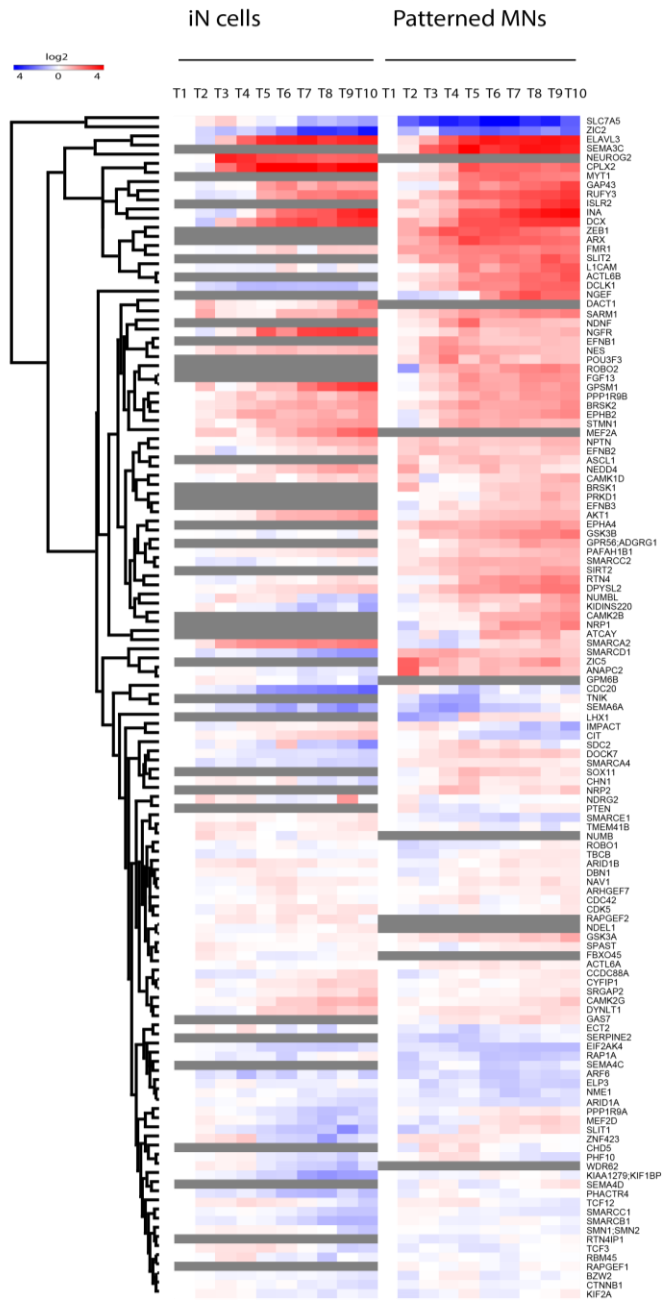
Figure S2. Neurogenesis. Neurogenesis associated proteins enriched using the Database for Annotation, Visualization, and Integrated Discovery (DAVID).
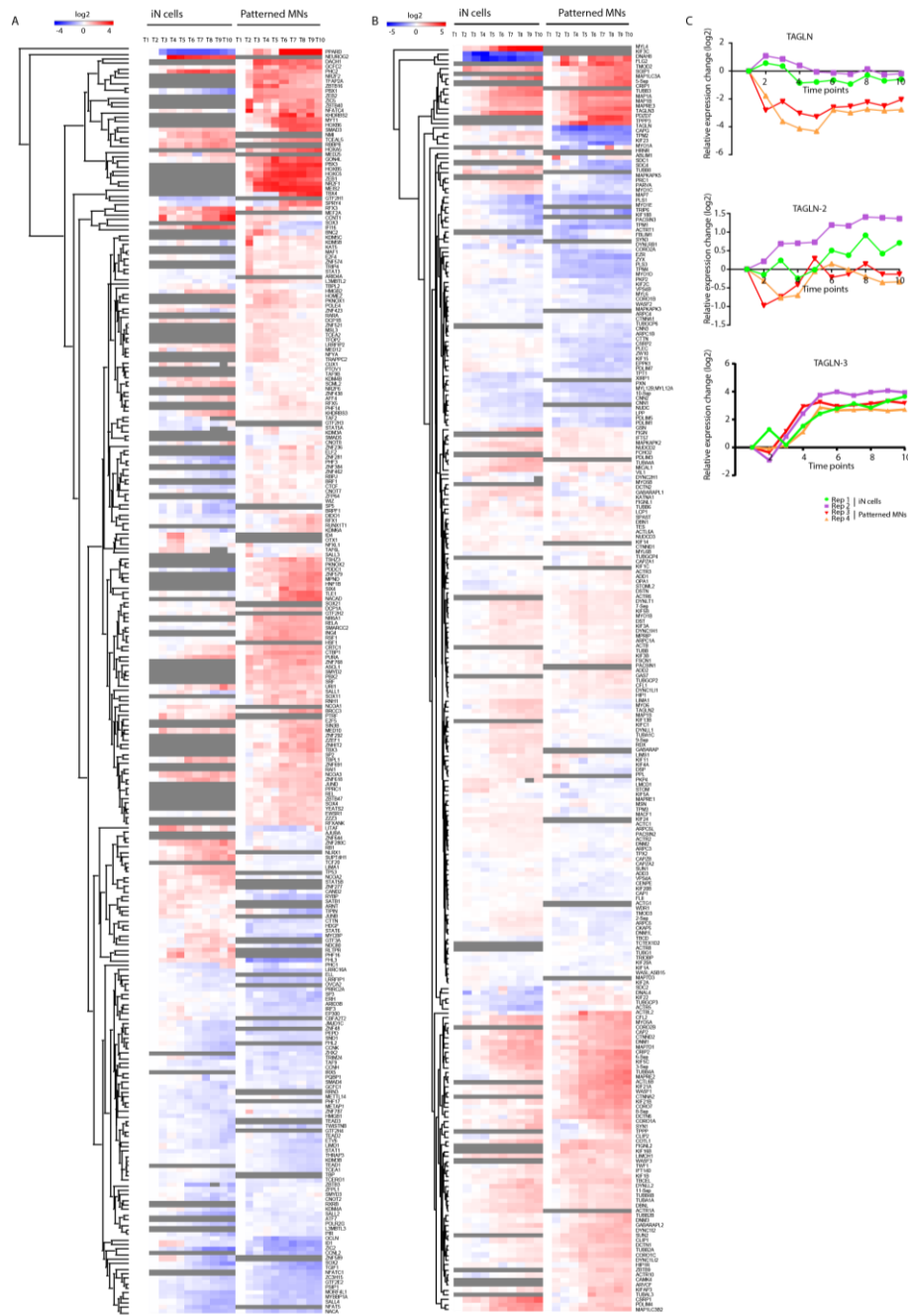
Figure S3. Transcription factors and cytoskeletal proteins. (A) Transcrioption factors enriched using the Database for Annotation, Visualization, and Integrated Discovery (DAVID). (B) Cytoskeletal proteins enriched using DAVID. (C) Example of the Transgelin family that undergo different expression during differentiation towards iN and patterned MNs.
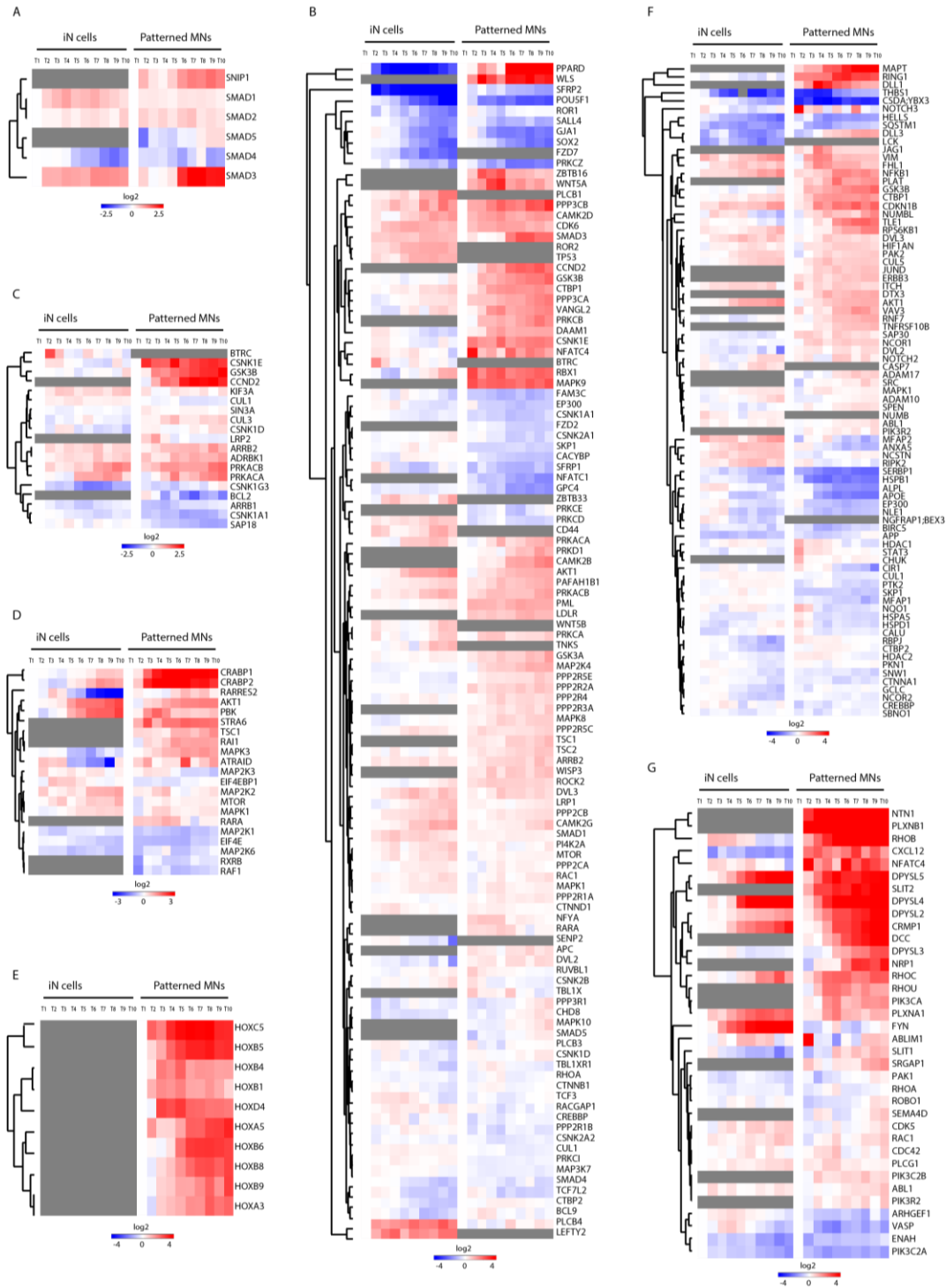
Figure S4. Signaling pathways. (A). Proteins related to TGFβ signaling enriched from KEGG and WIKI Pathways. (B). Proteins related to Wnt signaling. (C). Proteins related to Hedgehog signaling. (D). Proteins related to RA signaling. (E). Homeobox proteins all expressed in patterned MNs and none of the different members expressed in iN cells. (F). Proteins related to Notch signaling. (G). Proteins related to axon guidance
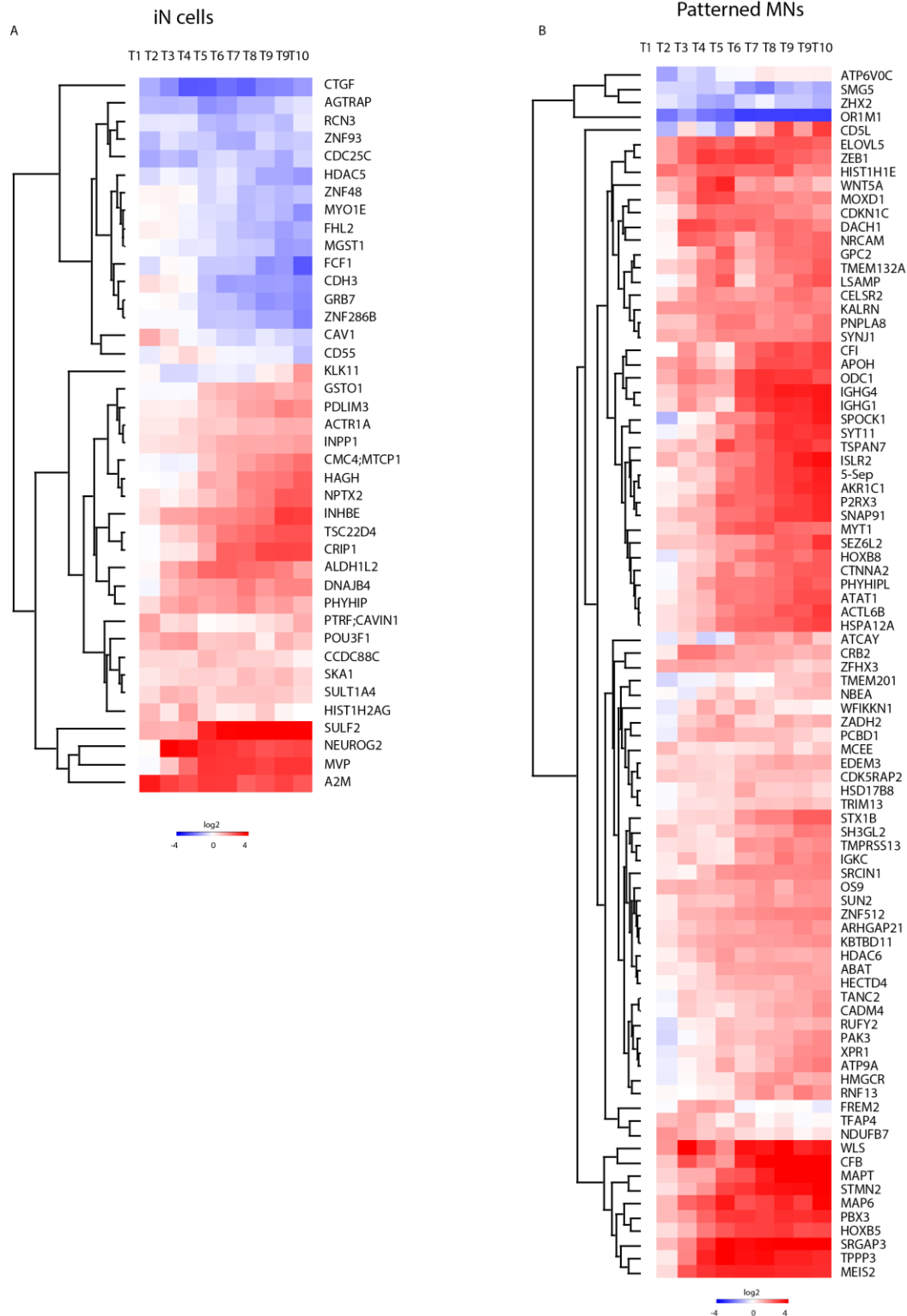
Figure S5. Proteins only identified in iN cells or patterned MNs. Proteins exclusively identified in (A) iN cells or (B) patterned MN differentiation. Proteins were selected for a twofold change in the two biological replicates in each group.

Quantitative proteomic alterations of human iPSC-based neuronal development indicate early onset of Rett syndrome.

Suzy Varderidou-Minasian[1, 2]‡, Lisa Hinz[3]‡, Dominique Hagemans[1, 2], Danielle Posthuma[3, 4], Maarten Altelaar[1, 2]‡, Vivi M. Heine[3, 5]‡*

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

[2] *Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, the Netherlands.*

[3] *Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, the Netherlands.*

[4] *Clinical Genetics, Amsterdam UMC, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands*

[5] *Pediatric Neurology, Emma Children's Hospital, Amsterdam UMC, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. vm.heine@vumc.nl.*

‡ These authors contributed equally to this work.
*Correspondence: vm.heine@vumc.nl

Abstract

Background

Rett syndrome (RTT) is a progressive neurodevelopmental disease that is characterized by abnormalities in cognitive, social and motor skills. RTT is often caused by mutations in the X-linked gene encoding methyl-CpG binding protein 2 (MeCP2). The mechanisms by which impaired MeCP2 induces the pathological abnormalities in the brain is not understood. Both patients and mouse models have shown abnormalities at molecular and cellular level before typical RTT-associated symptoms appear. This implies that underlying mechanisms are already affected during neurodevelopmental stages.

Methods

To understand the molecular mechanisms involved in disease onset, we used an RTT patient induced pluripotent stem cell (iPSC)-based model with isogenic controls and performed time-series of proteomic analysis using in-depth high-resolution quantitative mass spectrometry during early stages of neuronal development.

Results

We provide mass spectrometry-based quantitative proteomic data, depth of about 7000 proteins, at neuronal developmental stages of RTT patient cells and isogenic controls. Our data provided evidence of proteomic alteration at developmental stages long before the phase that symptoms of RTT syndrome become apparent. Changes in dendrite morphology or synaptic defects, become apparent at early developmental stages. Differential expression increased from early to late neural stem cell phases, although proteins involved in immunity, metabolic processes and calcium signaling were already affected at initial stages.

Limitations

The limitation of our study is the number of biological replicates (n=3). As the aim of our study was to investigate a large number of proteins, only limited amount of biological replicates were suitable to study as for larger scales a reduction of target proteins would be needed.

Conclusions

Our results can serve as valuable resource of target proteins that could be used as potential targets for early treatments to reduce the progression of RTT symptoms. We found global and time-point specific alterations during early neuronal differentiation in RTT cultures. Insight into altered protein levels can help development of new biomarkers and therapeutic approaches in RTT syndrome. Therefore, we hope that our results give awareness of the early pre-natal onset of RTT and emphasize the need of early testing followed by an effective treatment.

Keywords

Rett syndrome; iPSC; neuron differentiation; quantitative mass spectrometry; TMT-10plex

Background

Rett syndrome (RTT) is a severe neurodevelopmental disorder that mainly affects females with a frequency of ~1:10,000 [1]. Clinical features of RTT start presenting themselves around 6-18 months of age, and include deceleration of head growth, abnormalities in cognitive, social and motor skill development and seizures [2, 3]. Postmortem studies showed that RTT patient brains present with an increased density of neurons in combination with reduced soma sizes, compared to healthy control brains [4, 5]. RTT neurons further show a decrease in dendritic branching, and a reduced number of dendritic spines and synapses [6, 7], typical for an immature phenotype, therefore categorizing RTT as neurodevelopmental disorder [1]. The molecular mechanisms underlying the onset of neuronal development in RTT is however unknown.

In 90-95% of RTT cases, the disease is caused by a loss-of-function mutation in the X-linked gene encoding methyl-CpG binding protein 2 (MeCP2) [8]. MeCP2 mutations predominantly occur on the X chromosome, which upon random X chromosome inactivation in females results in somatic mosaics with normal and mutant MeCP2 [9]. Males carrying a MeCP2 mutation are not viable or suffer from severe symptoms and die early in life [10]. MeCP2 was first described as a nuclear protein modulating gene expression, via binding to methylated DNA and hundreds of target genes. These modulations take place through direct repression or activation of genes, or by means of DNA modulation and secondary gene regulation. Consequently, mutations in MeCP2 lead to miss-regulation of hundreds of genes, including those influencing brain development and neuronal maturation [11-14]. So far research in RTT focused on genomic and transcriptomic studies [15-17] and less so on proteome changes [18, 19]. Recent advances in mass spectrometry-based proteomics now facilitate the study of global protein expression and

quantification [20]. Considering the broad and complex regulating functions of MeCP2, modulating multiple cellular processes, we need insight into the final molecular effectors reflected by perturbation at the protein level to understand pathological states.

Here we used an iPSC-based RTT model and performed proteome analysis on iPSC-derived neuronal stem cells (NES cells) [21]. Earlier studies proved that iPSCs from RTT patients reflect disease-specific characteristics, including changes in neuronal differentiation at early stages of development [22, 23]. However, we lack knowledge on the precise molecular mechanisms underlying the progression of the disease. To study early alterations in the proteome of RTT cells compared to isogenic controls (iCTR), we performed a high-resolution mass spectrometry-based quantitative proteomics at different time points during neuronal stem cell development (**Fig. 1**). Interestingly, proteins differentially expressed in RTT versus iCTR are involved in cellular processes, implicated in classical features of typical RTT phenotypes, such as dendrite formation and axonal growth. We show that the difference between RTT and iCTR, in terms of the number of differentially expressed proteins, already begins at early stages and increases at later neural stem cell stages. We also show that a specific group of proteins involved in immunity and metabolic processes are differentially expressed in RTT at all time points studied. Here we provide evidence of target proteins that could be used as potential targets for early treatments to reduce the progression of RTT symptoms.

Methods

Cell culture and isogenic controls

RTT patient fibroblasts were derived from the *Cell lines and DNA bank of Rett syndrome, X-linked mental retardation and other genetic diseases* at the University Siena in Italy via the Network of Genetic Biobanks Telethon. Three fibroblast lines carrying different MECP2 mutations were selected. RTT#2282C2 showing a deletion in Exon 3 and 4 of the MeCP2 gene (RTT Ex3-4), RTT#2204 carrying a missense mutation within the methyl binding domain (RTT T158M) and RTT#2238 having a nonsense mutation in the transcription repression domain (RTT R255X). Fibroblasts were derived frozen and thawed and expanded in in fibroblast medium (DMEM-F12, 20% FBS, 1%NEAA, 1%Pen/Strep, 50 µM β-Mercaptoethanol). To generate pure RTT fibroblast lines and isogenic controls cells were detached from cell culture plate and single fibroblasts were seeded in wells of a 96-well plate. Cells were further expanded and characterized on their MeCP2 state by immunocytochemistry and PCR [24]. All of our experiments were exempt from the approval of the institutional review board.

Reprogramming

Reprogramming of fibroblasts was performed as described before [24]. In brief, fibroblasts were detached from cell culture plate and washed with PBS. $4x10^5$ cells were resuspended in 400 µl Gene Pulser® Electroporation Buffer Reagent (BioRad) with 23,4 µg of each episomal plasmid (Addgene, Plasdmid #27078, #27080, #27076) containing the reprogramming factors OCT4, SOX2, KLF4 and C-MYC. Cell solution was carefully mixed and electroporated with three pulses of 1.6 kV, capacitance of 3 µF and a resistance of 400 Ω (Gene Pulser II (BioRad)). Fibroblasts were left for recovery in Fibroblast medium without antibiotics containing 10 µM Rock inhibitor (Y-27632). After cells reached a confluence of 60-70%, medium was changed to TeSR™-E7™ (STEMCELL). Colonies appeared after 21-28 days. They were picked manually and maintained in TeSR™-E8™ (STEMCELL).

Differentiation of neuronal stem cells

In total, 8 iPSC lines were differentiated towards neuronal stem cells as described before [21, 57]. Four RTT lines and four isogenic controls from the same individual therefore, were plated in high-density on Geltrex®-coated wells of a 12-well plate in TeSR™-E8™ with 10 µM Rock inhibitor. Medium was changed daily for 2 days. Afterwards half of the medium was changed daily to Neuro-Maintenance-Medium (NMM) (1:1 DMEM/F12+GlutaMAX:Neurobasal Medium, 1x B27, 1xN2, 2,5 µg/ml Insulin, 1,5 mM L-Glutamin, 100 µM NEAA, 50 µM 2-Mercaptoethanol, 1% penicillin/streptomycin) containing 1 µM Dorsomorphin and 10 µM SB431542 up to day 12. At day 10-12 rosette structures appeared, which were manually picked and further cultured on Poly-L-Ornithin (0,01%)/Laminin (20 µg/ml) coated cell culture plates in NMM medium containing EGF (20 ng/ml) and FGF-2 (20 ng/ml). Half of medium was changed daily and cells were cultured up to day 25.

