Protein-protein interactions in human pluripotent stem cell-derived neural stem cells and their neuronal progeny

DISSERTATION

Zur Erlangung des Doktorgrades (Dr. rer. nat) der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

Vorgelegt von

Jonas Martin Doerr

aus Speyer

Bonn 2014

Anfertigung mit der Genehmigung der Mathematisch-Naturwissenschaftlichen
Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn
am Institut für Rekonstruktive Neurobiologie

1. Gutachter: Prof. Dr. Oliver Brüstle

2. Gutachter: Prof. Dr. Michael Hoch

Tag der Promotion: 27.8.2014

Erscheinungsjahr: 2014

Science cannot solve the ultimate mystery of nature.

And that is because, in the last analysis,
we ourselves are a part of the mystery
that we are trying to solve.

Max Planck

1. Introduction

1.1. Discovery of protein functions

Whenever two or more proteins come into contact, they also may interact. A protein-protein interaction (PPI) occurs between individual proteins if one of them influences the other, they affect each other reciprocally or a sheer spatial recruitment takes place. These mechanisms rest upon non-covalent interactions like van der Waals forces, hydrogen bonds as well as electrostatics of surface-resident amino acid residues of involved proteins. PPIs play a key role in virtually all biological processes where proteins are involved. Amongst them are processes of signal transduction, trafficking, enzymatic degradation and transcriptional regulation (Aebersold and Mann, 2003; Fields and Song, 1989). The entire set of proteins expressed by an analyzed cell, tissue or a complete organism at a particular time is called proteome and can change in its composition due to environmental or developmental influences. The sum of all molecular interactions present in an organism or a cell - concerning a certain protein - was termed interactome, which is partly synonymous with “biological network” (Sanchez et al., 1999). Generally, an interactome describes the molecular interactions within a given proteome.

In human genomics, remarkable progress has been made by identification of mutants linked to diseases, including neurological, amongst others. While these observations provided evidence for the involvement of a certain gene product in disease processes, its molecular function and integration into protein cascades or pathways often remained unattainable. As observations linking particular mutations of a gene with a well-known interaction network to a certain disease can help to establish the mechanism of its interactions, a similar approach for genes with a yet unknown interaction network remains a challenge. One approach to solve this problem is proteomic mapping, whereby physical relationships and pathways are identified that can serve as scaffolds for subsequent functional analysis. Since the proteome of a cell is subject to change due to immense dynamics, variations and fluctuations in the protein composition, the establishment of a universal PPI network, comparable to data generated by genome sequencing, is unattainable. Currently, the only suitable approach towards the generation of a universal PPI network is the description of individual protein core complexes – protein interactions present in the majority of cell types and states, often species-spanning – that are much less susceptible to decay

or variations. These core complexes then can be connected with each other to provide a core complex network. Such approaches were previously used in simple model organisms like yeast for the determination of the composition of hundreds of protein core complexes (Benschop et al., 2010; Cusack, 1999). A characterization of protein core complexes, the identification of their dynamics and interaction with other core complexes therefore should represent the first steps in any PPI analysis as such data then can serve as foundation and scaffold for understanding transient and regulated interactions.

1.1.1. Approaches for the determination of protein functions

Historically, basic mechanisms of the developmental processes and cellular machinery were studied by disturbing existing regulatory pathways through knockout or classical transgenic approaches in Drosophila melanogaster, Caenorhabditis elegans and mammalian cells (Folger et al., 1982), and by the establishment of embryonic stem (ES) cell technology including their subsequent manipulation to generate transgenic mice lacking, overexpressing or replacing specific genes and other more complex organisms (Nagy et al., 1993; Thomas and Capecchi, 1987; Zan et al., 2003). Here, site-directed mutagenesis was often used to investigate the function of a gene product by causing a complete ablation of the protein of interest (knock-out) rather than altering specific binding sites. The impact of these mutagenesis studies was rather observed on a phenotypic level than allocating the modified protein to existing complexes or pathways. By analyses of presumably affected proteins on a molecular level (e.g. Western immunoblotting, protein localization), metabolic and signaling pathways could be deduced by demonstration of variably regulated target proteins (Wang et al., 2009b). Many, now well established, pathways and regulatory mechanisms were initially described using knockout strategies. The circumstance that the depletion of a protein in a highly complex system like a cell or even a complete organism may elicit compensatory or secondary mechanisms complicates succeeding analyses: The observed effects do not necessarily reflect the physiological influence of the analyzed protein in a highly dynamic and fluctuating PPI network.