Immunocytochemistry

To perform immunocytochemistry cells were fixated with 4% Paraformaldehyde and blocked with blocking buffer containing 5% Normal Goat Serum (Gibco®), 0,1% bovine serum albumin (SigmaAldrich) and 0,3% Triton X-100 (SigmaAldrich). Primary antibody incubation for MeCP2 (D4F3, CellSignaling, 1:200, rabbit), OCT3/4 (C-10, Santa Cruz, 1:1000, mouse), SSEA4 (Developmental Studies Hybridoma Bank, 1:50, mouse), TRA1-60 (Santa Cruz, 1:200, mouse), TRA1-81 (Millipore, 1:250, mouse), SOX2 (Millipore, 1:1000, rabbit) was performed in blocking buffer over night at 4°C. Next day cells were washed and secondary antibody Alexa Fluor® 488 (ThermoFisher, 1:1000, mouse or rabbit) and Alexa Fluor® 594 (ThermoFisher, 1:1000, mouse or rabbit) were applied in blocking buffer for 1 h at room temperature. To identify cell nuclei DAPI was used for 5 min before cells were mounted with Fluoromount™ (Sigma-Aldrich).

RNA collection, Sequencing and PCR analysis

To isolate RNA samples, standard TRIzol®-Chloroform isolation was done. RNA was stored at -80°C until further processed. For RNA-Seq, RNA-quality was validated, using Agilent 2200 TapeStation system. When samples showed RNA Integrity Number (RIN) > 8 they were collected and RNA was further processed according to manufacturer's protocol (Illumina, Catalog # RS-122-9004DOC) followed by sequencing by Illumina HiSeq 4000 (50 base pair single read).

For PCR analysis RT-PCR was performed. cDNA was synthesized by using SuperScriptIV-Kit (ThermoFisher) following manufacturer's recommendations and could be stored until further processing at -20°C. To perform PCR different primer sets were used (**Table 1**) and PCR was executed with Phire Hot Start II DNA Polymerase (ThermoFisher).

Table 1. Primers used for iPSC characterization.

OCT3/4	Fwd:	**GAC AGG GGG AGG GGA GGA GCT AGG**
	Rev:	CTT CCC TCC AAC CAG TTG CCC CAA AC
SOX2	Fwd:	GGG AAA TGG GAG GGG TGC AAA AGA GG
	Rev:	TTG CGT GAG TGT GGA TGG GAT TGG TG
NANOG	Fwd:	CAG CCC CGA TTC TTC CAC CAG TCC C
	Rev:	CGG AAG ATT CCC AGT CGG GTT CAC C
C-MYC	Fwd:	GCG TCC TGG GAA GGG AGA TCC GGA GC
	Rev:	TTG AGG GGC ATC GTC GCG GGA GGC TG
TDGF1	Fwd:	TGC TGC TCA CAG GGC CCG ATA CTT C
	Rev:	TCC TTT CGA GCT CAG TGC ACC ACA AAA C
UTF1	Fwd:	CAG ATC CTA AAC AGC TCG CAG AAT
	Rev:	GCG TAC GCA AAT TAA AGT CCA GA
DNMT3B	Fwd:	CAG GAG ACC TAC CCT CCA CA
	Rev:	TGT CTG AAT TCC CGT TCT CC

Western blotting

Frozen cell pellets were lysed by adding WB-Lysate buffer (50mM Hepes ph 7,5, 150 mM NaCl, 1 mM EDTA, 2,5 mM EGTA, 0,1% TritonX-100, 10% Glycerol, 1 mM DTT). To determine protein concentration

Bradford-Test was performed and 30 ug of sample were used. For SDS-PAGE pre-casted Gels were used (Biorad) and ran in 10x Tris/Glycine Buffer for Western Blots and Native Gels (Biorad #1610734). Gels were blotted in tank-blotter (Biorad) on PVDF membranes (Biorad) according to manufactures protocol. After protein transfer blots were blocked in 5% BSA/TBS for 1 h and stained for Sox2 (1:100, Millipore AB5603), Sox9 (1:250; CellSignalling 82630) and ß-actin (1:1000; Chemicon, C4 MAB 1501) in 5% BSA/TBS over night at 4°C. Next day blots were washed and stained with secondary antibodies in 5% BSA/TBS for 1 h at RT. After another 3 TBS washes blots were stained with SuperSignal™ West Femto Maximum Sensitivity Substrate (ThermoFisher) and analysed with LiCor analyser.

Sample collection

Samples were collected at different days throughout the differentiation. First samples were taken at day 3 of protocol, one day after medium change towards NMM with Dorsomorphin and SB431542. Second samples were taken at day 9, before rosette structures were cut, followed by third sample collection at day 15, after rosettes were manually picked. Finally, fourth samples were taken at day 22, after first passage was performed and cells were recovered. To collect all, cells were washed once with PBS and then scraped off the cell culture plate. Solution was collected in an Eppendorf Microtube and centrifuged at maximum speed for 5 min. Supernatant was discarded and pellet was frozen at -80°C until further processed for mass spectrometry.

Cell lysis and protein digestion

Samples were lysed, reduced and alkylated in lysis buffer (1 % sodiumdeoxycholate (SDC), 10 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 40 mM chloroacetamide (CAA) and 100 mM TRIS, pH 8.0 supplemented with phosphatase inhibitor (PhosSTOP, Roche) and protease inhibitor (Complete mini EDTA-free, Roche). After sonication, samples were centifugated at 20,000 x g for 20 min. Protein concentration was estimated by a BCA protein assay. Reduction was done with 5 mM Ammonium bicarbonate and dithiothreitol (DTT) at 55°C for 30 min followed by alkylation with 10 mM Iodoacetamide for 30 min in dark. Proteins were then digested into peptides by LysC (Protein-enzyme ratio 1:50) at 37°C for 4 h and trypsin (Protein-enzyme ratio 1:50) at 37°C for 16 h. Peptides were then desalted using C18 solid phase extraction cartridges (Waters).

Tandem Mass Tag (TMT) 10 plex labelling

Aliquots of ~ 100 µg of each sample were chemically labeled with TMT reagents (Thermo Fisher) according to **Fig. 1**. In total three TMT mixtures were created for each biological replicate. Peptides were resuspended in 80 µl resuspension buffer containing 50 mM HEPES buffer and 12.5 % acetonitrile (ACN,

pH 8.5). TMT reagents (0.8 mg) were dissolved in 80 µl anhydrous ACN of which 20 µl was added to the peptides. Following incubation at room temperature for 1 hour, the reaction was then quenched using 5% hydroxylamine in HEPES buffer for 15 min at room temperature. The TMT-labeled samples were pooled at 1:1 ratios followed by vacuum centrifuge to near dryness and desalting using Sep-Pak C18 cartridges.

Off-line basic pH fractionation

Before the mass spectrometry analysis, the TMT mixture was fractionated and pooled using basic pH Reverse Phase HPLC. Samples were solubilized in buffer A (5% ACN, 10 mM ammonium bicarbonate, pH 8.0) and subjected to a 50 min linear gradient from 18 % to 45 % ACN in 10 mM ammonium bicarbonate pH 8 at flow rate of 0.8 ml/min. We used an Agilent 1100 pump equipped with a degasser and a photodiode array (PDA) detector and Agilent 300 Extend C18 column (5 µm particles, 4.6 mm i.d., and 20 cm in length). The peptide mixture was fractionated into 96 fractions and consolidated into 24. Samples were acidified with 10% formic acid and vacuum-dried followed by re-dissolving with 5% formic acid/5% ACN for LC-MS/MS processing.

Mass spectrometry analysis

We used nanoflow LC-MS/MS using Orbitrap Lumos (Thermo Fisher Scientific) coupled to an Agilent 1290 HPLC system (Agilent Technologies). Trap column of 20 mm x 100 µm inner diameter (ReproSil C18, Dr Maisch GmbH, Ammerbuch, Germany) was used followed by a 40 cm x 50 µm inner diameter analytical column (ReproSil Pur C18-AQ (Dr Maisch GmbH, Ammerbuch, Germany). Both columns were packed in-house. Trapping was done at 5 µl/min in 0.1 M acetic acid in H2O for 10 min and the analytical separation was done at 100 nl/min for 2 h by increasing the concentration of 0.1 M acetic acid in 80% acetonitrile *(v/v)*. The mass spectrometer was operated in a data-dependent mode, automatically switching between MS and MS/MS. Full-scan MS spectra were acquired in the Orbitrap from m/z 350-1500 with a resolution of 60,000 FHMW, automatic gain control (AGC) target of 200,000 and maximum injection time of 50 ms. Ten most intense precursors at a threshold above 5,000 were selected with an isolation window of 1.2 Da after accumulation to a target value of 30,000 (maximum injection time was 115 ms). Fragmentation was carried out using higher-energy collisional dissociation (HCD) with collision energy of 38% and activation time of 0.1 ms. Fragment ion analysis was performed on Orbitrap with resolution of 60,000 FHMW and a low mass cut-off setting of 120 m/z.

Data processing

Mass spectra were processed using Proteome Discover (version 2.1, Thermo Scientific). Peak list was searched using Swissprot database (version 2014_08) with the search engine Sequest HT. The following parameters were used. Trypsin was specified as enzyme and up to two missed cleavages were allowed. Taxonomy was set for Homo sapiens and precursor mass tolerance was set to 50 p.p.m. with 0.05 Da fragment ion tolerance. TMT tags on lysine residues and peptide N termini and oxidation of methionine residues were set as dynamic modifications, and carbamidomethylation on cysteine residues was set as static modification. For the reporter ion quantification, integration tolerance was set to 20 ppm with the most confident centroid method. Results were filtered to a false discovery rate (FDR) below 1%. Finally, peptides lower than 6 amino-acid residues were discarded. Within each TMT experiment, reporter intensity values were normalized by summing the values across all peptides in each channel and then corrected for each channel by having the same summed value. After that the normalized S/N values were summed for all peptides. Finally proteins were Log_2 transformed and normalized by median subtraction.

Data visualization

The software Perseus was used for data analysis and to generate the plots. Volcano plots for each time point was generated and up- or down-regulated proteins were considered significant with a fold change cut-off = 1.3. Functional analysis to enrich to GO terms were done on David Database and pathway enrichment analysis was done on Reactome Functional Interaction (http://www.reactome.org/). Furthermore, protein interaction network was performed using Cytoscape, Gnenmania plugin.

Results

Generation of iPSCs from RTT and isogenic controls

RTT patient and iCTR fibroblasts were reprogrammed into iPSCs via electroporation of reprogramming plasmids as published before [24]. Pluripotency was confirmed using classic assays, i.e. immunocytochemistry (Additional file **1**: Figure S1a) and RNA expression (Additional file **1**: Figure S1b). Expression for normal and affected MeCP2 in the RTT and iCRT lines was confirmed by immunocytochemistry and PCR for MeCP2 (**Fig. 2a, b**). Additionally, the protein expression level of MeCP2 was validated by mass spectrometry over time and showed a higher expression of MeCP2 in iCTR samples relative to RTT lines (**Fig. 2c**). Generation of NES cells was performed as described before [21]. Neuronal induction was confirmed by morphological alterations of the cells such as neuronal rosette appearance after approximately 12 days of neuronal development initiation (**Fig. 1**).
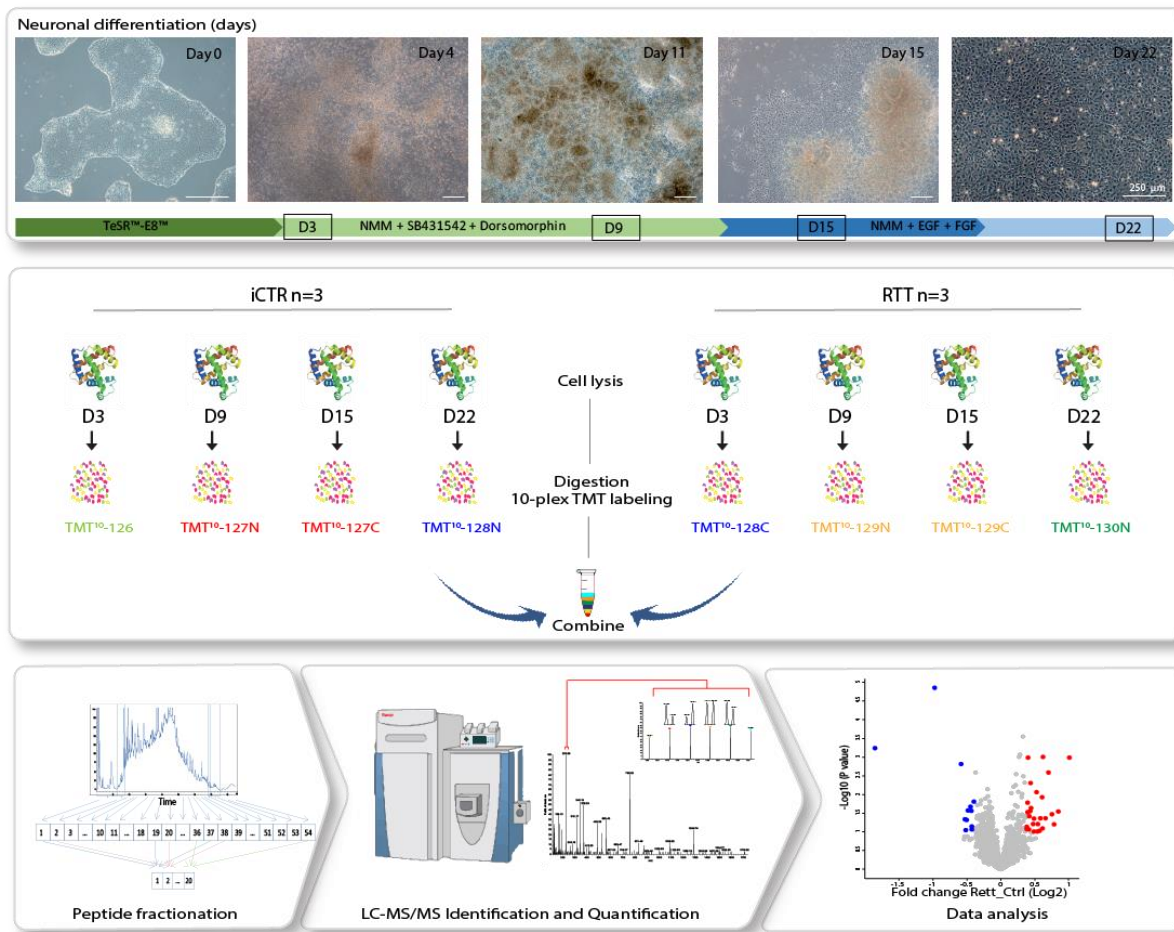
Fig. 1. Overview of experimental workflow. iPSC differentiation towards neurons. Different colors in arrows indicate change of medium. Squares mark days of sample collection. Samples at indicated time points for iCTR and RTT were processed for proteomic analysis (n=3). In total 24 samples were subjected for tryptic digestion, TMT-based isotope labeling, high-pH fractionation and LC MS/MS analysis. Different bioinformatic approaches were then used to analyze the data.

MS-based quantitative proteomics during neuronal development

To get an overview of the proteomic changes between RTT and iCTR during neuronal development, cell lysates at indicated time points were subjected to tryptic digestion, high-pH fractionation followed by high-resolution tandem mass spectrometry (LC-MS/MS) analysis and TMT-10plex quantification (**Fig. 1**). In total we identified 7720 proteins, of which 3658 proteins were quantified in all samples (Additional file **1**: Figure S2). Next, to determine protein expression changes over time-points, we compared RTT versus iCTR and considered proteins with a p-value ≤ 0.1 and ≥ 1.3 fold change in 2 out of 3 biological replicates as significantly regulated (**Fig. 3a**). This resulted in 23 significantly up or down regulated proteins between

RTT and iCTR on D3, 111 on D9, 72 on D15 and 243 on D22. We then compared the significantly up or down regulated proteins across different time points in a Venn diagram (Additional file **1**: Figure S3). We noticed that the majority of the significantly regulated proteins have only 0-2% overlap between the different time points. To better understand the biological processes of the proteins that are differentially expressed at each time point, we performed Gene Ontology (GO) enrichment analysis with respect to biological functions (**Fig. 3b**). GO analysis on the significant proteins revealed terms related to up regulation in RTT versus iCTR of 'neuron apoptotic process' and 'cellular response to hypoxia' on D3. Terms related to 'negative regulation of axon regeneration', 'positive regulation of filopodium assembly' and 'positive regulation of neuron projection development' were down regulated in RTT on D3. On D9 of neuronal differentiation, 'insulin receptor signaling pathway' was up regulated whereas 'excitatory postsynaptic potential' was down regulated in RTT. On D15, 'cell-cell adhesion' and 'acyl-CoA metabolic process' were up regulated and terms such as 'axon guidance', 'brain development' and 'histone acetylation' were down regulated. Furthermore, on D22, terms related to 'cell-cell adhesion' and 'actin cytoskeleton organization' were up regulated and 'nervous system development' and 'forebrain development' were down regulated in RTT. Terms such as brain development were down regulated in D15 as well as D22. To further verify the results of the mass spectrometry, we performed western blot analysis for proteins SOX2 and SOX9, transcription factors with pivotal role in development and differentiation [25, 26], which showed significant differences in expression levels between iCTR and RTT lines at D22 (**Fig. 3a, c**). In line with mass spectrometry data, western blot analysis showed a significant increase in SOX9 expression levels in RTT lines when compared to iCTR (p=0.0057, unpaired t-test), and a decrease in SOX2 expression in RTT lines at D22, although this did not reach statistical significance (p=0.07, unpaired t-test). Together both approaches demonstrate that SOX2 and SOX9 were differentially expressed in RTT versus iCTR. Overall, we show that proteins associated with neuronal development are differentially expressed in RTT at early stages of neuronal differentiation.
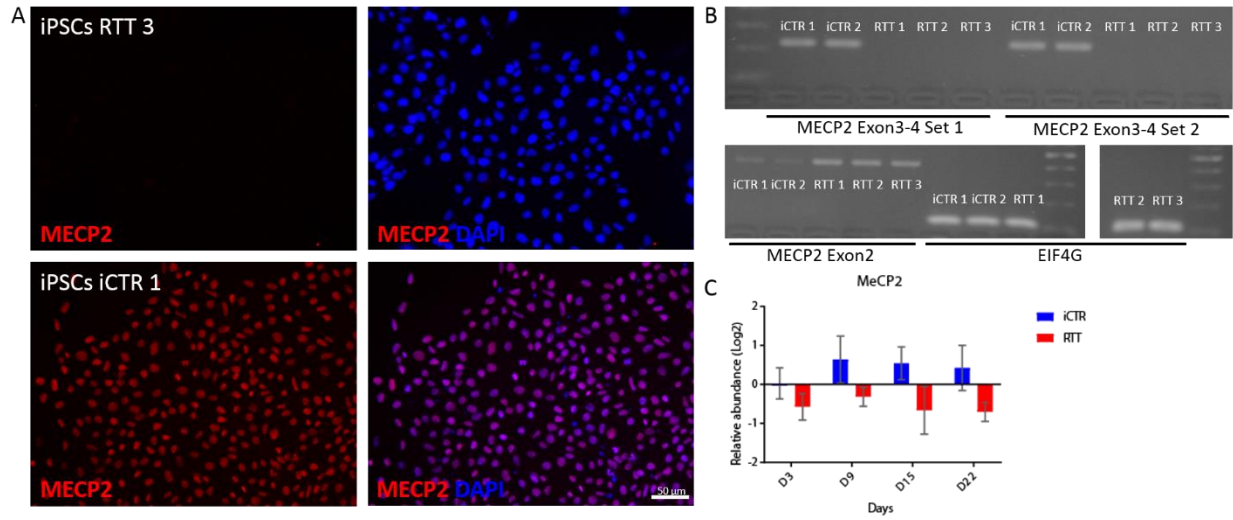
Fig. 2. iCTR and RTT cell line validation. **a**. Representative immunohistochemical staining for MeCP2 in RTT iPSC line (upper) and iCTR iPSC line (lower). **b**. PCR results of iCTR and RTT iPSC lines for two different primer sets spanning deletion Del_Ex3-4 and Exon2 as positive control. **c**. Relative abundance level of MECP2 in RTT NES and iCTR NES lines at different time points of sample collection.
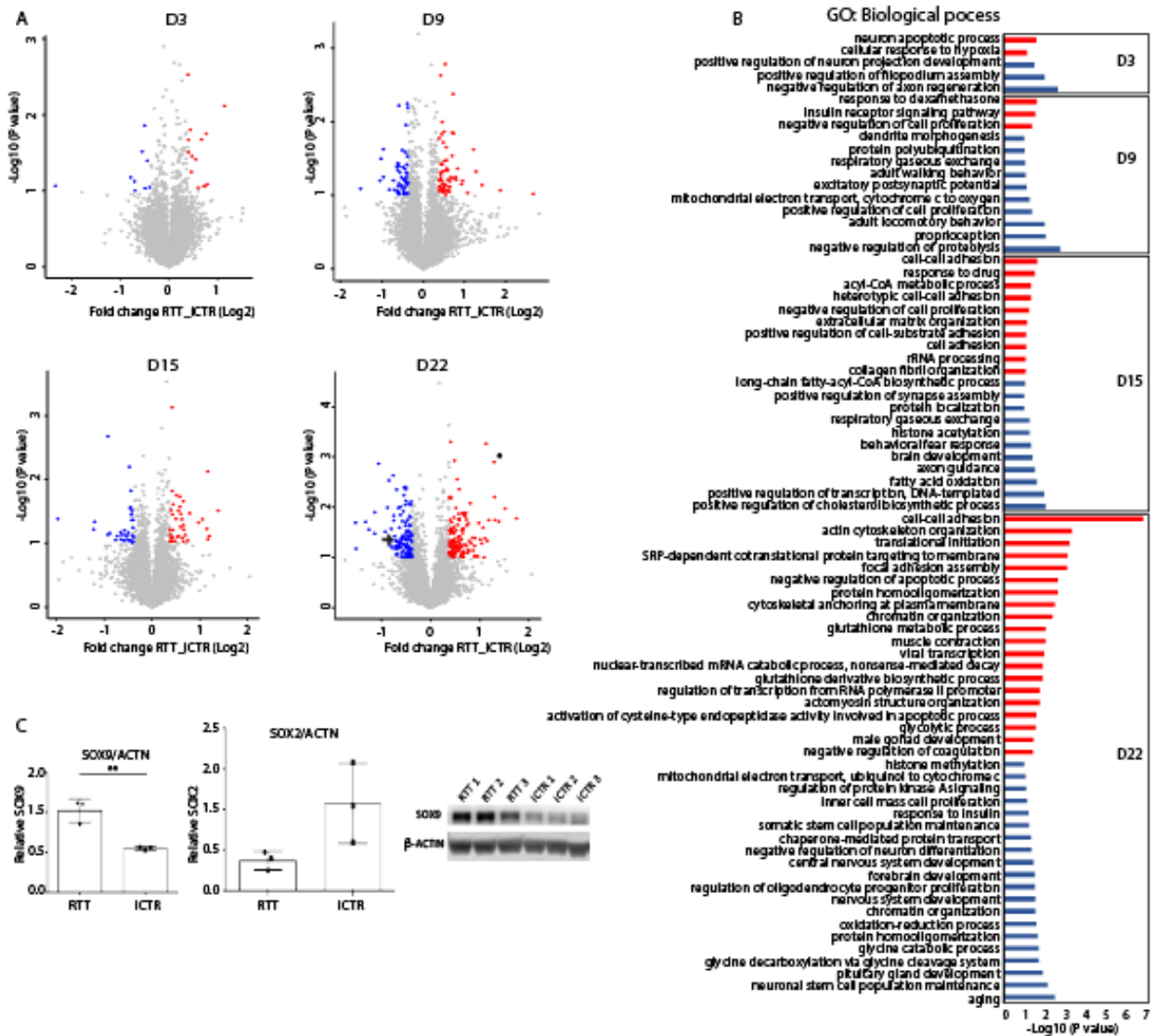
Fig. 3. Volcano plots each day and GO analysis. **a.** Volcano plot demonstrating proteins differentially regulated in RTT compared to iCTR in each time point of neuronal development. Each data point represents a single quantified protein. The x-axis represents the log2-fold change in abundance (RTT/iCTR) and y-axis the -log10 (p-value). Threshold for significant proteins is chosen for p- value cut-off (0.1) and fold change ≥ 1.3. Proteins in blue indicate for down regulation and in red indicate for up regulation in RTT. Arrow in day 22 points for SOX2 expression and asterisk points for SOX9 expression. **b**. Gene Ontology analysis of the significant proteins on their biological function in each time point of neuronal differentiation. Red indicates the up regulated and blue the down regulated biological processes. **c**. Western blot analysis showing SOX9 and SOX2 expression in iCTR NES lines at day 22. Significant increase in SOX9 expression in RTT samples and SOX2 shows a trend towards decrease in RTT samples.

79

Coordinated proteome alteration during neuronal development in RTT syndrome

To gain insight into how the differentially expressed proteins in RTT behave across time points, we further analysed all the significantly up or down regulated proteins at D3, D9, D15 and D22. This resulted in 234 significantly up and 190 down regulated proteins. The average log2 values of RTT were extracted with iCTR for each time point and the difference between RTT and iCTR is shown in a heat map (**Fig 4a**). To obtain an unbiased view of the differentially expressed proteins during neuronal differentiation, we performed cluster analysis on the significantly up or down regulated proteins. This resulted in four clusters for both the up or down regulated proteins with distinct expression profiles. Cluster 1 contains proteins strongly up regulated in D3 that are involved in neuron apoptotic processes and cytochrome c release from mitochondria-related GO terms. The majority of the up regulated proteins belong to cluster 2 whose expression levels increase in D22 and are involved in adhesion assembly and glutathione metabolic processes. Cluster 3 contains proteins up regulated in D9 and D15, whereas cluster 4 represents proteins strongly up regulated in D9. In the down regulated proteins, cluster 1 represents proteins that showed a strong down regulation at D15 and that are mainly involved in cholesterol biosynthesis and fatty acid oxidation (**Fig. 4b**). Cluster 2 represents proteins strongly down regulated in D9, which are associated with regulation of proteolysis and mRNA stability. Cluster 3 of the down regulated proteins in RTT shows a decrease expression profile in D3 which are involved in axon regeneration and neuron projection development and cluster 4 covers many proteins strongly down regulated in D22 involved in processes such as neuronal stem cell maintenance, pituitary gland development and aging. Collectively, our data reveals the expression changes of the differentially expressed proteins during neuronal development in RTT having a stage-specific expression pattern.
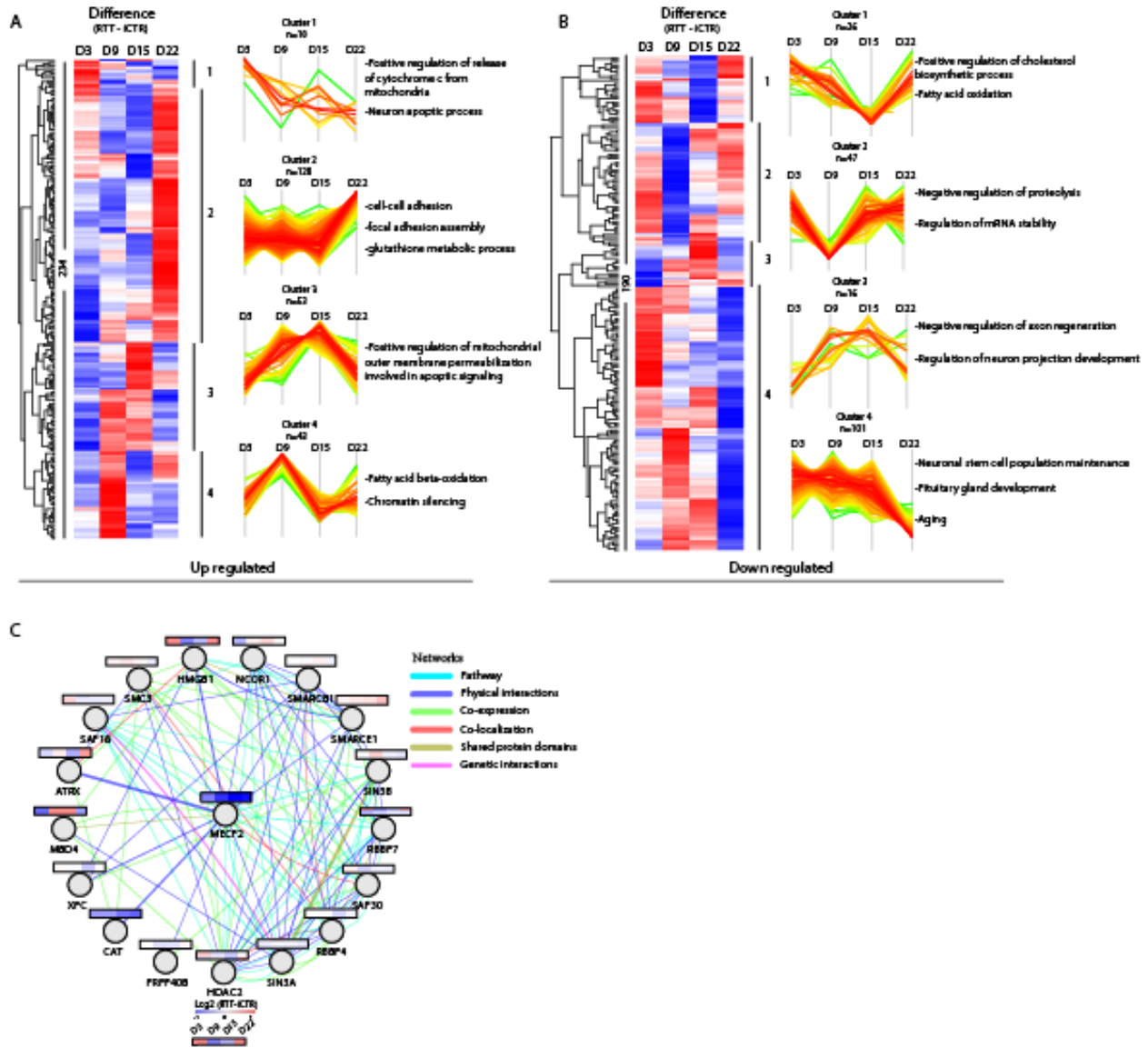
Fig. 4. Up and down regulated proteins and MeCP2-binding partners. **a**. Heat map of all significant up and **b**. down regulated proteins with a p-value ≤ 0.1 and ≥ 1.3 fold change between RTT and iCTR. The Z score of the difference between RTT - iCTR is given for each day with the corresponding cluster analysis and the GO terms for biological processes. **c**. Network analysis of MeCP2-binding proteins identified in our data. The average Log2 ratio RTT-iCTR over time is color coded for each protein over the time course of neuronal development. Edges are color coded according to the network as indicated.

MeCP2 network analysis

To further investigate the proteins that are targets of the MeCP2 protein, we drew a protein interaction network (Cytoscape, Genemania plugin) using MeCP2 protein as input (**Fig.4c**). The data covered 20 MeCP2-interacting proteins of which 18 proteins were identified in our data. To further investigate how the MeCP2- interacting proteins change over time in RTT, we extracted the average log2 values of RTT from iCTR at each time point. As expected, MeCP2 is down regulated along the course of neuronal differentiation. The proteins are tightly interconnected around HDAC2, SIN3A, RBBP4, SAP30, SMARCE1, SMARCB1, and MECP2, but to a lesser extend around CAT, XPC, PRPF40B, and MBD4. The network revealed several RNA/DNA binding proteins of which MECP2 and CAT are one of the most down regulated proteins in RTT. While some proteins in the network, such as SMARCB1 and SIN3A stayed constant over time, others showed changing levels, such as MBD4 and HMGB1. Interestingly, MBD4 is a member of the methyl-CpG-binding domain (MBD) family of proteins together with MeCP2. Overall, we searched for MeCP2-binding partners and showed how these proteins change along the course of neuronal differentiation in RTT and iCTR.
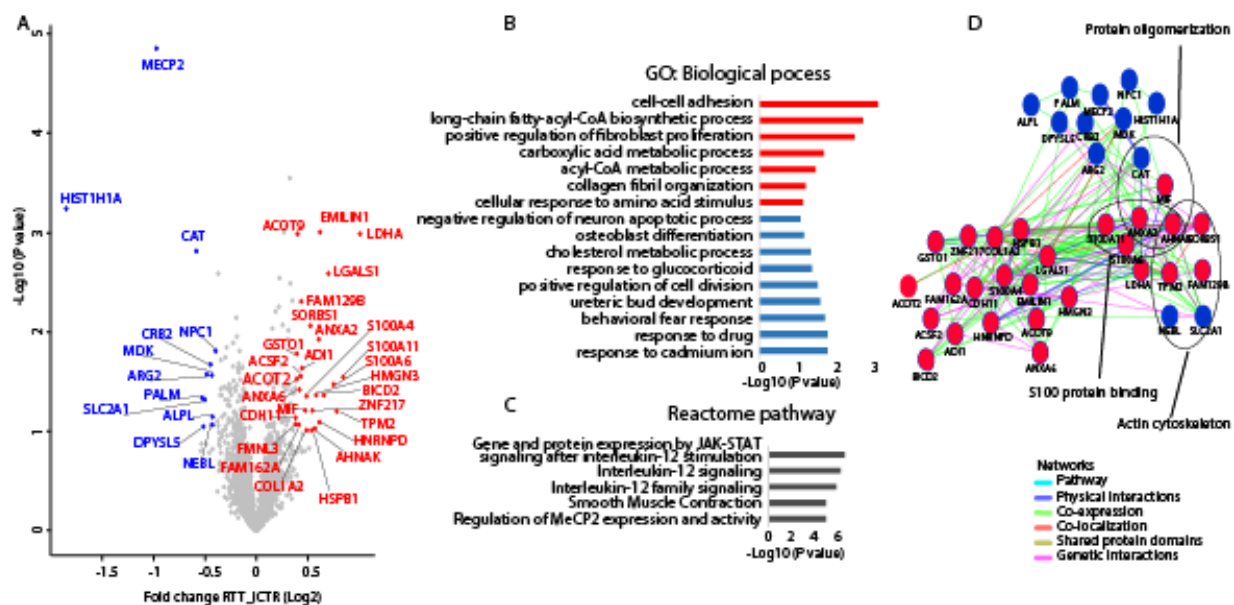

Fig. 5. Volcano plot all days and network analysis. **a**. Volcano plot demonstrating proteins differentially expressed in RTT versus iCTR after pooling all time points of neuronal development. The x-axis represents the log2 fold change in abundance (RTT/iCTR) and y-axis the -log10 (p-value). Threshold for significant proteins is chosen for p- value cutt-off (0.1) and fold change ≤ 1.3. Up regulated proteins in RTT are shown in red and down regulated are shown in blue. **b**. GO analysis on the Biological Process of the significant proteins. X-axis represents the -Log10 (p-value) and red and blue colors indicate for up and down regulated proteins in RTT respectively. **c**. Reactome pathway analysis of

the significant proteins. **d**. Network analysis of the significant proteins by Cytoscape plugin GeneMania. Red indicates for up regulated and blue indicates for down regulated proteins in RTT.

Protein subsets differentially expressed at all-time points

To study proteins differentially expressed between RTT and iCTR regardless of the time point of differentiation, we grouped the RTT samples derived from all time points together and iCTR derived from all time points together. Due to the tight ratios typically observed in TMT quantification [27], we selected a cut-off for proteins being up or down regulated with a p-value ≤ 0.1 and ≥ 1.3 fold change difference in RTT compared to iCTR based on the observed distribution in the volcano plot (**Fig. 5a**). We identified 27 proteins being up and 12 proteins being down regulated in RTT compared to iCTR. As expected, MeCP2 was one of the most strongly down-regulated proteins in RTT. GO analysis revealed biological processes such as 'cell-cell adhesion' and 'acyl-CoA metabolic processes' to be up regulated, which was up regulated as well in D15 and D22 (**Fig. 4a, 5b**). In contrast, several processes such as 'response to cadmium ion', 'response to drug' and 'behavioral fear response' were down regulated in RTT (**Fig. 5b**). Analysis of the differentially regulated proteins using Reactome pathway analysis revealed among others, 'JAK/STAT signaling after Interleukin-12 stimulation' and 'regulation of MeCP2 expression and activity' to be differentially expressed in RTT versus iCTR (**Fig. 5c**). To further visualize the connectivity among these significant proteins, we analyzed their protein networks in the Cytoscape tool (Genemania plugin). A high degree of connectivity, such as being co-expressed and having shared genetic interactions, around these proteins was identified. Interestingly, the majority of the proteins are involved in immunity, actin cytoskeleton organization and calcium binding (**Fig. 5d**). Together, we show that proteins associated with immunity and metabolic processes are differentially expressed in RTT in a time-point independent manner during differentiation towards neurons.

Discussion

RTT samples present protein expression changes that became more apparent from early to late neuronal stem cell phases.

Already at D3 of neural induction, we found decreased protein levels associated with axon regeneration and filopodium assembly, and increased protein levels in neuronal apoptotic processes. Furthermore we show a down regulation of proteins associated with dendrite morphogenesis and excitatory postsynaptic potential at D9, and down regulation of axon guidance and brain development at D15 of neural induction. By D22 a clear set of proteins was altered in RTT with up regulated proteins associated with 'cell adhesion', 'cytoskeleton organization' and 'translation initiation'. Proteins that are down regulated in RTT are

involved in 'nervous system development', 'forebrain development' as well as 'histone methylation', which has classical roles with MeCP2 in its organization [35]. While the proteomic alterations found are relatively small (1.3 fold-change), they are in line with previous studies showing dysregulation of dendrites and axons in RTT patients/mouse models [33, 36-38] as well as in the juvenile RTT brain [39-41]. Although at these time points of neuronal stem cells did not develop dendrites or axons yet, our findings indicate that proteins involved in these processes are already expressed and altered in RTT at early developmental stages. During day 3-22, the number of proteins that are differentially expressed increases, although this set of protein changes over the time points. The increase of the fold change over time is an indicator of the manifestation of RTT, starting from early brain development towards the mature central nerve system. Furthermore, we noticed a small overlap of the altered proteins across time points which might indicate that proteins associated with RTT are specifically expressed at distinct time-points during the course of neural differentiation and reflect the robust cellular identity changes during early developmental stages. Overall, we show that dysregulation of MeCP2 affects protein expression changes associated with neurodevelopmental functions at early stages of neuronal differentiation.

Expression changes of MeCP2-interacting proteins in RTT and iCTR

Network analysis using GeneMANIA [42] revealed part of these proteins to be interacting with MeCP2. Interestingly, of these interacting proteins, MBD4 is a member of the methyl-CpG-binding domain (MBD) family of proteins together with MeCP2. Mutations in these functional important domains tend to cause RTT associated phenotypes [43-45]. Furthermore, HMGB1, which was down regulated in D9 and D15 in our data, was previously shown to be lower expressed in hippocampal granule neurons of MECP2 KO mice [46].

Robust protein profile changes along the course of neuronal differentiation

When all RTT versus iCTR samples from different time points were pooled, data revealed differentially expressed proteins in RTT involved in immunity, calcium binding and metabolism. This analysis allowed us to compensate for limited amount of biological replicates in such high-throughput technology. Several proteins associated in metabolic processes were differentially expressed in RTT, including ACSF2, ACOT2, ACOT9, LDHA, MIF, NPC1, CAT and GSTO1. Current evidence in perturbed lipid metabolism in the brain and the peripheral tissue of RTT patients and mouse models now also supports the metabolic dysfunctions as a component of RTT [34]. Also mRNA *GSTO1* was up regulated in RTT patients' lymphocytes together with several other mitochondrial related genes [47]. Furthermore, our results indicate dysregulated proteins associated with Interleuking-12 signalling. Interestingly, children with MeCP2 duplication show

immunological abnormalities and suppressed IFN-Y [48]. Although RTT is more often classified as a neurodevelopmental disorder, recent studies in cytokine release also suggest involvement of the immune system [49]. We further identified several proteins associated with calcium signalling to be altered. A disturbance in calcium homeostasis during early postnatal development was reported in MeCP2 knockout model and altered calcium signalling RTT-iPSC-derived cells [50, 51]. Altogether, this indicates that next to dysregulation in neurodevelopmental processes, disease mechanisms underlying RTT phenotypes could also involve immunity, calcium signaling and metabolism.

Early treatment

The finding, that neuronal stem cells of RTT patients do show altered protein expressions responsible for neuronal development and maturation, indicates that RTT influences the patients much earlier than first symptoms actually appear. Therefore, possibilities for early treatment need to be discussed. As there is no cure for RTT yet, mutations on MeCP2 are not screened for in pre-natal diagnostics. However, an early testing post-natal could make sense, as we did show that the disease has its onset already before symptoms appear. The time window after birth until approximately 18 month could therefore be used to support neuronal stimulation and maturation with treatments such as IGF-1 or Bumetanide [52-54]. This could support or extend the compensatory process patients' show before the start of symptoms and maybe extenuate the disease onset. The progress of RTT is based on different aspects. There is evidence that mosaicism and therefore the ratio between cells expressing healthy compared to mutated MeCP2 plays a major role [55]. Furthermore, the mutation itself has a huge impact on severity of the disease [56]. However, in almost all cases of patients with MeCP2 mutation a post-natal phase of compensation can be observed. We are convinced that this critical period can be used to start an efficient treatment and prevent disease progression. Therefore, we hope that our results give awareness of the early pre-natal onset of RTT and emphasize the need of early testing followed by an effective treatment.

Limitations

A limitation of our study is the number of biological replicates that was used (n=3). Because of the incredible work that was needed, this number of replicates was feasible for this study. A possible avenue for future research would test few target proteins of a larger scale of biological replicates.

Conclusion

Before typical RTT-associated symptoms appear, both RTT patients and mouse models of RTT already show abnormalities [28, 29]. This implies that underlying mechanisms are already affected during early

neurodevelopmental stages. Much of our understanding of how MeCP2 deficiency contributes to RTT disease is derived from genomic and transcriptomic studies. So far, only a few proteomic studies have been performed on RTT human derived tissue [19, 30, 31]. The current study provides mass spectrometry-based quantitative proteomic data, depth of about 7000 proteins, using an earlier developed iPSC-based models involving RTT patient cells and isogenic controls [24]. We showed that changes in dendrite morphology or synaptic defects, previously associated with RTT [22, 32], already become apparent at early developmental stages. Proteins involved in immunity and metabolism, also in line with previous studies on RTT pathology [33, 34], are differentially expressed at all time points. This indicates we found time-point specific alterations as well as differentially expressed proteins at all time points during early neuronal differentiation in RTT cultures. Insight into differentially expressed protein levels could support identification of novel biomarkers as well as therapeutic strategies.

Abbreviations

iPSC: induced pluripotent stem cell, **MeCP2**: methyl-CpG binding protein 2, **RTT**: Rett syndrome, **iCTR**: isogenic controls, **GO**: Gene ontology, **TMT**: Tandem mass tag,

Declarations

Acknowledgements

Funding

Ethics approval

All experiments were exempt from approval of Medical Ethical Toetsingscommissie (METC), Institutional Review Board of the VU medical centre.

Consent for publication

Not applicable

Competing interest

The authors declare that they have no competing interest

Data availability

All mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD013327

Authors contributions

Suzy Varderidou-Minasian processed all mass spectrometry data and analysis, Lisa Hinz prepared all samples. Dominique Hagemans helped with the cell lysis and TMT labeling. Danielle Postuma, Maarten Altelaar and Vivi Heine helped with the manuscript preparation. All authors read and approved the final draft.

References

1. Chahrour, M. and H.Y. Zoghbi, *The story of Rett syndrome: from clinic to neurobiology.* Neuron, 2007. **56**(3): p. 422-37.
2. Neul, J.L., et al., *Rett syndrome: revised diagnostic criteria and nomenclature.* Ann Neurol, 2010. **68**(6): p. 944-50.
3. Percy, A.K., et al., *Rett syndrome diagnostic criteria: lessons from the Natural History Study.* Ann Neurol, 2010. **68**(6): p. 951-5.
4. Akbarian, S., *The neurobiology of Rett syndrome.* Neuroscientist, 2003. **9**(1): p. 57-63.
5. Marchetto, M.C., B. Winner, and F.H. Gage, *Pluripotent stem cells in neurodegenerative and neurodevelopmental diseases.* Hum Mol Genet, 2010. **19**(R1): p. R71-6.
6. Qiu, Z., et al., *The Rett syndrome protein MeCP2 regulates synaptic scaling.* J Neurosci, 2012. **32**(3): p. 989-94.
7. Smrt, R.D., et al., *Mecp2 deficiency leads to delayed maturation and altered gene expression in hippocampal neurons.* Neurobiol Dis, 2007. **27**(1): p. 77-89.
8. Amir, R.E., et al., *Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2.* Nat Genet, 1999. **23**(2): p. 185-8.
9. Trappe, R., et al., *MECP2 mutations in sporadic cases of Rett syndrome are almost exclusively of paternal origin.* Am J Hum Genet, 2001. **68**(5): p. 1093-101.
10. Schanen, C. and U. Francke, *A severely affected male born into a Rett syndrome kindred supports X-linked inheritance and allows extension of the exclusion map.* Am J Hum Genet, 1998. **63**(1): p. 267-9.
11. Lewis, J.D., et al., *Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA.* Cell, 1992. **69**(6): p. 905-14.
12. Nan, X., et al., *Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex.* Nature, 1998. **393**(6683): p. 386-9.
13. Chahrour, M., et al., *MeCP2, a key contributor to neurological disease, activates and represses transcription.* Science, 2008. **320**(5880): p. 1224-9.
14. Gabel, H.W., et al., *Disruption of DNA-methylation-dependent long gene repression in Rett syndrome.* Nature, 2015. **522**(7554): p. 89-93.
15. Rodrigues, D.C., et al., *MECP2 Is Post-transcriptionally Regulated during Human Neurodevelopment by Combinatorial Action of RNA-Binding Proteins and miRNAs.* Cell Rep, 2016. **17**(3): p. 720-734.
16. Shovlin, S. and D. Tropea, *Transcriptome level analysis in Rett syndrome using human samples from different tissues.* Orphanet J Rare Dis, 2018. **13**(1): p. 113.

17. Colak, D., et al., *Genomic and transcriptomic analyses distinguish classic Rett and Rett-like syndrome and reveals shared altered pathways.* Genomics, 2011. **97**(1): p. 19-28.
18. Matarazzo, V. and G.V. Ronnett, *Temporal and regional differences in the olfactory proteome as a consequence of MeCP2 deficiency.* Proc Natl Acad Sci U S A, 2004. **101**(20): p. 7763-8.
19. Cortelazzo, A., et al., *A plasma proteomic approach in Rett syndrome: classical versus preserved speech variant.* Mediators Inflamm, 2013. **2013**: p. 438653.
20. Altelaar, A.F.M., J. Munoz, and A.J.R. Heck, *Next-generation proteomics: towards an integrative view of proteome dynamics.* Nature Reviews Genetics, 2012. **14**: p. 35.
21. Nadadhur, A.G., et al., *Multi-level characterization of balanced inhibitory-excitatory cortical neuron network derived from human pluripotent stem cells.* PLoS One, 2017. **12**(6): p. e0178533.
22. Andoh-Noda, T., et al., *Differentiation of multipotent neural stem cells derived from Rett syndrome patients is biased toward the astrocytic lineage.* Mol Brain, 2015. **8**: p. 31.
23. Kim, K.C., et al., *MeCP2 Modulates Sex Differences in the Postsynaptic Development of the Valproate Animal Model of Autism.* Mol Neurobiol, 2016. **53**(1): p. 40-56.
24. Hinz, L., et al., *Generation of Isogenic Controls for In Vitro Disease Modelling of X-Chromosomal Disorders.* Stem Cell Reviews and Reports, 2018.
25. De Santa Barbara, P., et al., *Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Mullerian hormone gene.* Mol Cell Biol, 1998. **18**(11): p. 6653-65.
26. Rizzino, A., *Sox2 and Oct-3/4: a versatile pair of master regulators that orchestrate the self-renewal and pluripotency of embryonic stem cells.* Wiley Interdiscip Rev Syst Biol Med, 2009. **1**(2): p. 228-236.
27. Altelaar, A.F.M., et al., *Benchmarking stable isotope labeling based quantitative proteomics.* Journal of Proteomics, 2013. **88**: p. 14-26.
28. Charman, T., et al., *Regression in individuals with Rett syndrome.* Brain Dev, 2002. **24**(5): p. 281-3.
29. De Filippis, B., L. Ricceri, and G. Laviola, *Early postnatal behavioral changes in the Mecp2-308 truncation mouse model of Rett syndrome.* Genes Brain Behav, 2010. **9**(2): p. 213-23.
30. Cheng, T.-L., et al., *Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain.* Scientific Reports, 2017. **7**: p. 42790.
31. Pecorelli, A., et al., *Proteomic analysis of 4-hydroxynonenal and nitrotyrosine modified proteins in RTT fibroblasts.* Int J Biochem Cell Biol, 2016. **81**(Pt B): p. 236-245.
32. Kim, K.Y., E. Hysolli, and I.H. Park, *Neuronal maturation defect in induced pluripotent stem cells from patients with Rett syndrome.* Proc Natl Acad Sci U S A, 2011. **108**(34): p. 14169-74.
33. Kaufmann, W.E., M.V. Johnston, and M.E. Blue, *MeCP2 expression and function during brain development: implications for Rett syndrome's pathogenesis and clinical evolution.* Brain Dev, 2005. **27 Suppl 1**: p. S77-s87.
34. Kyle, S.M., N. Vashi, and M.J. Justice, *Rett syndrome: a neurological disorder with metabolic components.* Open Biol, 2018. **8**(2).
35. Martinez de Paz, A. and J. Ausio, *MeCP2, A Modulator of Neuronal Chromatin Organization Involved in Rett Syndrome.* Adv Exp Med Biol, 2017. **978**: p. 3-21.
36. Lee, L.J., V. Tsytsarev, and R.S. Erzurumlu, *Structural and functional differences in the barrel cortex of Mecp2 null mice.* J Comp Neurol, 2017. **525**(18): p. 3951-3961.
37. Armstrong, D.D., *Neuropathology of Rett syndrome.* Ment Retard Dev Disabil Res Rev, 2002. **8**(2): p. 72-6.
38. Johnston, M.V., et al., *Neurobiology of Rett syndrome: a genetic disorder of synapse development.* Brain Dev, 2001. **23 Suppl 1**: p. S206-13.

39.   Armstrong, D.D., *Rett syndrome neuropathology review 2000.* Brain Dev, 2001. **23 Suppl 1**: p. S72-6.

40.   Jellinger, K. and F. Seitelberger, *Neuropathology of Rett syndrome.* Am J Med Genet Suppl, 1986. **1**: p. 259-88.

41.   Boggio, E.M., et al., *Synaptic determinants of rett syndrome.* Frontiers in synaptic neuroscience, 2010. **2**: p. 28-28.

42.   Montojo, J., et al., *GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop.* Bioinformatics, 2010. **26**(22): p. 2927-8.

43.   Moretti, P. and H.Y. Zoghbi, *MeCP2 dysfunction in Rett syndrome and related disorders.* Curr Opin Genet Dev, 2006. **16**(3): p. 276-81.

44.   Shahbazian, M.D. and H.Y. Zoghbi, *Rett syndrome and MeCP2: linking epigenetics and neuronal function.* Am J Hum Genet, 2002. **71**(6): p. 1259-72.

45.   Cukier, H.N., et al., *Novel variants identified in methyl-CpG-binding domain genes in autistic individuals.* Neurogenetics, 2010. **11**(3): p. 291-303.

46.   Smrt, R.D., et al., *Mecp2 deficiency leads to delayed maturation and altered gene expression in hippocampal neurons.* Neurobiology of disease, 2007. **27**(1): p. 77-89.

47.   Pecorelli, A., et al., *Genes related to mitochondrial functions, protein degradation, and chromatin folding are differentially expressed in lymphomonocytes of Rett syndrome patients.* Mediators Inflamm, 2013. **2013**: p. 137629.

48.   Yang, T., et al., *Overexpression of methyl-CpG binding protein 2 impairs T(H)1 responses.* Sci Transl Med, 2012. **4**(163): p. 163ra158.

49.   Leoncini, S., et al., *Cytokine Dysregulation in MECP2- and CDKL5-Related Rett Syndrome: Relationships with Aberrant Redox Homeostasis, Inflammation, and omega-3 PUFAs.* Oxid Med Cell Longev, 2015. **2015**: p. 421624.

50.   Mironov, S.L., et al., *Remodelling of the respiratory network in a mouse model of Rett syndrome depends on brain-derived neurotrophic factor regulated slow calcium buffering.* J Physiol, 2009. **587**(Pt 11): p. 2473-85.

51.   Marchetto, M.C., et al., *A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells.* Cell, 2010. **143**(4): p. 527-39.

52.   Castro, J., et al., *Functional recovery with recombinant human IGF1 treatment in a mouse model of Rett Syndrome.* Proc Natl Acad Sci U S A, 2014. **111**(27): p. 9941-6.

53.   Kahle, K.T., et al., *Decreased seizure activity in a human neonate treated with bumetanide, an inhibitor of the Na(+)-K(+)-2Cl(-) cotransporter NKCC1.* J Child Neurol, 2009. **24**(5): p. 572-6.

54.   Tarquinio, D.C., et al., *Age of diagnosis in Rett syndrome: patterns of recognition among diagnosticians and risk factors for late diagnosis.* Pediatr Neurol, 2015. **52**(6): p. 585-91.e2.

55.   Metcalf, B.M., et al., *Temporal shift in methyl-CpG binding protein 2 expression in a mouse model of Rett syndrome.* Neuroscience, 2006. **139**(4): p. 1449-60.

56.   Cuddapah, V.A., et al., *Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome.* J Med Genet, 2014. **51**(3): p. 152-8.

57.   Shi, Y., P. Kirwan, and F.J. Livesey, *Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks.* Nature Protocols, 2012. **7**: p. 1836.
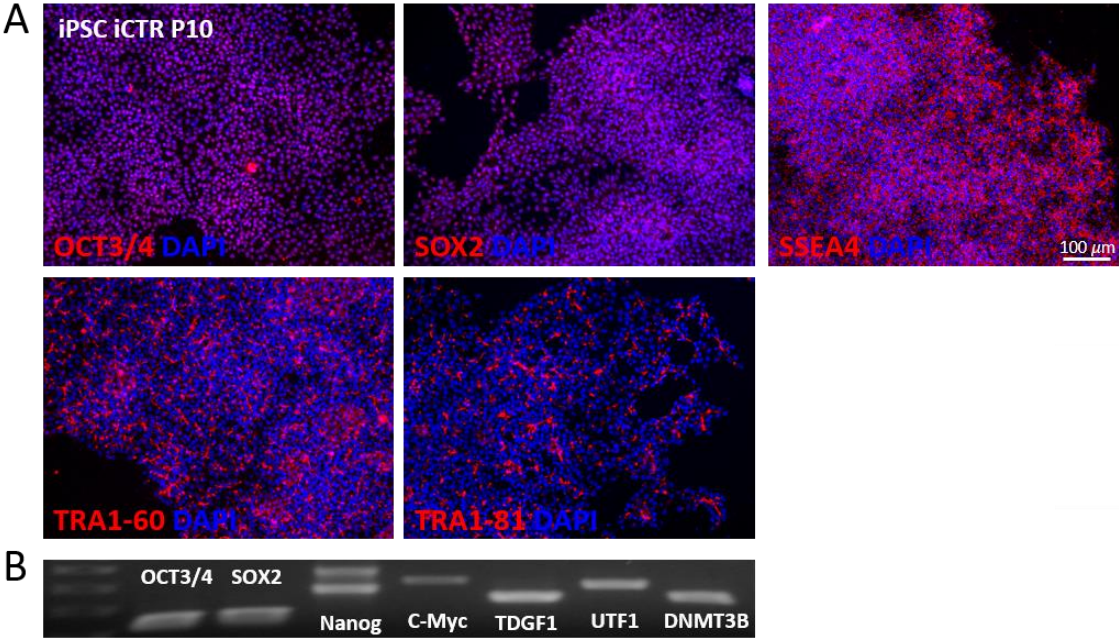
Supplementary data

Figure S1. iPSC characterisation. **a**. Exemplary characterisation of iPSC lines. Immunocytochemistry for pluripotency marker (OCT3/4, SOX2, SSEA4, TRA1-60, TRA1-81). **b**. PCR-analysis for pluripotency marker.
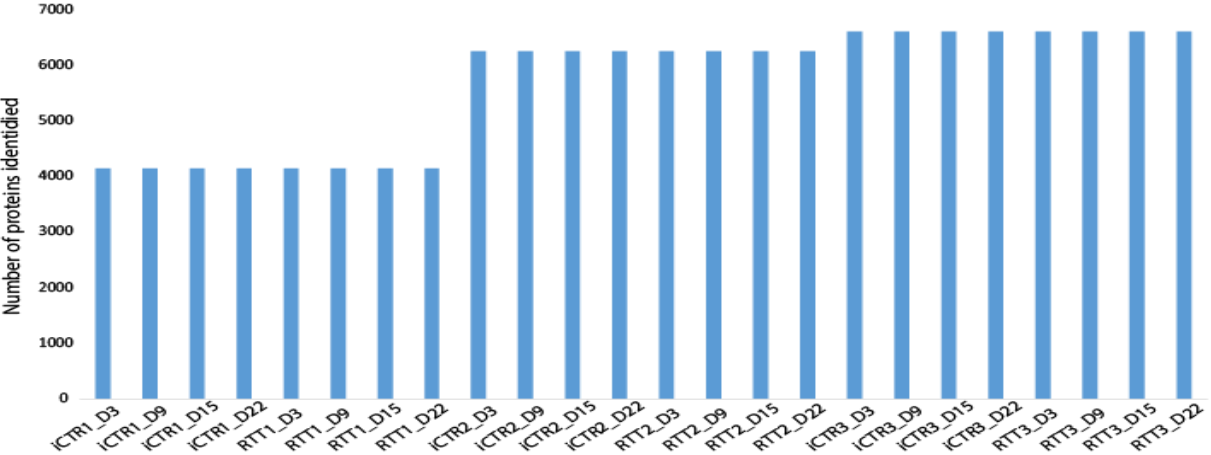

Figure S2. Number of proteins identified. A bar chart showing the number of proteins identified in each biological replicate and time point when analyzed as 20x high-pH fractions.
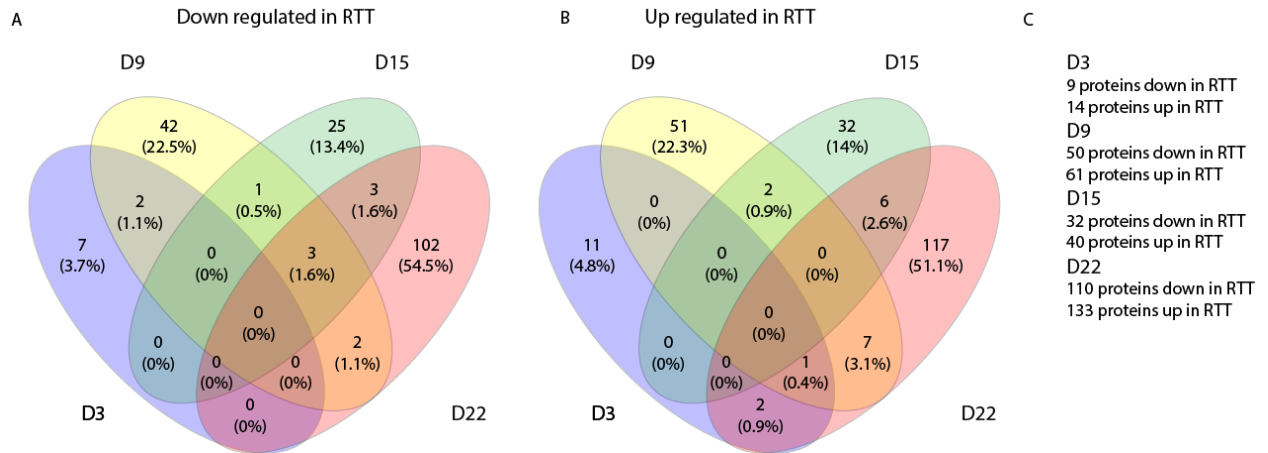
Figure S3. Venn diagram. **a**. Number of proteins decreased in expression in RTT at different time points. **b**. Number of proteins increased in expression in RTT at different time points. **c**. Overview of the number of proteins altered in RTT.

Spinal Muscular Atrophy Patient iPSC-derived Motor Neurons Display Altered Proteins at Early Stages of Differentiation.

Suzy Varderidou-Minasian[1,2], Bert M. Verheijen[3], Henk Karst[3], W. Ludo van der Pol[5], R. Jeroen Pasterkamp[3,4], Maarten Altelaar[1,2]

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

[2] *Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, the Netherlands.*

[3] Department of Translational Neuroscience, UMC Brain Center, University Medical Center Utrecht, Utrecht University, 3584 CG Utrecht, The Netherlands.

[4] UMC Utrecht MIND Facility, UMC Brain Center, University Medical Center Utrecht, Utrecht University, 3584 CG Utrecht, The Netherlands.

[5] *UMC Brain Center, department of Neurology and Neurosurgery,* University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, The Netherlands

Abstract

Spinal muscular atrophy (SMA) is characterized by low levels of survival motor neuron (SMN) protein and loss of motor neurons (MN); however, the underlying mechanism that links SMN deficiency to selective motor neuronal dysfunction is still largely unknown. We present here, for the first time, a comprehensive quantitative mass spectrometry study that covers the development of iPSC-derived MNs from both healthy individuals and SMA patients. We show an altered proteomic signature in SMA already at early stages during MN differentiation, associated with ER to Golgi transport, mRNA splicing and protein ubiquitination, in line with known SMA phenotypes. These alterations in the SMA proteome increase further towards later stages of MN differentiation. In addition, we find differences in altered protein expression between SMA patients, which however, have similar biological functions. Finally, we highlight several known SMN-binding partners as well as proteins associated with ubiquitin-mediated proteolysis and evaluate their expression changes during MN differentiation. Altogether, our work provides a rich resource of molecular events during early stages of MN differentiation, containing potentially therapeutically interesting protein expression profiles for SMA.

Introduction

Spinal muscular atrophy (SMA) is an autosomal recessive neuromuscular disease with onset in childhood, affecting around 1 in 6,000 to 10,000 births [1]. SMA is characterized by the degeneration of motor neurons (MNs) in the spinal cord and the atrophy of muscles, leading to early death [2]. The disease is subdivided into four main types, depending on age and the level of acquired motor milestones, ranging from the most severe *Type I*, characterized by neonatal onset, severe weakness and limited life expectancy, to *Type IV*, with adult onset and mild muscular weakness [3]. SMA is caused by reduced levels of survival motor neuron (SMN) protein, encoded by two, almost identical, genes: SMN1 and SMN2, both located in chromosome 5q13.2 [1, 4]. Most SMA cases are caused by loss or mutation in the SMN1 gene, for which SMN2 cannot fully compensate [5, 6]. SMN2 differs from SMN1 at a single nucleotide and lacks exon 7, which leads to an altered splicing pattern that underlies the production of mostly shortened and unstable SMN protein that can be rapidly degraded by the ubiquitin/proteasome system [7], [8] SMN2 splicing also allows the production of low levels of functional full length protein [9-12] and the SMN2 copy number is the most important determinant of disease severity. SMN is a ubiquitously expressed protein that functions in the assembly of small nuclear ribonucleic proteins (snRNPs) and is important in RNA biogenesis [13]. Here, snRNPs function in the recognition and removal of introns from the pre-mRNA in the nucleus. Small nuclear RNA (snRNA) is transcribed in the nucleus and transported to the cytoplasm.

Together with several SM proteins and snRNP-specific proteins, a snRNP complex is formed. This complex consists of SMN, together with GEMIN2-8 and UNR-interacting protein (UNRIP); it assembles the snRNPs and mediates transport into the nucleus. In the nucleus, the snRNPs undergo further maturation [14-16].

Remarkably, in SMA, the reduced levels of SMN protein affects mainly the lower MNs [17]. To date, it remains unclear why SMA pathology is largely restricted to this selected cell population. Several reasons have been suggested, such as a cellular function of SMN specific to MNs or increased sensitivity of MNs to reduced levels of SMN due to the extended size of their axons, and unique interactions with skeletal muscles. It is known that SMN is expressed in growing neurites and neuromuscular junctions of MNs during neuronal differentiation [18]. It modulates axon growth by interacting with β-actin mRNA, suggesting that SMN functions in the survival of MNs by allowing normal axon growth and transport, and interaction with the neuromuscular junction [19-21].

Although SMA has been studied extensively and the first therapy focused on SMN2 upregulation has been approved by the Food and Drug Administration (FDA) in December 2016, there is a general lack of understanding how molecular pathways are acting downstream of SMN. To date, the only biochemical biomarker for SMA is the level of SMN protein. Furthermore, SMA is often studied at the end stage of MN death, but the molecular pathways active during MN development and how these are affected in SMA remain unclear. To study early neurodevelopmental processes in the light of disorders such as SMA has become feasible through the generation of induced pluripotent stem cells (iPSCs), derived from e.g. human fibroblasts, which holds great promise in the investigation of neurological diseases [22]. This has made it possible to generate human neurons, including MNs, to understand disease mechanisms and potentially identify (better) treatments. However, until now iPSCs have seldomly been employed to recapitulate SMA or to provide more information about the onset of the disease during MN generation [23-27]. While fibroblasts and iPSC-derived MNs in SMA have been compared in previous studies, the underlying process during MN differentiation in SMA remains elusive [25]. Analyzing proteome changes during MN development in a genetic SMA background can improve our understanding of SMA in an unbiased manner and unravel the associated molecular pathways. Doing so will require large-scale mass spectrometry approaches with quantitative assessment.

The aim of our study was to evaluate specific proteome changes during human iPSC-derived MN differentiation between SMA patients and healthy controls. We quantitatively monitored the MN proteome over ten time points during differentiation (using TMT-10plex), revealing affected molecular processes in SMA, such as the upregulation of ER to Golgi vesicle-mediated transport and the

downregulation of mRNA splicing and protein polyubiquitination. Furthermore, we quantitatively monitored known SMN binding partners that function in RNA splicing and visualized their expression levels during MN differentiation.

Results

Generation of iPSCs from SMA and control samples that differentiate towards spinal motor neurons.

SMA (SMA1, SMA2) and control (Ctrl1, Ctrl2) iPSC lines were generated from skin fibroblasts using retroviral transduction of the four transcription factors OCT4, KLF4, SOX2 and c-MYC as previously described [28] (See **supplementary Table 1** for clinical history and origin [29]), and characterized using a range of standardized pluripotency assays (**Figure 1A**). Resulting iPSCs were able to differentiate into different cell lineages (mesoderm, endoderm and ectoderm), and karyotype analysis showed no chromosomal aberrations **(Figure 1B, C)**. Using a slightly modified version of the procedure described by Yves Maury *et al* [30], iPSCs were differentiated into lower spinal motor neurons (MNs). Briefly, from the iPSCs, embryoid bodies (EBs) were formed to induce loss of pluripotency and to promote differentiation into neuronal fate. For neutralization, dual-SMAD signaling was inhibited for four days followed by addition of caudo-ventralizing factors containing retinoic acid and a sonic hedgehog agonist. On day 15, EBs were dissociated into single cells and allowed further maturation for another 9 days. At this point, the MNs were positive for the neuronal marker βIII-tubulin (>80%)) and motor neuron marker ISL1 (~40 %) (**Figure 1D**). Electrophysiological recordings of MNs further showed that MNs were able to fire repetitive action potentials. Together, these data demonstrate that both SMA and control iPSCs can differentiate towards lower MNs that can be used as *in vitro* model to study SMA disease.

Pluripotency

Ctrl_1   Ctrl_2   SMA_1   SMA_2

Oct4a
SSEA4

Nanog
Tra-1-81

Sox2
Tra-1-60

A

Spontaneous differentiation

Ctrl_1   Ctrl_2   SMA_1   SMA_2

AFP

SMA

TUJ1

B

Karyotype

C

Ctrl_1
46, XY

Ctrl_2
46, XY

SMA_1
46, XX

SMA_2
46, XY

MN differentiation

D

T1 iPSC

T4 Embryoid bodies
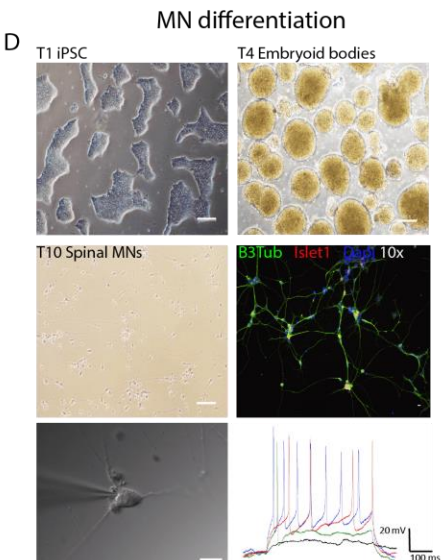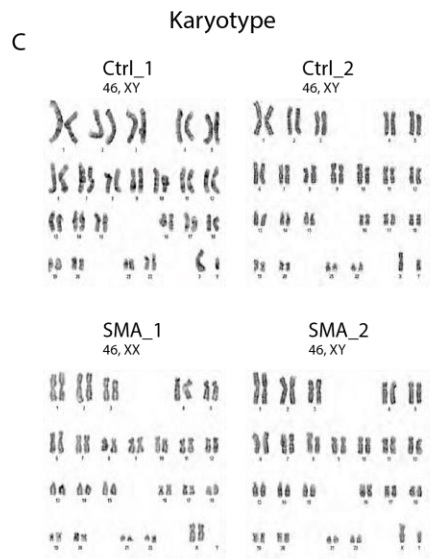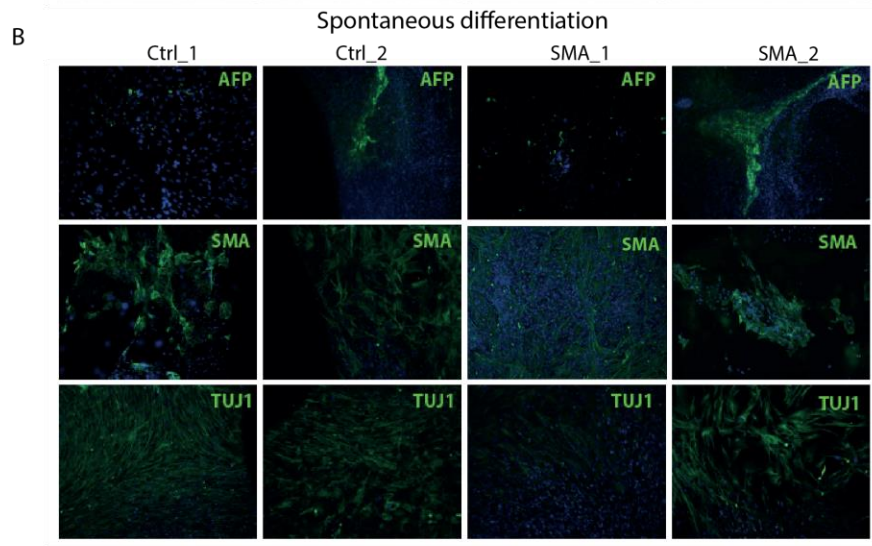
T10 Spinal MNs

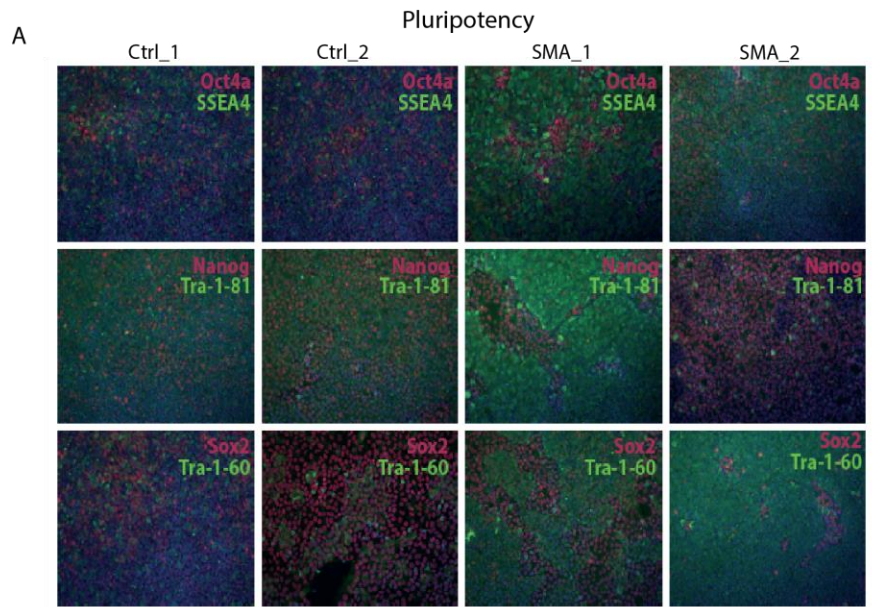B3Tub Islet1 Dapi 10x

20 mV

100 ms

Figure 1. Characterization of iPSC lines and MN differentiation of healthy and SMA lines. (A) Representative positive immunostainings for nuclear and surface pluripotency antigens. (B) Spontaneous differentiation towards endoderm (AFP), mesoderm (SMA) and ectoderm (TUJ1) of all lines. (C) Normal G-band karyotype of the iPSCs for healthy and SMA lines. (D) MN differentiation at indicated time points. Phase contrast image of iPSCs at T1, EB's at T4 and MNs at T10. Scale bar 100 µm. MNs at T10 were immunostained with neuronal marker (B3Tubulin), motor neuron marker (Islet1) and nuclear marker (Dapi). Patched MNs and example of traces showing action potential generation. Scale bar 25 µm.

SMA iPSC differentiation towards MNs show altered proteomic signature

To resolve the proteome changes in SMA during MN differentiation, samples from 10 distinct time points within the differentiation timeline were subjected to in-depth quantitative proteome analysis (**Figure 2**). Samples at indicated time points were digested into peptides, labeled with TMT-10plex, fractionated by high-pH fractionation and analyzed using liquid chromatography coupled to high-resolution mass spectrometry (LC-MS/MS) analysis. We identified 8319 protein groups, with a false discovery of 1% and quantified 3638 proteins across all time points in all biological replicates (**Supplementary Figure S1**). This approach allowed us to identify (and subsequently quantify) low abundant components of the cellular proteome, including the SMN protein and components of the SMN core complex [31], which will be discussed in more detail later. For our quantitative data analysis, we normalized all values of each time point to the reference value of iPSC protein expression before initiation of differentiation (T1). This allowed us to quantitatively compare the protein changes of MN differentiation between the different samples. All data throughout the study reflects a log2 transformed ratio change relative to T1 for each biological replicate.
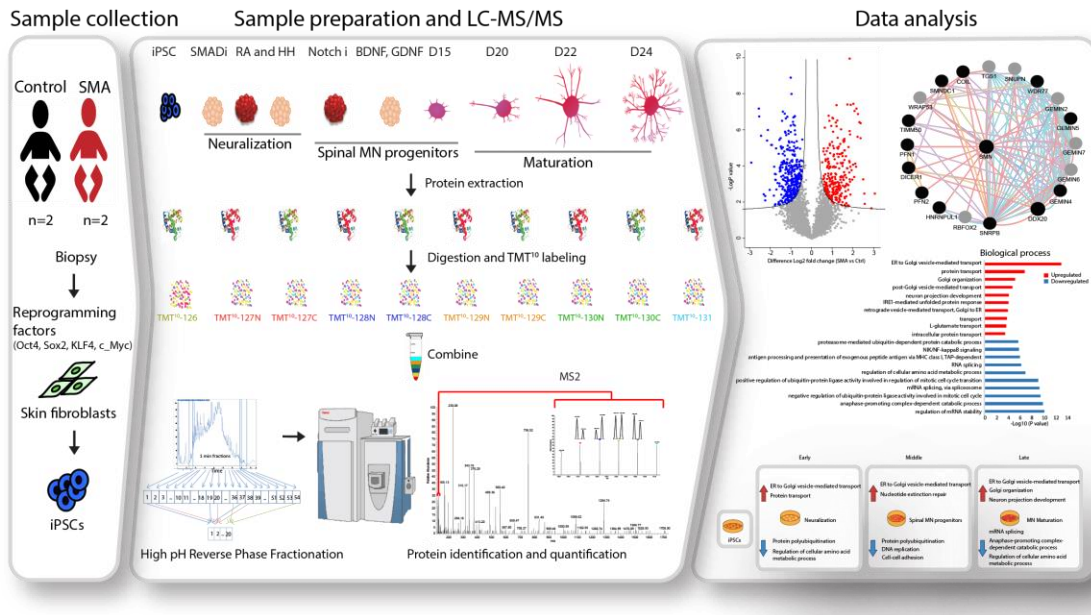
Figure 2. Experimental design. Skin biopsy was obtained from healthy controls and SMA patients. Fibroblasts were reprogrammed into iPSCs and differentiated into spinal MNs. Samples were lysed, proteins were digested into peptides and TMT-10plex labeled. Mixed peptides were pre-fractionated and analyzed by LC-MS/MS. The MS results were analyzed and data was presented to identify proteins differentially expressed in SMA.

During MN differentiation, levels of SMN were reduced in SMA, compared to healthy controls, with SMA1 showing the largest reduction (**Figure 3A**). Interestingly, the difference between SMA and controls increased from T7 onwards, at which point neurite outgrowth and maturation occur. We have previously verified lower SMN levels in SMA compared to control iPSC-derived MNs using western blotting [32]. To get an overview of the whole-proteome changes between SMA and control samples, we performed principal component analysis (PCA) at all time points (**Figure 3B C**). This revealed that samples were largely segregated due to variation between different cell lines and time points of differentiation (component 1 and component 2, which account for 36.3% and 24.7% of variability respectively). However, SMA samples segregate from controls in component 3 (with 12.5% of variability), demonstrating for proteomic alterations between SMA and controls. But the heterogeneity between samples and time points accounting for larger variation needs to be taken into account. The global proteomic changes of the 3638 quantified proteins at all time points revealed that SMA2 cell line expressed the large majority changes in expression followed by SMA1. Furthermore, within each sample, time point variation is observed, segregating early time points (T2, 3, 4), middle time points (T5, 6, 7) and late time points (T8, 9 and 10) (**Figure 3D**).
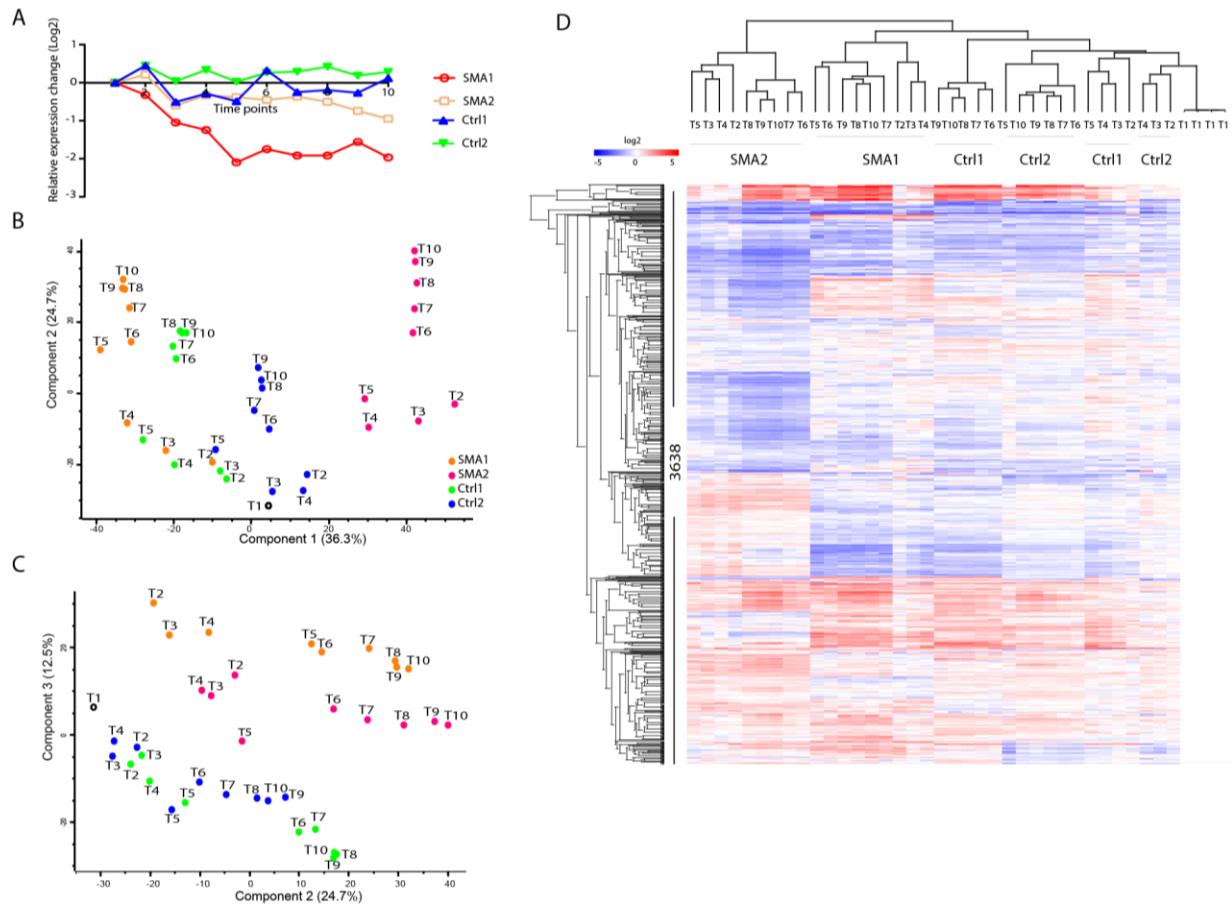
Figure 3. SMN protein expression and proteome clustering. (A) SMN protein expression changes during spinal MN differentiation of healthy controls and SMA patient derived material. (B) The proteome of all time-points during differentiation of healthy controls and SMA patient derived material segregated into biological variation in component 1, time-points in component 2 and (C) healthy vs SMA in component 3. (D) Hierarchical clustering of each sample and time-points.

Proteome alterations at different stages of MN differentiation

To further investigate changes in the proteome of SMA compared to controls during MN development, we separated the timeline into three periods of MN differentiation: –neutralization (early), -spinal motor neuron progenitors (middle) and –maturation (late). Within the indicated time points, we considered proteins significant with an FDR ≤ 0.05. This resulted in 194 significant proteins in early, 130 in middle and 602 in late between SMA and control (**Figure 4A**). Of these, an approximately equal number of proteins were up- or downregulated in SMA. When the differentially expressed proteins in SMA in the three timelines (early, middle and late) are compared, they show a large overlap, especially of the middle time point with the early and late (**Figure 4B**). A large majority of the differentially regulated proteins are in the maturation stage (late) of MN differentiation. The differentially expressed proteins in SMA were examined

using Gene Ontology (GO) enrichment analysis in terms of potential biological processes (**Figure 4C, Supplementary Table 2**). Processes such as 'ER to Golgi vesicle-mediated transport', 'DNA damage', 'neuron projection development', and 'membrane and endoplasmic reticulum proteins functioning in protein binding and SNAP receptor activity' were upregulated in SMA. Processes related to 'protein ubiquitination', 'mRNA splicing' and proteins linked to extracellular exosome and nucleoplasm that function in RNA binding were downregulated in SMA. Dysregulation of mRNA splicing is well described as a hallmark in SMA disease, adding further validity to this developmental model's ability to capture elements of the patient condition [15]. Furthermore, previous studies have shown altered ER to Golgi vesicle transport and protein ubiquitination in relation to SMA, which in this case, reflects similar findings in the iPSC-derived MNs [33, 34]. These results seem to indicate that the SMA genetic background has a substantial impact on the development of MNs and that alterations in cellular processes associated with late stage SMA already manifest themselves at early onset of the disease. Conversely, while previous studies have shown dysregulation of neuronal development at later stages of mature MNs [31], here, we observed upregulation of proteins associated with neuron projection development during maturation. Because the samples in our current studies reflect early time points during MN differentiation, especially compared to post-mortem tissue conventionally studied in SMA related research, this might be a consequence of cells trying to compensate for the lack of SMN protein.
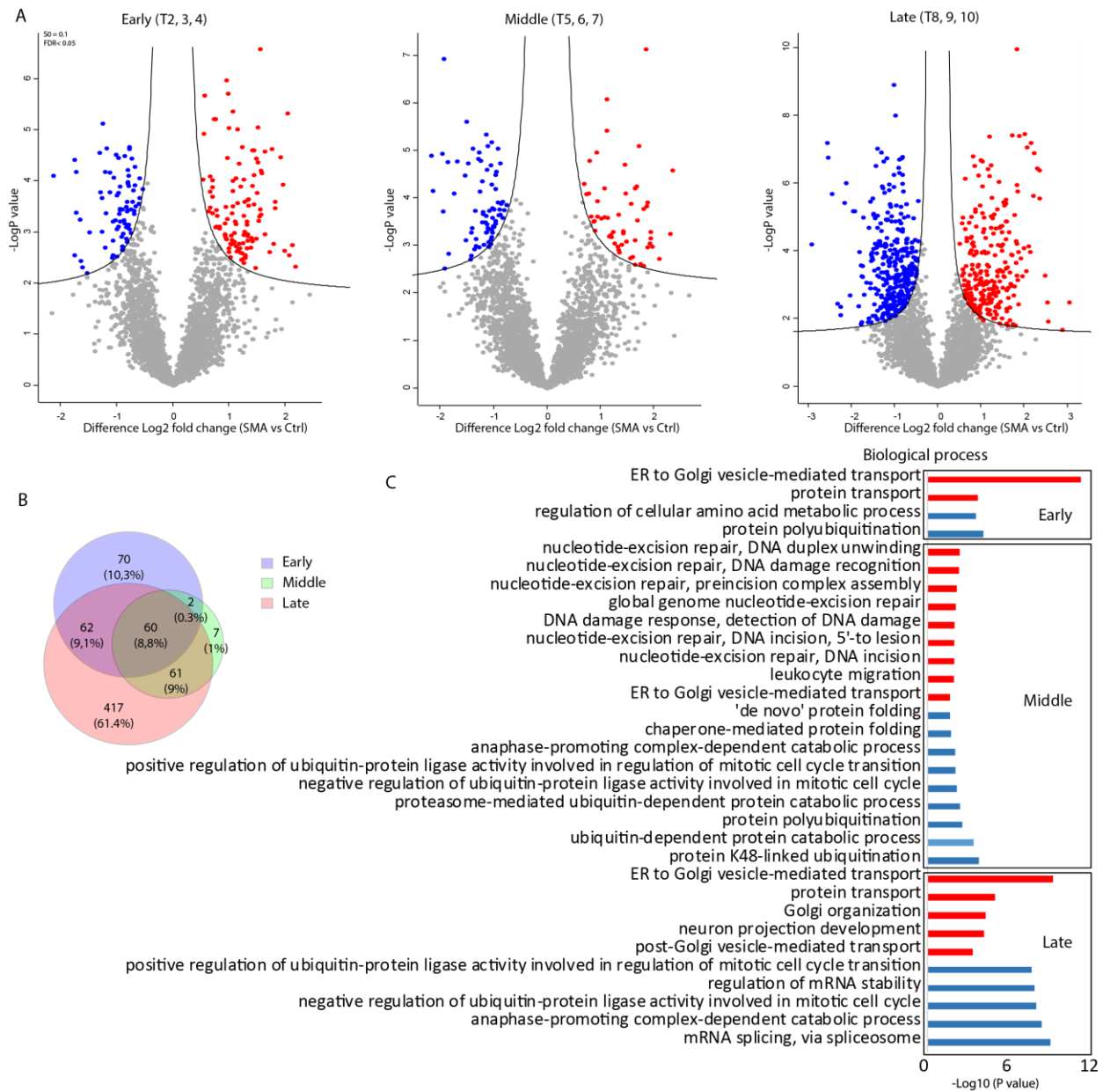
Figure 4.Significance analysis and GO biological processes. (A) Volcano plots illustrating differentially expressed proteins in SMA compared to healthy controls. The log2 fold change is plotted against the –log10 P value during neutralization (early), spinal MN progenitors (middle) and MN maturation (late) with significant threshold FDR < 0.05 and S = 0.1. Red represents proteins upregulated in SMA and blue represents proteins downregulated in SMA. (B) Venn diagram showing the overlay of the significant proteins in the three periods during MN differentiation. (C) GO analysis for biological processes of the significant proteins during MN differentiation at indicated time points. Red represents biological processes upregulated in SMA and blue represents downregulated processes in SMA. (D)

Profile plot showing the difference between SMA and healthy controls of the significant proteins along the time-points.

Different proteins in each SMA line have similar biological functions.

Because of the variability of SMN between the two SMA patients already at the early time points, we further aimed to compare each SMA line separately to the controls to further investigate their individual biological processes. We have shown that SMA2 was the most diverging cell line followed by SMA1. Therefore, we considered proteins significant with FDR ≤ 0.001 for SMA2 and FDR ≤ 0.01 for SMA1 line during neutralization and spinal MN progenitor stages. We then compared the overlap of the significant proteins between the two lines **(Supplementary Figure S2, Supplementary Table 3)**. This showed a minor (~9%) overlap of the significant proteins between the two lines. However, a large majority of the proteins that were either significant in SMA1 or in SMA2 showed similar biological processes. For example, different proteins that were upregulated were associated with Golgi-mediated transport and different proteins that were downregulated were associated with mRNA splicing. Furthermore, proteins that were differentially expressed in both SMA lines were involved in neuromuscular junction development, MAP kinase activity and ubiquitination. Taken together, these results demonstrate that different proteins in each SMA line with minor overlap have similar biological processes. This reveals that there is heterogeneity between patients at the molecular level however, the altered proteins lead to the same cellular processes being affected. In **Figure 5** we summarize our proteomic results in a schematic model to get an overview of events taking place during MN differentiation in SMA. ER to Golgi vesicle-mediated transport is the first step in the secretory pathway where newly synthesized proteins are packaged into vesicles and designed for secretion via exocytosis or use in the cell. Up regulation of protein synthesis and protein transport might indicate that ER to Golgi mediated transport is altered. Notably, a prior study in transcriptome profiling has shown increased activation of ER stress in SMA iPSC-derived MNs [35]. Inhibition of ER stress improved motor function *in vitro* and *in vivo*. Here, we demonstrate that this pathway is affected already at early time points during MN differentiation. In addition, down regulation of protein polyubiquitination and mRNA splicing is a known cellular pathway dysregulated in SMA [36]. Here, we demonstrate that these altered processes become apparent at early developmental stages.
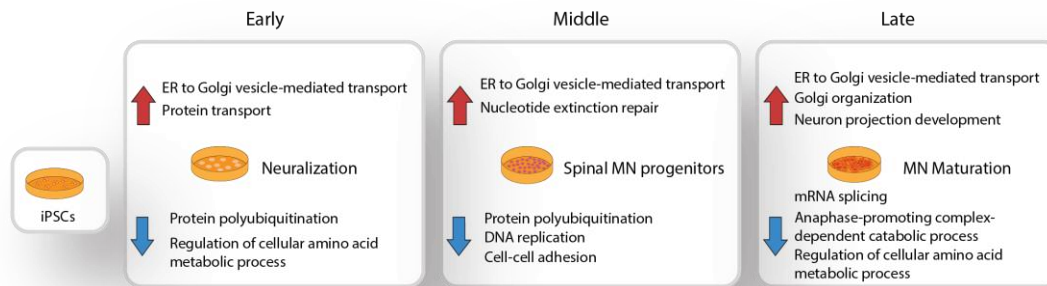
Figure 5. Summary of biological processes during MN differentiation in SMA. Biological processes that are either upregulated or downregulated in SMA in the course of spinal MN differentiation. The differentiation timeline is represented as neutralization (early), spinal MN progenitors (middle) and MN maturation (late).

SMN binding partners

To further investigate the mechanism by which SMN influences SMA, we examined known SMN interaction partners (Cytoscape, Genemania plugin) along MN differentiation. By drawing a protein interaction network around SMN as input, we identified 19 connected binding partners having a role in RNA splicing of which 13 were identified in our mass spectrometry measurements during MN differentiation (**Figure 6A**). To get an overview of the altered expression profiles of these 13 proteins during MN development in SMA, we determined the difference intensity values observed in SMA versus control iPSC lines, which we visualized in a heatmap (**Figure 6B**). Of these, SMN, SMNDC1, SNRPB and PFN1 proteins were significantly (FDR ≤ 0.05, as indicated in **Figure 4**) downregulated in the MN maturation stage. Interestingly, SMNDC1, which is a paralog of SMN and is involved in mRNA splicing shows significant downregulation. However, its association with SMA has so far not been reported, making it a potential candidate for further study. SNRPB is a member of the Sm family of snRNP-associated proteins and SMN protein has a crucial role in the biogenesis of these snRNPs, potentially explaining the concerted downregulation of these two proteins in SMA. Finally, the mutation of PFN1 has been shown to cause Amyotrophic Lateral Sclerosis (ALS) resulting in MNs having smaller axon growth and growth cones [37], but its association with SMA has not yet been studied. Among the molecular processes downregulated in the SMA iPSC line, 'protein ubiquitination', has been strongly associated with SMA before [38, 39]. To further explore the role of ubiquitination in SMA during MN differentiation, the expression profile of proteins associated with KEGG term 'ubiquitin mediated proteolysis' were examined

in more details (**Figure 6C**). We identified 36 proteins of which more than half were downregulated in SMA. CDC34 and UBE2G1 were significantly downregulated at all stages during differentiation. Interestingly, reduced levels of UBA1 were identified in neuromuscular system of SMA mice and iPSC-derived MN of SMA patients [26, 40]. In contrast, we observed a slight upregulation of UBA1 in SMA (not significant). Variable UBA1 levels, as well as SMN levels, were previously observed in iPSC-derived MNs of SMA compared to healthy controls [41]. Overall, we show a specific regulation of functionally distinct subgroups of the proteome around SMN binding partners as well as the ubiquitin mediated proteolysis at specific stages during MN differentiation.
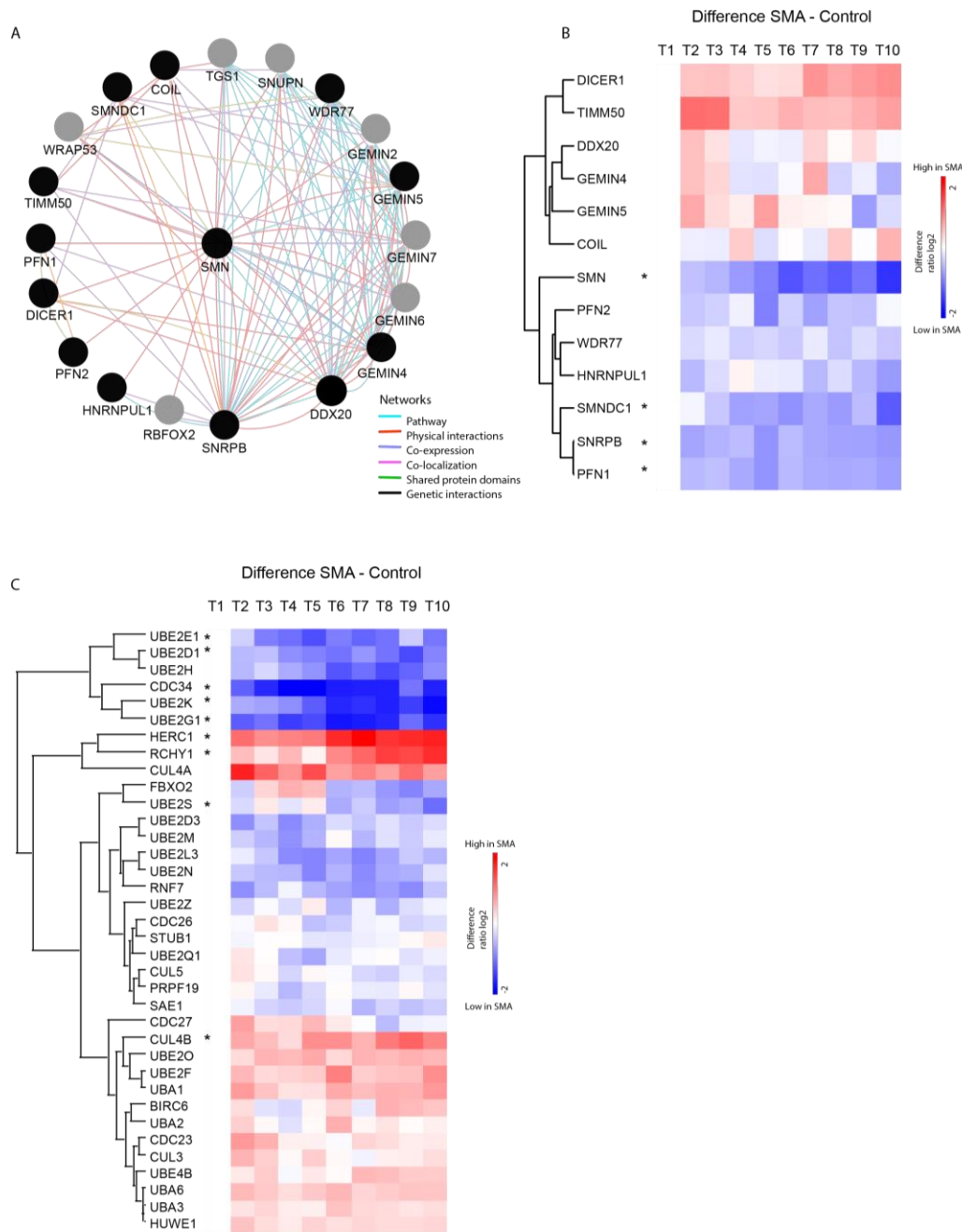
Figure 6. Network analysis around SMN. (A) Network analysis of SMN binding partners derived from Cytoscape (Genemania plugin). Grey color indicates that the protein is not identified and black color indicates that the protein is identified by mass spectrometry measurements. (B) Heat map showing the difference value between SMA and control for each protein. Red color indicates for higher expression in SMA compared to controls and blue indicates for lower expression level in SMA. Asterisk indicates that the protein was significantly different (FDR ≤ 0.05) in iPSC derived spinal MNs of SMA patients. (C) Heatmap of protein expression changes associated to ubiquitin mediated proteolysis.

Discussion

Because lower levels of SMN protein in MNs cause SMA, major efforts are ongoing in both academia and industry to discover SMN-elevating therapeutics [36]. While the effects of reduced SMN are well documented in MNs, studies assessing the impact of SMA at early stages of MN development are less common [42]. Here, we describe the use of human iPSC technology to quantitatively study proteome changes during the entire timeline of MN differentiation in SMA and controls. Evaluation of our quantitative proteomics data indicated altered protein expression already occurring during neuralization and in spinal MN progenitors, and revealed the largest differences between protein expression profiles in mature MNs. In particular, proteins associated with ER to Golgi vesicle-mediated transport were upregulated and mRNA splicing and polyubiquitination were downregulated, consistent with previous studies describing defects in late stages of MNs [15, 34, 35, 43, 44]. Here we provide evidence of altered proteomic changes in these pathways at early developmental stages. This could serve as valuable resource of potential targets for early treatment in SMA to reduce the progression of symptoms.

We showed heterogeneity in SMN protein levels among SMA patients during MN differentiation and to a lesser extent between the two control lines. Differences in SMN protein level of MN cultures were previously observed as well, indicating that SMN heterogeneity is not restricted to SMA [27]. Interestingly, here we note that the difference in SMN levels between SMA and controls started to increase from T7 onwards, when dendrites and axons start to develop and mature. This may suggest that difference in SMN levels in SMA compared to controls may be restricted to later stages of MN differentiation. While a previous study reported such a delayed neurite outgrowth, and decreased neurites in iPSC derived MNs from SMA patients, the difference in SMN was not obvious, due to the low percentage of MNs [45]. This may indicate MN vulnerability to lower levels of SMN. Furthermore, decreased SMN levels in *in vivo* mouse MNs were observed only in SMA mice during maturation whereas the SMN levels of the controls remained stable [46]. These results may indicate that the level of SMN is important during MN maturation.

When we compared protein expression trends during MN development in each SMA iPSC line separately, we found different proteins being regulated between the SMA samples. However, these proteins were largely involved in similar biological processes. Moreover, these processes were more strongly enriched at later stages of development, namely during MN maturation. Since there is selective degeneration of lower MNs in SMA patients with reduced SMN protein expression, and given that SMN has a neuron-specific role in mRNA processing, it is not surprising that the difference between SMA and healthy controls on the corresponding proteins is increased during the generation of MNs [47, 48].

106

Efforts to improve SMA therapy can potentially benefit from combinatorial strategies next to SMN-elevating therapeutics and could benefit from exploring novel targets at early stages of disease. In our dataset we therefore investigated the protein expression behavior of known SMN-binding partners during MN differentiation. This identified several SMN-binding partners that show significant alterations in SMA and are potential useful candidates to consider further. Especially SMNDC1, which is a paralog of SMN1 and is a constituent in the spliceosome complex, would be an interesting candidate for further study in relation to SMA [49]. Furthermore, the expression behavior of the here highlighted proteins should be further studied in the light of the disease-ameliorating effect of increasing SMN levels, using compounds such as MLN4924 and splicing modulator C3, which were previously used to increase SMN levels in SMA, [27, 50-52]. This combination may prioritize protein candidates for targeting in combinatorial therapy strategies.

Despite the fact that SMA is being considered to mainly affect spinal MNs, recent studies have suggested that other neuronal subtypes might be affected as well, such as pyramidal neurons and hippocampal neurons in the cortex [53, 54]. Therefore, studying different neuronal subtypes, and 3D organoid models will be essential to analyze the development and neuronal maturation phenotypes in SMA. In addition, given that human iPSC-derived MNs were recently co-cultured with endothelial cells on a microfluidic 'organ-on-chip' platform to model cellular interactions involved in MN diseases [55], SMA iPSC-derived spinal MNs grown on an organ-on-chip could model SMA closer to an *in vivo* condition and advance our understanding of disease origin and progression.

In conclusion, iPSCs generated from SMA patients were differentiated here into spinal MNs in culture to examine proteins with altered neurodevelopmental signatures. This revealed that, early time points during MN differentiation can be used both to further examine the role of SMN in neuronal development and to screen drugs for novel therapeutics before the onset of disease.

Materials and methods

Ethic statement

Skin fibroblasts were obtained from 2 SMA type I patients and 2 healthy controls and stored at -80 °C. A summary of the patients and control lines can be found in **Supplementary Table 1**. Control fibroblasts were provided by Dr. Vivi M. Heine (VU University, Amsterdam, the Netherlands). Both SMA patients were type I defined by the presence of 2 copies of SMN2. All the cell lines and protocols in this study were

carried out in accordance with the guidelines approved by the Medical Ethical Committee of the University Medical Centre Utrecht.

Generation of iPSCs

Primary human fibroblasts were maintained in mouse embryonic fibroblast (MEF) containing Dulbecco's modified Eagle's medium glutamax (Life Technologies), supplemented with 10% fetal bovine serum (Sigma-Aldrich), and 1% penicillin/streptomycin (Life Technologies). Viral transduction was performed as described previously [56]. Briefly, a lentiviral vector expressing OCT4, KLF4, SOX2, c-MYC and a mixture containing MEF medium and 4 mg/ml hexadimethrine bromide (Sigma) was used. Cells were incubated for 24 hours in this mixture followed by MEF medium for 5 days. Hereafter, cells were transferred to irradiated MEFs in human embryonic stem cell (huES) medium containing DMEM-F12 (Life Technologies), knockout serum replacement (Life Technologies), penicillin/streptomycin, L-glutamine (Life Technologies), non-essential amino acids (Life Technologies), β-mercaptoethanol (Meck Millipore) and 20 ng/ml recombinant human fibroblast growth factor-basic (bFGF; Life Technologies). Potential iPSC colonies were selected on the basis of their embryonic stem cell like morphology. Feeder-free iPSCs were cultured on Geltrex coated dishes (Life Technologies) in mTeSR1 medium (Stem Cell Technologies) and passaged enzymatically with Accutase (Innovative Life Technologies). All cell lines were tested for mycoplasma contamination every other week.

Karyotype

All iPSC lines were incubated for 30 min at 37°C in Colcemid (100 ng/mL; Life Technologies) and dissociated with trypsin (TrypLE) for 10 min. Following this, cells were washed with phosphate buffered Seline (PBS) and incubated for 30 min in 5 mL hypotonic solution (1 g potassium chloride, 1 g sodium citrate in 400 mL H2O). This was followed by centrifugation for 3 min at 1500 rpm and fixation for 5 min at room temperature with methanol: acetic, 3:1. Cells were then resuspended and submitted for G-Band karyotyping.

Motor neuron differentiation

Motor neuron differentiation was performed using a slightly modified version of previously described protocol [30]. Briefly, on day 0, iPSCs were gently lifted by Accutase treatment for 5 min at 37 °C and resuspended in differentiation medium (DMEM F-12, Neurobasal vol:vol, N2 supplement (Life Technologies), B27 without vitamin A (Life Technologies), Pen-strep 1%, ascorbic acid 0.5 µM (Sigma-Aldrich) and 5 µM Y27632 (STemGent). Embryoid bodies (EBs) were formed through a standardized

microwell assay by seeding at a density of 150 cells/microwell in differentiation medium [57]. For neutralization, dual-SMAD signaling was inhibited for four days with 3 µM Chir-99021 (Tocris), 0.2 µM LDN193189 (Miltenyi Biotec, Bergisc Gladbach) and 40 µM SB-431542 (Axon Medchem). Hereafter, EBs were flushed out and transferred to a non-adherent 10 cm petri dish (Greiner Bio-one) in differentiation medium with 500 nM Smoothened Agonis; SAG (Calbiochem) and 100 nM retinoic acid; RA (Sigma-Aldrich). On day 9, 10 µM DAPT (Tocris) was added to the medium and on day 10, 20 ng/ml BDNF (Peprotech) and 10 ng/ml GDNF (Peprotech) were added. Medium was changed every other day. On day 15, EBs were dissociated into single cells using Papain (Worthington Biochemical Corporation) and DNAse (Worthington Biochemical Corporation). Cells were plated on PDL (20 µg/ml, Sigma-Aldrich) and laminin (5 µg/ml, Invitrogen) coated coverslips at 60-70% confluency.

Immunohistochemistry

Cells were fixed with 4% paraformaldehyde (10 min at room temperature) and rinsed with PBS. For intracellular stainings, cells were permeabilized with 0.1% Triton X-100 (Sigma-Aldrich) and blocked with 20% goat serum in 2% BSA/PBS (45 min at room temperature). Primary antibodies were diluted in 2% FBS/0.1% Triton in PBS (overnight at 4 °C). The following primary antibodies were used: rabbit anti-Tubulin-β3 (Sigma-Aldrich) and mouse anti-Isl-1 (DSHB). After a washing step with PBS, secondary antibodies were added (1 hour at room temperature). The following secondary antibodies were used: Streptavidin-Alexa555, Alexa499 and Alexa568 (Invitrogen). Cells were then washed and mounted with Prolong Gold reagent with Dapi (Invitrogen). Samples were visualized and imaged on Zeiss AxioScope microscope and images were exported and analyzed with Photoshop CS5.

Electrophysiological recordings

We performed electrophysiological recordings on the MNs at day 25 after differentiation as described above. Individual MNs were selected for patch clamp recordings and bathed in artificial cerebrospinal fluid containing (in mM) 120 NaCl, 3.5 KCl, 1.3 MgSO4, 1.25 NaH2PO4, 2.5 CaCl2, 10 D-glucose and 25 NaCO3. We used an upright microscope (Axoiskop, Zeiss) and intracellular recordings were obtained using 4-5 MΩ borosilicate glass pipettes filled with an internal solution containing (in mM) 140 K-methanesulfonate, 10 HEPES, 0.1 EGTA, 4 MgATP, 0.3 NaGTP. Traces were collected using an Axopatch 200 amplifier (Molecular Devices), filtered with a 5 kHz filter, digitalized at 10 kHz using a Digidata 1322A (Axon Instrumments, USA) and analyzed on a PC using pClamp 9.0 and Clampfit 9.2 (Axon Instruments). Recordings with a series resistance < 2.5 times the pipette resistance were accepted. Cells were depolarized to induce spike trains in 10 steps of 10 nA with an interval of 30 seconds and duration of 500 ms.

Proteomics

Cell lysis and protein digestion

Samples were collected at days 0, 2, 6, 8, 10, 13, 15, 20, 22 and 24 from two healthy and two SMA biological replicates. Cells were lysed in lysis buffer containing 8 M urea in 50 mM ammonium bicarbonate (pH 8.0), 1 complete mini protease inhibitor (Roche) and phosphoSTOP phosphatase inhibitor mixture (Roche). Cells were sonicated on ice and debris was removed by centrifugation at 2000 g for 15 min at 4 °C. Protein concentration was determined with Bradford Assay (BioRad) followed by reduction with 4 mM DTT (25 min at 56 °C) and alkylation with 8 mM Iodoacetamide (30 min at room temperature in dark). Proteins were digested into peptides using 1 µg Lys-C per 75 µg protein (4 hours at 37 °C). The solution was diluted to a final Urea concentration of 2 M with 50 mM ammonium bicarbonate and further digested with 1 µg trypsin per 100 µg protein (overnight at 37 °C). The digestion was quenched with 5% formic acid and peptides were desalted using Sep-Pak C18 cartridges (Waters) and vacuum centrifuged to dryness.

Tandem Mass Tag (TMT) 10-plex labelling

Digested aliquots of ~ 100 µg of each sample were chemically labeled according to instructions outlined in the TMT reagent labeling kit (Thermo Fisher). Each label reagent tag was assigned to samples illustrated in **Supplementary Table 4**. Peptides were resuspended in 80 µl resuspension buffer containing 50 mM HEPES buffer and 12.5% acetonitrile (ACN, pH 8.5). TMT reagents (0.8 mg) were dissolved in 80 µl anhydrous ACN of which 20 µl was added to the peptides. Following incubation at room temperature for 1 hour, the reaction was then quenched using 5% hydroxylamine in HEPES buffer for 15 min at room temperature. The TMT-labeled samples were pooled with equal protein ratio, followed by vacuum centrifuge to near dryness and desalting using Sep-Pak C18 cartridges.

Off-line basic pH fractionation

Peptides were separated by basic pH Reverse Phase HPLC. Samples were solubilized in buffer A (5% ACN, 10 mM ammonium bicarbonate, pH 8.0) and subjected to a 50 min linear gradient from 18% to 45% ACN in 10 mM ammonium bicarbonate pH 8 at flow rate of 0.8 ml/min. An Agilent 1100 pump equipped with a degasser and a photodiode array (PDA) detector was used with an Agilent 300 Extend C18 column (5 µm particles, 4.6 mm inner diameter, and 20 cm length). The peptide mixture was then fractionated into 96 fractions and consolidated into 24 fractions. Samples were acidified with 10% formic acid and vacuum-dried followed by redissolving with 5% formic acid/5% ACN for LC-MS/MS processing.

Mass spectometry analysis

Each fraction was analysed by nanoLC ESI MSMS using an Orbitrap Fusion (Thermos Fisher Scientific) coupled to an Agilent 1290 HPLC system (Agilent Technologies). Peptides were separated on a double frit trap column of 20 mm x 100 µm inner diameter (ReproSil C18, Dr Maisch GmbH, Ammerbuch, Germany). This was followed by a 40 cm x 50 µm inner diameter analytical column (ReproSil Pur C18-AQ (Dr Maisch GmbH, Ammerbuch, Germany). Both columns were packed in house. Trapping was done at 5 µl/min in 0.1 M acetic acid in H2O for 10 min and the analytical separation was done at 100 nl/min for 2 hours by increasing the concentration of 0.1 M acetic acid in 80% acetonitrile (v/v). The instrument was operated in a data-dependent mode to automatically switch between MS and MS/MS. Full-scan MS spectra were acquired in the Orbitrap from m/z 350-1500 with a resolution of 60,000 FHMW, automatic gain control (AGC) target of 400 000 and maximum injection time of 50 ms. For the MS/MS analysis, the ten most intense precursors at a threshold above 5,000 were selected for MS/MS with an isolation width of 0.7 Th after accumulation to a target value of 30,000 (maximum injection time was 115 ms). Fragmentation was carried out using higher-energy collisional dissociation (HCD) with collision energy of 38% and activation time of 0.1 ms. Fragment ion analysis was performed on Orbitrap with resolution of 60,000 FHMW and a low mass cut-off setting of 120 m/z. Data were acquired using Xcalibur software (Thermo Scientific).

Data processing

To process MS raw files, we employed Proteome Discover (version 2.2, Thermo Scientific). Peak list was searched using Swissprot database (version 2017_02) with the search engine Mascot (version 2.3, Matrix Science). Enzyme specificity was set to trypsin and allowed cleaving N-terminal to proline up to two missed cleavages. Peptides had to have a minimum length of seven amino acids to be considered for identification. Taxonomy was chosen for Homo sapiens and precursor mass tolerance was set to 50 ppm with 0.05 Da fragment mass tolerance. TMT tags on lysine residues and peptide N termini (+229.163 Da) and oxidation of methionine residues (+15.995 Da) were set as dynamic modifications, while carbamidomethylation on cysteine residues (+57.021 Da) was set as static modification. For the reporter ion quantification, integration tolerance was set to 20 ppm with the most confident centroid method. The mass analyzer was done with FTMS with MS2 order. Activation type was done with HCD with minimum collision energy of 0 and Maximum of 1000. Results were filtered with a Mascot score of at least 20 and Percolator was used to adjust the peptide-spectrum matches (PSMs) to a false discovery rate (FDR) below 1%.

Bioinformatics analysis

The open PERSEUS environment was used for statistical and bioinformatics analysis and to generate the plots and figures. For several plots we also used GraphPad Prism (version 7.04). To compare the relative protein ratios within samples, the values of each time point were normalized to the reference value of T1 (iPSCs) and log2 transformed. All peptide ratios were then normalized against the median. To identify the most discriminating proteins between SMA and controls, we applied a t test statistics with permutation-based FDR of 5% and S0 of 0.1 (the S0 parameter sets a threshold for minimum fold change [58]) The significantly enriched proteins were then analyzed for annotation enrichments for Gene Ontology (GO) using DAVID database [59]. Network analysis was performed using Cytoscape [60] with GeneMania plugin [61].

References

1. Lefebvre, S., et al., *Identification and characterization of a spinal muscular atrophy-determining gene.* Cell, 1995. **80**(1): p. 155-65.
2. Nash, L.A., et al., *Spinal Muscular Atrophy: More than a Disease of Motor Neurons?* Curr Mol Med, 2016. **16**(9): p. 779-792.
3. Crawford, T.O. and C.A. Pardo, *The neurobiology of childhood spinal muscular atrophy.* Neurobiol Dis, 1996. **3**(2): p. 97-110.
4. Lefebvre, S., et al., *Correlation between severity and SMN protein level in spinal muscular atrophy.* Nat Genet, 1997. **16**(3): p. 265-9.
5. Lorson, C.L., et al., *SMN oligomerization defect correlates with spinal muscular atrophy severity.* Nat Genet, 1998. **19**(1): p. 63-6.
6. Pellizzoni, L., B. Charroux, and G. Dreyfuss, *SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins.* Proc Natl Acad Sci U S A, 1999. **96**(20): p. 11167-72.
7. Chang, H.C., et al., *Degradation of survival motor neuron (SMN) protein is mediated via the ubiquitin/proteasome pathway.* Neurochem Int, 2004. **45**(7): p. 1107-12.
8. McAndrew, P.E., et al., *Identification of proximal spinal muscular atrophy carriers and patients by analysis of SMNT and SMNC gene copy number.* Am J Hum Genet, 1997. **60**(6): p. 1411-22.
9. Lorson, C.L., et al., *A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy.* Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6307-11.
10. Monani, U.R., et al., *A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2.* Hum Mol Genet, 1999. **8**(7): p. 1177-83.
11. Cho, S. and G. Dreyfuss, *A degron created by SMN2 exon 7 skipping is a principal contributor to spinal muscular atrophy severity.* Genes Dev, 2010. **24**(5): p. 438-42.
12. Le, T.T., et al., *SMNDelta7, the major product of the centromeric survival motor neuron (SMN2) gene, extends survival in mice with spinal muscular atrophy and associates with full-length SMN.* Hum Mol Genet, 2005. **14**(6): p. 845-57.
13. Liu, Q., et al., *The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins.* Cell, 1997. **90**(6): p. 1013-21.
14. Pellizzoni, L., et al., *A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing.* Cell, 1998. **95**(5): p. 615-24.
15. Burghes, A.H. and C.E. Beattie, *Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick?* Nat Rev Neurosci, 2009. **10**(8): p. 597-609.
16. Matera, A.G., R.M. Terns, and M.P. Terns, *Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs.* Nat Rev Mol Cell Biol, 2007. **8**(3): p. 209-20.
17. Sleigh, J.N., T.H. Gillingwater, and K. Talbot, *The contribution of mouse models to understanding the pathogenesis of spinal muscular atrophy.* Dis Model Mech, 2011. **4**(4): p. 457-67.
18. Fan, L. and L.R. Simard, *Survival motor neuron (SMN) protein: role in neurite outgrowth and neuromuscular maturation during neuronal differentiation and development.* Hum Mol Genet, 2002. **11**(14): p. 1605-14.
19. Chan, Y.B., et al., *Neuromuscular defects in a Drosophila survival motor neuron gene mutant.* Hum Mol Genet, 2003. **12**(12): p. 1367-76.
20. Sasongko, T.H., et al., *Hypomutability at the polyadenine tract in SMN intron 3 shows the invariability of the α-SMN protein structure.* Ann Hum Genet, 2008. **72**(Pt 2): p. 288-91.
21. McWhorter, M.L., et al., *Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding.* J Cell Biol, 2003. **162**(5): p. 919-31.
22. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.

23. Ebert, A.D. and C.N. Svendsen, *Stem cell model of spinal muscular atrophy.* Arch Neurol, 2010. **67**(6): p. 665-9.

24. Sareen, D., et al., *Inhibition of apoptosis blocks human motor neuron cell death in a stem cell model of spinal muscular atrophy.* PLoS One, 2012. **7**(6): e39113.

25. Barrett, R., et al., *Reliable generation of induced pluripotent stem cells from human lymphoblastoid cell lines.* Stem Cells Transl Med, 2014. **3**(12): p. 1429-34.

26. Fuller, H.R., et al., *Spinal Muscular Atrophy Patient iPSC-Derived Motor Neurons Have Reduced Expression of Proteins Important in Neuronal Development.* Front Cell Neurosci, 2015. **9**: p. 506.

27. Rodriguez-Muela, N., et al., *Single-Cell Analysis of SMN Reveals Its Broader Role in Neuromuscular Disease.* Cell Rep, 2017. **18**(6): p. 1484-1498.

28. Harschnitz, O., et al., *Autoantibody pathogenicity in a multifocal motor neuropathy induced pluripotent stem cell-derived model.* Ann Neurol, 2016. **80**(1): p. 71-88.

29. Rademacher, S., et al., *Metalloprotease-mediated cleavage of PlexinD1 and its sequestration to actin rods in the motoneuron disease spinal muscular atrophy (SMA).* Hum Mol Genet, 2017. **26**(20): p. 3946-3959.

30. Maury, Y., et al., *Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes.* Nature Biotechnology, 2014. **33**: p. 89.

31. Fuller, H.R., et al., *Valproate and bone loss: iTRAQ proteomics show that valproate reduces collagens and osteonectin in SMA cells.* J Proteome Res, 2010. **9**(8): p. 4228-33.

32. Brandes, G., et al., *Metalloprotease-mediated cleavage of PlexinD1 and its sequestration to actin rods in the motoneuron disease spinal muscular atrophy (SMA).* Human Molecular Genetics, 2017. **26**(20): p. 3946-3959.

33. Li, H., et al., *alpha-COP binding to the survival motor neuron protein SMN is required for neuronal process outgrowth.* Hum Mol Genet, 2015. **24**(25): p. 7295-307.

34. Powis, R.A., et al., *Systemic restoration of UBA1 ameliorates disease in spinal muscular atrophy.* JCI Insight, 2016. **1**(11): p. e87908.

35. Ng, S.-Y., et al., *Genome-wide RNA-Seq of Human Motor Neurons Implicates Selective ER Stress Activation in Spinal Muscular Atrophy.* Cell Stem Cell, 2015. **17**(5): p. 569-584.

36. Groen, E.J.N., K. Talbot, and T.H. Gillingwater, *Advances in therapy for spinal muscular atrophy: promises and challenges.* Nat Rev Neurol, 2018. **14**(4): p. 214-224.

37. Wu, C.-H., et al., *Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis.* Nature, 2012. **488**(7412): p. 499-503.

38. Abera, M.B., et al., *ML372 blocks SMN ubiquitination and improves spinal muscular atrophy pathology in mice.* JCI Insight, 2016. **1**(19): p. e88427.

39. Shorrock, H.K., et al., *UBA1/GARS-dependent pathways drive sensory-motor connectivity defects in spinal muscular atrophy.* Brain, 2018. **141**(10): p. 2878-2894.

40. Wishart, T.M., et al., *Dysregulation of ubiquitin homeostasis and beta-catenin signaling promote spinal muscular atrophy.* J Clin Invest, 2014. **124**(4): p. 1821-34.

41. Powis, R.A., et al., *Systemic restoration of UBA1 ameliorates disease in spinal muscular atrophy.* JCI insight, 2016. **1**(11): p. e87908-e87908.

42. Monani, U.R., *Spinal muscular atrophy: a deficiency in a ubiquitous protein; a motor neuron-specific disease.* Neuron, 2005. **48**(6): p. 885-96.

43. Peter, C.J., et al., *The COPI vesicle complex binds and moves with survival motor neuron within axons.* Human Molecular Genetics, 2011. **20**(9): p. 1701-1711.

44. Pellizzoni, L., J. Yong, and G. Dreyfuss, *Essential role for the SMN complex in the specificity of snRNP assembly.* Science, 2002. **298**(5599): p. 1775-9.

45. Boza-Moran, M.G., et al., *Decay in survival motor neuron and plastin 3 levels during differentiation of iPSC-derived human motor neurons.* Sci Rep, 2015. **5**: p. 11696.

46.   Groen, E.J.N., et al., *Temporal and tissue-specific variability of SMN protein levels in mouse models of spinal muscular atrophy.* Hum Mol Genet, 2018. **27**(16): p. 2851-2862.

47.   Rossoll, W., et al., *Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of beta-actin mRNA in growth cones of motoneurons.* J Cell Biol, 2003. **163**(4): p. 801-12.

48.   Carrel, T.L., et al., *Survival motor neuron function in motor axons is independent of functions required for small nuclear ribonucleoprotein biogenesis.* J Neurosci, 2006. **26**(43): p. 11014-22.

49.   Neubauer, G., et al., *Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex.* Nat Genet, 1998. **20**(1): p. 46-50.

50.   Naryshkin, N.A., et al., *Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy.* Science, 2014. **345**(6197): p. 688-93.

51.   Soucy, T.A., et al., *An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer.* Nature, 2009. **458**(7239): p. 732-6.

52.   Emanuele, M.J., et al., *Global identification of modular cullin-RING ligase substrates.* Cell, 2011. **147**(2): p. 459-74.

53.   Taylor, A.S., et al., *Restoration of SMN to Emx-1 expressing cortical neurons is not sufficient to provide benefit to a severe mouse model of Spinal Muscular Atrophy.* Transgenic Res, 2013. **22**(5): p. 1029-36.

54.   Wishart, T.M., et al., *SMN deficiency disrupts brain development in a mouse model of severe spinal muscular atrophy.* Hum Mol Genet, 2010. **19**(21): p. 4216-28.

55.   Sances, S., et al., *Human iPSC-Derived Endothelial Cells and Microengineered Organ-Chip Enhance Neuronal Development.* Stem Cell Reports, 2018. **10**(4): p. 1222-1236.

56.   Warlich, E., et al., *Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming.* Mol Ther, 2011. **19**(4): p. 782-9.

57.   Rivron, N.C., et al., *Tissue deformation spatially modulates VEGF signaling and angiogenesis.* Proc Natl Acad Sci U S A, 2012. **109**(18): p. 6886-91.

58.   Tusher, V.G., R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response.* Proc Natl Acad Sci U S A, 2001. **98**(9): p. 5116-21.

59.   Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

60.   Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.

61.   Montojo, J., et al., *GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop.* Bioinformatics, 2010. **26**(22): p. 2927-8.
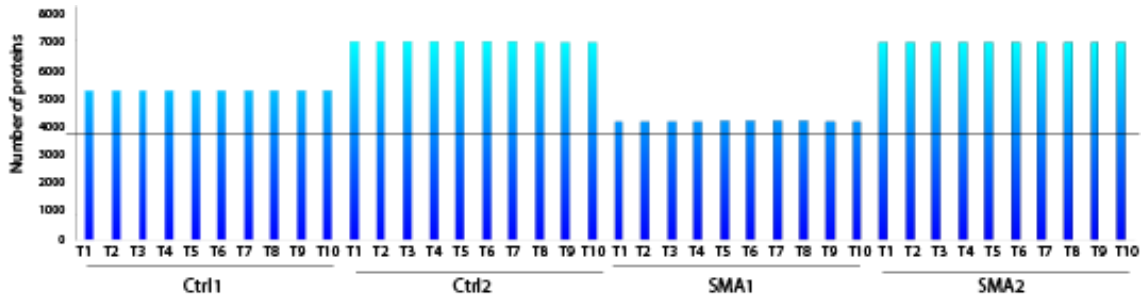
Supplementary data

Figure S1. Protein identification. Bar graph showing the number of proteins identified in each biological replicate and in each time point.
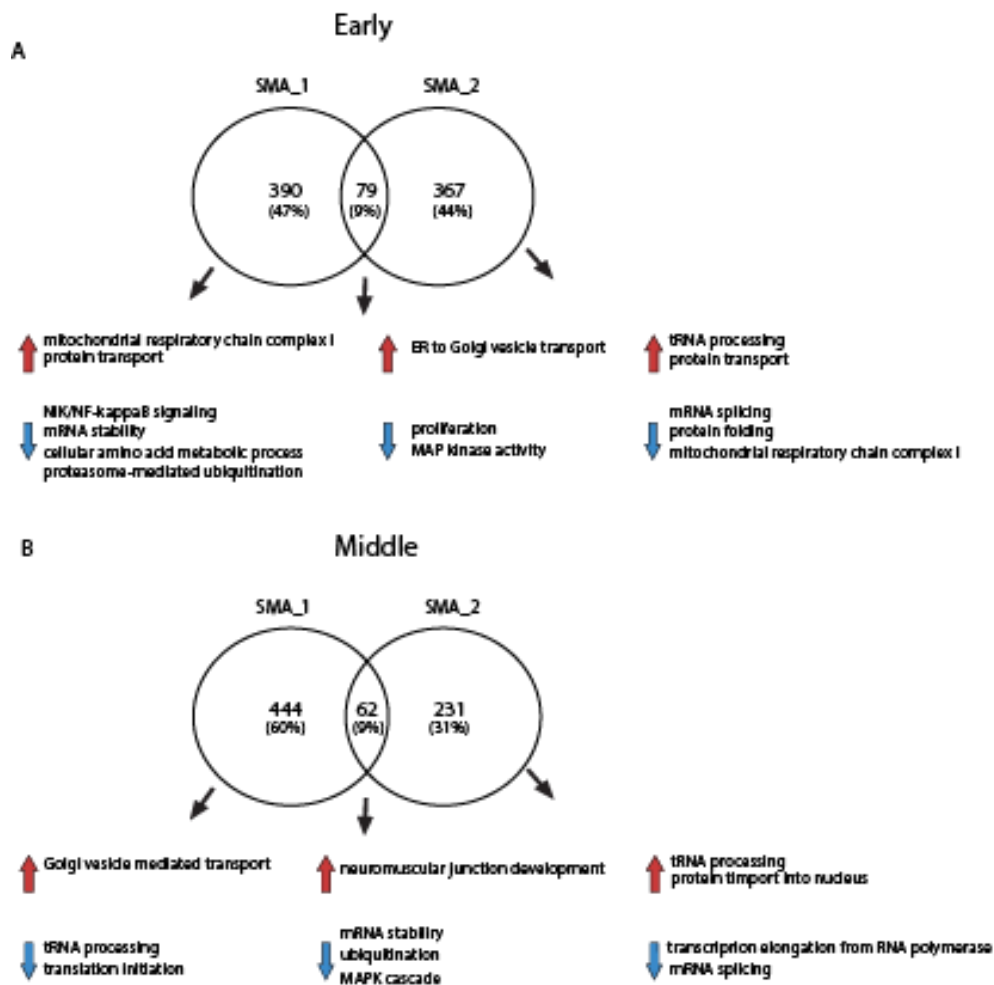
Figure S2. Difference between the SMA patients. (A) Venn diagram illustrating the number of differentially expressed proteins in the two SMA lines compared to healthy controls during the neutralization (early) and (B) during the spinal MN progenitor (middle) time points. GO analysis with respect to biological processes are shown in red for the upregulated and in blue for the downregulated proteins.

Supplementary Table 1. Information of patient and control lines used for this study.

Donor	Condition	Gender	Age at biopsy
Control-I	Healthy	M	20 days
Control-II	Healthy	M	74 days
SMA-I	SMA type I	F	109 days
SMA-II	SMA type I	M	4 months

Outlook

Future outlook

Crosstalk between multi-omic studies may enhance our understanding of the biological mechanism.

In the last few years, mass spectrometry-based proteomics has matured and is being routinely used to identify and quantify thousands of proteins. However proteomics alone, as it is done in this thesis, can only provide limited insight into the biological mechanisms during development and in disease. In order to understand a complex system more thoroughly, integration of multi-omics approaches are needed. In multi-omics studies, different molecular layers (e.g. genomics, transcriptomics, proteomics and metabolomics) are combined. The integration of such cross talk between multiple molecular layers may enhance our understanding of the molecular dynamics. Integrating multi-omics approaches are becoming popular, however this has several challenges [1]. Because a huge amount of data that is being generated by these techniques, integrating these data together is a complex procedure and challenging. Therefore, a collaboration within the different multi-omic layers is essential. I believe that the scientific research community should invest in subject areas such as statistics and bioinformatics tools to fully use and understand the big data.

In addition, it is important that in a single multi-omics study, the same sample is measured and analyzed to further understand the relationship between the different omics layers. Technological advances in genomic and transcriptomic field now allow to measure multiple 'omes' from the same cell [2, 3]. With measurements from the same cell, correlative evidence between genetic and transcriptomic diversity can be explained. A study comparing the transcriptome / proteome was done by Schwanhausser *et al* (2011) in mouse fibroblasts showing a moderate correlation (R^2 = 0.41) between mRNA and protein [4]. This poor correlation could be explained due to regulatory processes such as protein degradation and post-translational modifications [5, 6]. By comparing genomics, transcriptomics with proteomics, one has to take into account that in next generation sequencing (NGS) the whole genome and total RNA is sequenced. This allows the generation of near complete genetic variations and its transcribed repertoire, however, this is not the case for proteomics [7]. A recent study by Kim *et al*, (2014) demonstrated a 'complete' human proteome using 30 human samples and over 200 cell types and identified proteins corresponding to approximately 84% of the annotated protein-coding genes. The not observed proteins could be explained by novel protein-coding genes in the human genome that are missing from current genome annotations due to translation of pseudogenes as well as non-coding RNAs [8]. Furthermore, because proteins can span 10 orders of magnitude in abundance, it is difficult to detect the low abundant

proteins. I believe that further into the future, we should be able to obtain multi-omics measurements derived from the same single cell. These experiments will combine omics-measurements from the same single cell with to fully provide knowledge of the cell constituents [9, 10]. This will allow us for example to determine the identity of a cell, discriminate cells from healthy and disease, study dynamics during cell differentiation and to discover new cell types.

Spatial and temporal resolution

When focusing on the brain which consist of more than 100 billion neurons and even more interactions between each other, it is very challenging to study the function of this complex organ. Over the decades, different technological approaches have been widely used to study the brain using for example several imaging tolls such as magnetic resonance imaging (MRI), electron microscopy (EM) and immunohistochemistry. On the molecular level, the study of neuroscience has also emerged. Besides genomics and transcriptomics, proteomics is also used and has its own niche called *neuroproteomics* [11]. There are several challenges however in this field. The enormous heterogeneity of different brain regions as well as different neuronal subpopulations increase the spatial complexity that needs attention [12]. Therefore, several sorting tools such as laser capture microdissection and fluorescence activated cell sorting (FACS) are often used to enrich a specific area or cells. Because MS-based proteomics requires larger starting material compared to genomics and transcriptomics, this can be a challenge for studies with low input material. For example in the study of low abundant sub-populations or smaller cell compartments such as the post-synaptic density, dendrites or axons.

In addition, the proteome of the brain can be studied in time series experiments (temporal resolution) such as cell differentiation, tissue development and disease progression [11]. For instance, different time point samples were taken during neuronal development to visualize different clusters of proteins having a specific expression profile [13]. However, one needs to take into account that the cells are not all synchronized in similar way. One cell has the potential to differentiate faster than the other in the same culture dish. The derived clusters of different protein expressions is therefore an average of multiple cells. Therefore, increasing temporal resolution by taking more time points during neuronal development, for example in **chapter 2**, will add little information. Ideally, to achieve a high temporal resolution, the study of single cell omics approaches is needed to include unique expression profiles on neuronal differentiation. A large majority of studies on single neurons is focused on molecular markers for subtype specification rather than on developmental time lines [14, 15]. A temporal study on the development of mouse spinal cord was done by William A. Alaynick *et al,* showing the transcriptomic

profile of mitotic and postmitotic neurons [16]. Furthermore, Ludovic Telley *et al,* performed a single cell transcriptomic study showing a series of transcriptional waves providing discrete time window during the development of newborn neurons in the mouse neocortex [17]. These results could be used as a genetic targets for the identification or directed differentiation of progenitors and neurons. To enrich for a subpopulation of cells during neuronal development, FACS is commonly used to isolate the cells with specific markers. For example, specific set of progenitor markers (e.g. Vimentin, Nestin) could be used to enrich the immature neurons and postmitotic markers (e.g. NeuN, PSD-95) could be used to enrich the fully differentiated neurons. There are few caveats to consider when using this enrichment tool to enrich a subpopulation. One major caveat may concern the morphology disruption of a neuron. In a differentiated neuronal cell culture, dendrites and axons are heavily interconnected with other (distant) neurons. Dissociating neurons into single cells and cell sorting them will disrupt its cellular connectivity and complex morphology. Therefore, unique molecular markers that encode for a specific timepoint in neuronal differentiation is still lacking and further research is required. Several groups presented protocols for extraction of intact cells from the brain using a vibratome and manual dissection of a brain region. This was followed by enzymatic digestion, which loosens the extracellular matrix and a mechanical pipetting pieces of tissues into single cells [18, 19]. Although these protocols are often used to isolate specific neuronal populations from the brain or CNS, they cannot isolate a single neuron without disrupting the complex morphology. Furthermore, other studies seed the cells as 'single cell' for further differentiation, however this is challenging in the field of neuronal differentiation because neurons don't differentiate as single cells and need each other for proper development [20, 21]. Furthermore, cell type-specific differences such as cell cycle state (e.g. proliferating or G0) must be carefully considered when comparing the expression patterns collected from multiple cells. Overall, to detect low input material, technological improvements in the MS workflow as well as in the cell culture system needs attention.

Human neuronal development

The neurodevelopmental study in **chapter 2,** characterizing human iPSC derived neurons during differentiation, would further benefit from a comparison with the development of neurons from other species. Rodent or primate derived iPSCs for example could be differentiated towards neurons and in parallel compared to the human derived neurons. This would emphasize human specific properties during neuronal development. Specific proteomic waves with discrete time window or group of proteins with altered expression during differentiation might provide evidence unique to human neuronal development. In the field of developmental neuroscience rodents are frequently used and several studies

compared the differences between human and rodent brain in terms of maturation, functionality and electrophysiological differences [22, 23]. They provided evidence of complexity in developmental timing and more spontaneous and robust electrical activity in rodents compared to human. Therefore, proteomic research in both human and other species in parallel during differentiation will provide a greater understanding of human neuronal development.

Neurodevelopmental disorders

Most studies performed on neurodevelopmental disorders have been in the field of genomics rather than proteomics. Recently, mass spectrometry-based proteomic studies have been used on neurodevelopmental disorders in the search for diagnostic and prognostic biomarkers [24, 25]. Because proteins are the effectors of many biological and pathological processes, search for protein biomarkers and drug targets could provide future therapeutics. Therefore, we used our mass-spectrometry proteomic approach to study SMA and Rett syndrome.

In **chapter 3** we studied RTT syndrome which is a neurodevelopmental disorder caused by a loss of function mutation in X-linked gene that encodes MeCP2 protein. Mutations in MeCP2 lead to miss-regulation of other genes including those involved in brain development and neuronal maturation. Our results indeed provided evidence of altered protein expression at early stages of neuronal development. Although at the time points that we used during neuronal differentiation dendrites or axons did not develop yet, our findings indicate that proteins involved in these processes are already expressed and altered in RTT at early developmental stages. Furthermore, we showed that proteins involved in immunity, calcium signaling and metabolism are differentially expressed at all time points of neuronal development. As there is no cure for RTT yet, this data could support the development of new biomarkers or therapeutic approaches. Currently, mutations on MeCP2 are not screened after birth, however, as we did show that the disease has its onset before the symptoms appear, efficient treatment to support neuronal stimulation and maturation could extenuate the disease onset. Since we show that a group of proteins associated to neuronal, dendrite and axon development are down regulated in RTT, stimulating the production of these proteins might have beneficial effects on neuronal development. These findings would then first have to be validated in a more complex model system. The 3D brain organoid for example is becoming an important tool in the study of brain development and neurological disorders such as autism, schizophrenia or brain defects such as caused by the Zika virus [26-28]. A recent study by Mellios *et al*, showed defects in neurogenesis and neuronal differentiation in RTT derived cerebral organoids, which is in line with our study [29, 30]. Furthermore, the brain organoids are great models that could

potentially be used in drug testing and development of innovative therapeutics. Also in terms of personalized medicine, where different patient-derived brain organoids can be tested with multiple drugs for optimal therapy selection.

In **chapter 4** we studied SMA, which is a neuromuscular disease caused by mutation in SMN gene that results in reduced levels of SMN protein. In this study, we evaluated the proteome changes during spinal motor neuron differentiation of iPSCs derived from SMA patients compared to healthy controls. We revealed altered proteomic changes already at early stages of motor neuron differentiation associated with ER to Golgi transport, mRNA splicing and ubiquitination. Dysregulation of these processes were described in previous studies at a later stage of SMA [31-33]. However, our results indicated that the altered expression of protein involved in these processes occurred already during neutralization and the spinal motor neuron progenitor stage and the severity increased over time, showing maximum levels in mature MNs. Different studies have shown that an early age treatment has beneficial effect for therapeutic success indicating that the therapeutic window is critical for effective response [34, 35]. The ultimate aim however is to diagnose SMA before onset of disease/symptoms therefore advancement in presymptomatic diagnosis and genetic newborn screening programs will be beneficial.

The current therapies for SMA in the clinical trial is based on; -SMN1 gene replacement therapy (intravenous administration using adenovirus vector), -modulating SMN2 encoded full-length protein (small molecules RG7916 and LMI070 and antisense oligonucleotide nusinersen), -neuroprotection (olesoxime) and –muscle strenghthening (CK-2127107, aminopyridine and pyridostigmine). Furthermore, SMA modifiers that target cellular processes such as; -ubiquitin homeostasis (UBA1), -actin dynamics (PLS3, ROCK), -endocytosis (NCALD), and –motor neuron stability (CTNNB1) are in preclinical stages. Very recently, the FDA (Food and Drug Administration) approved Zolgensma (made by Novartis) as the first gene therapy for SMA. This product is an adeno-associated virus vector based gene therapy targeting SMN. A one-time intravenous administration of a fully functional copy of SMN gene is being delivered into the motor neurons. This improves child's survival, muscle movement and function. However, the safely and effectiveness of this drug is still ongoing but based on clinical trial involving 36 patients, serious liver injury was one of the side effects. Current study in SMA focuses mainly on biomarkers predicting SMA disease severity and progression to determine future therapeutic trials. What is critical for biomarker screening is that the levels of such biomarkers must change in a temporally traceable and predictable manner during disease progression. Our study in measuring the proteome changes along neuronal differentiation of SMA meets this criteria. One has to further validate whether the altered proteins from

our study can be monitored and tracked in other samples such as skin biopsy or blood as well.

The data provided in our study were derived from two SMA patients and two healthy controls. As we have shown in **chapter 4**, there is a heterogeneity between the cell lines (presumably due to biological variation). In the future, I would recommend to generate isogenic control lines for each SMA line to deal with the high biological variability. This can be done for example using CRISPR/Cas9 system. Furthermore, as recent studies have suggested that other neuronal subtypes in the cortex might be affected in SMA, future study in different cell types will be essential [36, 37]. In addition, the molecular mechanisms during pathology of the neuromuscular junction (NMJ) in SMA is incompletely understood. A recent study in mice showed that SMN depletion in neonatal mice led to pathogenic changes at the NMJ while depletion in adult mice did not, highlighting for temporal requirements during the degenerative changes [38]. To further fine tune this effect in human iPSC derived samples, an *in vitro* co-culture system with motor neurons together with skeletal muscles can be used to recapitulate the formation and disruption of the NMJ in SMA.

We further compared the two SMA lines separately to the controls and noticed that, although different proteins were affected, these played a role within similar biological processes. This suggest that the search for a potential biomarkers as well as a general treatment for all SMA patients suffers from both the complexity of SMA pathology and heterogeneity between SMA patients. Novel approaches in personalized medicine in testing the different drugs for optimal therapy can be used. A better approach will also combine therapies to ameliorate the symptoms in every SMA patient. Combinatorial therapies for SMA, also termed as next-generation 'SMN-plus 'combinatorial therapies, are strongly suggested in the field for the near future to increase SMN levels, preserving neuromuscular junction and preventing motor neuron death.

In **chapter 3** and **4** we studied two different neurodevelopmental disorders. Comparing the data in **chapter 3** (RTT syndrome) with **chapter 4** (SMA), we noticed that the altered proteins in SMA are gradually up/down regulated in expression. In contrast to SMA, the altered protein expression in RTT syndrome shows a time specific expression pattern. This can be partly explained by the pathology of each disease, which develops differently but also the differences in the experimental workflow might contribute to this observation. In SMA, the time point samples are measured within one biological replicate and compared relative to the iPSC stage. This allowed us to quantitatively compare the protein changes of MN differentiation compared to their respective genetically matched iPSCs. In RTT syndrome however, less time points were taken and therefore, four time points from each RTT sample and four time points from

its isogenic controls could be measured at ones. The relative changes are compared within one TMT-10plex. Therefore the absolute intensity values are directly comparable. Future research might need to combine SMA and healthy controls within one TMT-10plex experiment to clarify whether the alteration in SMA are indeed gradually affected during neuronal development.

Trends in proteomics

Quantitative proteomics

Although MS is generally not a quantitative-based method due to differences in ionization processes, significant efforts made it possible to obtain quantitative information. Label-free quantification is currently the most widely used method in quantitative proteomics. However, this method has several major limitations such as the requirement for multiple runs which increases MS time. Another major limitation is the missing value problem between runs [39]. The group of Jürgen Cox and Matthias Mann developed tools such as MaxLFQ and BoxCar which imputes missing values and improves MS1 peak picking in samples with a high dynamic range, resulting in increased protein identification, thereby reducing the need for pre-fractionation [40, 41]. However, the missing values of peptides are still hard to interpret and MS1-based quantification suffers from a complex MS1 spectrum when multiplexing. To overcome these limitations, multiplexed proteomics using isobaric mass tags such as TMT and iTRAQ are widely used [42]. Here a major advance is that the complexity of MS1 spectra does not increase with the number of samples. Also, the ability to analyze multiple samples at ones, circumvents the missing value problem. A major advance here is the increased throughput (up to 11 samples) that can be analyzed. A drawback in multiplexed proteomics however is the co-isolation of peptides with similar m/z values, distorting the true peptide ratio. To address this problem, recently, many groups developed methods to correct for this, such as gas-phase purification by alleviating the isolated contaminants from the precursor, MS3 fragmentation and improved software [43-45]. Although multiplexed proteomics approaches are attractive for many studies, the current limit is 11 samples, which can be analyzed at ones. Therefore, this approach is not suited for studies that require comparison of hundreds up to thousands of samples. For these kind of studies, it is desired to focus on only a few (~100) proteins of interest also known as targeted proteomics [46]. Here, an additional benefit is that also low abundant proteins can be analyzed that might have been missed by shotgun proteomics.

Post-translational modifications (PTMs)

PTMs add an additional level to proteomics studies, and many methods have been developed that target different PTMs over the last decade [47-50]. There are different PTMs such as phosphorylation, ubiquitination glycosylation that regulate many biological processes. They are highly dynamic, which allows them to adapt to perturbations. The fact that multiple PTMs can be present in a protein, even at the same site, makes them very challenging to study. Over the years many enrichment methods have been developed improving their identification and quantification in MS-based proteomics. However, this still requires large amounts of starting material. In addition, PTMs have different mechanisms such as regulating the stability of a protein, the binding, localization or activity etc. To be able to understand these processes, additionally advances in bioinformatics approaches are needed. Models that for example predict functionality can then be integrated in the study of specific disorders [51].

Validation

MS-based proteomics is a powerful technology with a lot of potential in biomarker discovery, which in principle, analyzes the proteome in an unbiased and hypothesis-free manner [52]. The majority of proteomic studies have a 'discovery' role and therefore follow up validation experiments are required. Large scale shotgun proteomics experiments can be followed by targeted proteomics for verification of several candidate proteins. MRM or SRM are often used for a larger cohort since they are more selective by allowing targeted analysis of low abundant proteins. This verification step is then followed by a validation step which often consists of antibody-based methods. Immunoassays for example such as Elisa can be used in the study of thousands of samples. Such assays are restricted to only those proteins with a high antigen-antibody recognition. In the last few years improvements in the proteomics workflow, (e.g. automated) allow high throughput studies with high sensitivity. I believe that the role of MS-based proteomics studies will increase in clinical studies and in-time will partly replace other technologies, such as antibody-based methods.

References

1.   Low, T.Y., et al., *Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis.* Cell Rep, 2013. **5**(5): p. 1469-78.
2.   Dey, S.S., et al., *Integrated genome and transcriptome sequencing of the same cell.* Nat Biotechnol, 2015. **33**(3): p. 285-289.
3.   Macaulay, I.C., et al., *G&T-seq: parallel sequencing of single-cell genomes and transcriptomes.* Nat Methods, 2015. **12**(6): p. 519-22.
4.   Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.

5.   Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nat Rev Genet, 2012. **13**(4): p. 227-32.

6.   Maier, T., M. Guell, and L. Serrano, *Correlation of mRNA and protein in complex biological samples.* FEBS Lett, 2009. **583**(24): p. 3966-73.

7.   Ozsolak, F. and P.M. Milos, *RNA sequencing: advances, challenges and opportunities.* Nat Rev Genet, 2011. **12**(2): p. 87-98.

8.   Kim, M.S., et al., *A draft map of the human proteome.* Nature, 2014. **509**(7502): p. 575-81.

9.   Grun, D. and A. van Oudenaarden, *Design and Analysis of Single-Cell Sequencing Experiments.* Cell, 2015. **163**(4): p. 799-810.

10.   Stegle, O., S.A. Teichmann, and J.C. Marioni, *Computational and analytical challenges in single-cell transcriptomics.* Nat Rev Genet, 2015. **16**(3): p. 133-45.

11.   Hosp, F. and M. Mann, *A Primer on Concepts and Applications of Proteomics in Neuroscience.* Neuron, 2017. **96**(3): p. 558-571.

12.   Cagnetta, R., et al., *Rapid Cue-Specific Remodeling of the Nascent Axonal Proteome.* Neuron, 2018. **99**(1): p. 29-46.e4.

13.   Frese, C.K., et al., *Quantitative Map of Proteome Dynamics during Neuronal Differentiation.* Cell Rep, 2017. **18**(6): p. 1527-1542.

14.   Sugino, K., et al., *Molecular taxonomy of major neuronal classes in the adult mouse forebrain.* Nature Neuroscience, 2005. **9**: p. 99.

15.   Molyneaux, B.J., et al., *Neuronal subtype specification in the cerebral cortex.* Nat Rev Neurosci, 2007. **8**(6): p. 427-37.

16.   Alaynick, W.A., T.M. Jessell, and S.L. Pfaff, *SnapShot: spinal cord development.* Cell, 2011. **146**(1): p. 178-178.e1.

17.   Telley, L., et al., *Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex.* Science, 2016. **351**(6280): p. 1443.

18.   Legroux, L., et al., *An optimized method to process mouse CNS to simultaneously analyze neural cells and leukocytes by flow cytometry.* J Neurosci Methods, 2015. **247**: p. 23-31.

19.   Ho, H., et al., *A Guide to Single-Cell Transcriptomics in Adult Rodent Brain: The Medium Spiny Neuron Transcriptome Revisited.* Frontiers in Cellular Neuroscience, 2018. **12**(159).

20.   Kim, H., et al., *Single-neuronal cell culture and monitoring platform using a fully transparent microfluidic DEP device.* Scientific Reports, 2018. **8**(1): p. 13194.

21.   Gomperts, S.N., et al., *Postsynaptically Silent Synapses in Single Neuron Cultures.* Neuron, 1998. **21**(6): p. 1443-1451.

22.   Ellenbroek, B. and J. Youn, *Rodent models in neuroscience research: is it a rat race?* Disease models & mechanisms, 2016. **9**(10): p. 1079-1087.

23.   Pressler, R. and S. Auvin, *Comparison of Brain Maturation among Species: An Example in Translational Research Suggesting the Possible Use of Bumetanide in Newborn.* Frontiers in neurology, 2013. **4**: p. 36-36.

24.   Darie, C., *Mass Spectrometry and its Applications in Life Sciences*. Vol. 66. 2013.

25.   Taurines, R., et al., *Proteomic research in psychiatry.* J Psychopharmacol, 2011. **25**(2): p. 151-96.

26.   Mariani, J., et al., *FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders.* Cell, 2015. **162**(2): p. 375-390.

27.   Ye, F., et al., *DISC1 Regulates Neurogenesis via Modulating Kinetochore Attachment of Ndel1/Nde1 during Mitosis.* Neuron, 2017. **96**(5): p. 1041-1054.e5.

28.   Qian, X., et al., *Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure.* Cell, 2016. **165**(5): p. 1238-1254.

29.   Mellios, N., et al., *Human cerebral organoids reveal deficits in neurogenesis and neuronal migration in MeCP2-deficient neural progenitors.* Molecular Psychiatry, 2018. **23**: p. 791.

30. Mellios, N., et al., *MeCP2-regulated miRNAs control early human neurogenesis through differential effects on ERK and AKT signaling.* Molecular psychiatry, 2018. **23**(4): p. 1051-1065.

31. Peter, C.J., et al., *The COPI vesicle complex binds and moves with survival motor neuron within axons.* Hum Mol Genet, 2011. **20**(9): p. 1701-11.

32. Ng, S.-Y., et al., *Genome-wide RNA-Seq of Human Motor Neurons Implicates Selective ER Stress Activation in Spinal Muscular Atrophy.* Cell Stem Cell, 2015. **17**(5): p. 569-584.

33. Pellizzoni, L., J. Yong, and G. Dreyfuss, *Essential role for the SMN complex in the specificity of snRNP assembly.* Science, 2002. **298**(5599): p. 1775-9.

34. Pechmann, A., et al., *Evaluation of Children with SMA Type 1 Under Treatment with Nusinersen within the Expanded Access Program in Germany.* J Neuromuscul Dis, 2018. **5**(2): p. 135-143.

35. Pane, M., et al., *Nusinersen in type 1 SMA infants, children and young adults: Preliminary results on motor function.* Neuromuscul Disord, 2018. **28**(7): p. 582-585.

36. Taylor, A.S., et al., *Restoration of SMN to Emx-1 expressing cortical neurons is not sufficient to provide benefit to a severe mouse model of Spinal Muscular Atrophy.* Transgenic Res, 2013. **22**(5): p. 1029-36.

37. Wishart, T.M., et al., *SMN deficiency disrupts brain development in a mouse model of severe spinal muscular atrophy.* Hum Mol Genet, 2010. **19**(21): p. 4216-28.

38. Kariya, S., et al., *Requirement of enhanced Survival Motoneuron protein imposed during neuromuscular junction maturation.* J Clin Invest, 2014. **124**(2): p. 785-800.

39. Webb-Robertson, B.J., et al., *Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics.* J Proteome Res, 2015. **14**(5): p. 1993-2001.

40. Cox, J., et al., *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.* Mol Cell Proteomics, 2014. **13**(9): p. 2513-26.

41. Meier, F., et al., *BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes.* Nature Methods, 2018. **15**(6): p. 440-448.

42. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. **75**(8): p. 1895-904.

43. Wenger, C.D., et al., *Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging.* Nat Methods, 2011. **8**(11): p. 933-5.

44. Ting, L., et al., *MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics.* Nat Methods, 2011. **8**(11): p. 937-40.

45. O'Brien, J.J., et al., *Compositional Proteomics: Effects of Spatial Constraints on Protein Quantification Utilizing Isobaric Tags.* J Proteome Res, 2018. **17**(1): p. 590-599.

46. Lange, V., et al., *Selected reaction monitoring for quantitative proteomics: a tutorial.* Mol Syst Biol, 2008. **4**: p. 222.

47. Posewitz, M.C. and P. Tempst, *Immobilized gallium(III) affinity chromatography of phosphopeptides.* Anal Chem, 1999. **71**(14): p. 2883-92.

48. Villén, J. and S.P. Gygi, *The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry.* Nature Protocols, 2008. **3**: p. 1630.

49. Pinkse, M.W., et al., *Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns.* Anal Chem, 2004. **76**(14): p. 3935-43.

50. Sylvestersen, K.B., C. Young, and M.L. Nielsen, *Advances in characterizing ubiquitylation sites by mass spectrometry.* Curr Opin Chem Biol, 2013. **17**(1): p. 49-58.

51. Ochoa, D., et al., *An atlas of human kinase regulation.* Mol Syst Biol, 2016. **12**(12): p. 888.

52. Aebersold, R. and M. Mann, *Mass-spectrometric exploration of proteome structure and function.* Nature, 2016. **537**: p. 347.

Summary

Samenvatting

Summary

In this thesis we applied mass spectrometry-based proteomics to study neuronal differentiation derived from induced pluripotent stem cells (iPSCs) and to study neurodevelopmental disorders such as Spinal Muscular Atrophy (SMA) and Rett syndrome (RTT).

In **chapter one** we give a general introduction to study human brain development. We introduce the generation of iPSCs that can be differentiated into neurons for several research areas. We give an overview of several neurological disorders that are being studied using this model system. Furthermore, we give an introduction to mass spectrometry-based proteomics describing general workflows and mass spectrometers. In addition, we describe several strategies used for quantitative mass spectrometry.

In **chapter two**, we study the protein determinant during neuronal differentiation. Here, we used iPSCs and differentiated these cells into induced neurons (iN) and into small molecule-derived patterned motor neurons. We provide a mass spectrometry-based quantitative proteomic signature at temporat resolution. We identified key proteins showing significant expression changes during neuronal differentiation. Furthermore, we provide a rich source of information on several signaling pathways.

Chapter three, describes the study of RTT syndrome using iPSC-based model with isogenic controls and a time-series of mass spectrometry-based proteomic analysis during early stages of neuronal development. We provide evidence of proteomic alterations long before symptoms of RTT syndrome become apparent. Such that proteins associated with dendrite morphology and synapses are differentially expressed at early neuronal stages of development. These changes increase from early to late neuronal phases giving awareness of the protein alterations at early onset of RTT syndrome.

In **chapter four**, we again used quantitative mass spectrometry to study the proteomic changes during early stages of iPSC-derived motor neurons (MN) in SMA. Also here, we show altered proteins at early stages of MN differentiation that are associated with known SMA phenotypes such as defective ER to Golgi transport, mRNA splicing and protein ubiquitination. In addition, these proteins increase in differential expression in SMA towards later stages of MN differentiation. Furthermore, we highlight that differences in altered protein expression between SMA patients have similar biological functions. In this chapter we provide a rich source of molecular events during MN differentiation that might be interesting in the development of new biomarkers and therapeutic approaches.

Chapter five is a future outlook describing the cross-talk between the multi-omic studies to better understand a biological mechanism. We introduce several challenges in the study of human brain

development associated with temporal and spatial resolution. Furthermore we discuss the latest studies done in SMA and RTT syndrome and provide a future perspective to further improve our understanding of these devastating disorders.

Samenvatting

In dit proefschrift hebben we massa spectrometrie toegepast om de eiwitten van neuronale differentiatie te bestuderen vanuit geïnduceerde pluripotente stamcellen (iPSCs). Dit is gebruikt om neurologische aandoeningen zoals Spinal Muscular Atrophy (SMA) en Rett syndroom (RTT) te bestuderen.

In **hoofdstuk één** geven we een algemene inleiding van de ontwikkeling van het menselijk brein. We introduceren de ontwikkeling van iPSC's die worden gebruikt als model systeem om neuronen te bestuderen. We geven een overzicht van verschillende neurologische aandoeningen die worden bestudeerd met behulp van dit modelsysteem. Bovendien geven we een inleiding op massa-spectrometrie en beschrijven we de algemene stappen voor het toepassen. Daarnaast beschrijven we verschillende strategieën in de kwantitatieve massaspectrometrie.

In hoofdstuk twee bestuderen we de eiwitdeterminant tijdens neuronale differentiatie. Hier hebben we iPSC's gebruikt en deze cellen gedifferentieerd in geïnduceerde neuronen (iN) en in van kleine moleculen afgeleide patroon-motorneuronen. We bieden een op massa-spectrometrie gebaseerde kwantitatieve proteomische signatuur bij temporat-resolutie. We identificeerden belangrijke eiwitten die significante expressieveranderingen vertoonden tijdens neuronale differentiatie. Bovendien bieden we een rijke bron van informatie over verschillende signaleringsroutes.

Hoofdstuk drie beschrijft de studie van RTT-syndroom met behulp van iPSC en de 'isogenic' controles. We gebruiken een tijdreeks experiment om de eiwitten tijdens vroege stadia van neuronale ontwikkeling te bestuderen. We laten veranderde eiwitten zien die verschillend in expressie zijn in RTT lang voordat de symptomen duidelijk worden. Deze eiwitten zijn geassocieerd met dendrietmorfologie en synapsen en die zijn verschillend in expressie in de vroege neuronale stadia van ontwikkeling. Deze veranderingen nemen toe in de latere stadia van neuronale ontwikkeling. Hiermee willen we aantonen dat se eiwitten verandert kunnen zijn in het vroege stadia van RTT-syndroom.

In **hoofdstuk vier** hebben we opnieuw kwantitatieve massa spectrometrie gebruikt om de eiwit veranderingen in SMA te bestuderen. We hebben iPSC gebruikt om ze te differentieren tot motor neuronen. Ook hier laten we veranderde eiwitten zien in vroege stadia van neuronale differentiatie ie geassocieerd zijn met bekende SMA-fenotypen. Deze zijn onder anderen 'defectief ER- naar Golgi-transport', 'mRNA-splitsing' en 'ubiquitinatie'. Bovendien zijn deze eiwitten meer verandert in expressie in de latere stadia van neuronale differentiatie in SMA. Naarnaast willen we benadrukken dat verschillende veranderde eiwitten tussen tussen SMA-patiënten een vergelijkbare biologische functies

132

hebben. In dit hoofdstuk bieden we een bron van moleculaire gebeurtenissen tijdens neuronale differentiatie die misschien interessant zou kunnen zijn in de ontwikkeling van nieuwe biomarkers en behandelingen.

Hoofdstuk vijf is een toekomstperspectief in de 'multi-omic' studies en beschrijft een aantal strategien om een biologisch mechanisme beter te begrijpen. We introduceren verschillende uitdagingen om de ontwikkeling van menselijke hersenen te bestuderen. Een van de uitdalingen is de resolutie op meerdere tijdstippen en de veschillende cell types. Verder bespreken we de nieuwste studies die zijn gedaan in SMA en RTT-syndroom en bieden we een toekomstperspectief om deze zietkes beter te begrijpen.

Curriculum Vitae

Suzy Varderidou was born in Kirovakan (Armenia) and moved to the Netherlands in 1998. She finished her secondary school (VWO) in Libanon Lyceum in Rotterdam. Afterwards, she studied Biomedical Sciences at the Vrije University of Amsterdam (VU) and did her masters in Neuroscience (Amsterdam VU). Suzy started het PhD in 2014 in the group of Dr. Maarten Altelaar and Prof. Albert Heck. The results of this research are presented in this thesis. In April 2019, Suzy will continue her career in academia as a postdoctoral researcher in the group of Prof. Paul Coffer together with Dr. Magdalena Lorenowicz.