For that reason, alternative methods for the analysis of protein functions have been developed over the years, directly or indirectly detecting the physical interaction of proteins *in vitro* or *in vivo*. Hereby, differential focus has been laid on the aspects of sensitivity and specificity. While the focus of studies investigating whether protein **a**

interacts with protein **b** lies on high specificity, approaches for high sensitivity try to detect as many interaction partners of protein **a** as possible. Methodically, two strategies can be distinguished: **(A)** One strategy intends to provide evidence that two given proteins interact, while **(B)** the other strategy aims at the identification of the identity of all interactors of one specific protein.

In typical approaches to provide answers to **(A)**, both proteins to be investigated have to be labeled. Then, their spatial interaction can be analyzed by biological screening methods like the Yeast-2-hybrid (Y2H) technique or fluorescence resonance energy transfer (FRET). Y2H relies on the fusion of a bait protein – the protein that is being used to screen for potential interactors – to a GAL4-BD binding site whereas prey proteins – the proteins binding to the bait - possess the GAL4 transcription factor-activating domain (GAL4-AD), together reconstituting the expression of a reporter gene (Fig. 1.1 a). While proteins investigated by Y2H can originate from any species, this system is restricted to the detection of direct PPI of small proteins that are able to pass the nuclear pore (Suter et al., 2008).

However, proteins isolated from their physiological compartment may have altered affinities for interactors and could give rise to false-positive results. Therefore, bait and prey interaction also can be analyzed in any tissue that can be modified to express tagged proteins in order to provide a suitable proteomic environment. A common technique applied here is FRET, which measures the energy transfer from one fluorophore to another fluorophore. For PPI analysis, the protein pair of cyano-fluorescent protein (CFP) and yellow-fluorescent protein (YFP) is typically used due to their overlapping emission/excitation spectra (Fig. 1.1 b). Because of the mechanism of dipole-dipole coupling, the efficiency of transfer is highly sensitive to small distances and represents a reliable methodology to assess distances of proteins exclusively present during direct interactions (C. Harris, 2010; Ciruela et al., 2010; Helms, 2008).

A commonly used assay for the analysis of PPI is co-immunoprecipitation. Here, the bait is isolated from whole cell lysate using a specific antibody bound to the bait protein (Dwane and Kiely, 2011; Elion, 2006). Physically interacting proteins then co-purify with this protein and can subsequently be identified by immunoblotting techniques. A prerequisite for the applicability of this technique is the correct anticipation of a PPI as well as the availability of highly specific antibodies for both

the protein of interest as well as the anticipated interactor, rendering this technique unsuitable for screening approaches.

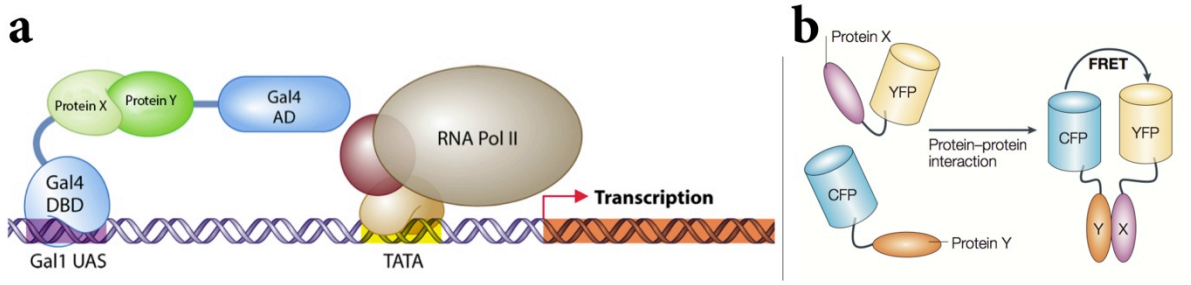


Figure 1.1: Yeast two-hybrid and Fluorescent resonance energy transfer (FRET)
a: *In a yeast two-hybrid (Y2H) experiment, the interaction of two proteins can be studied by fusion of one protein to the DNA-binding domain (DBD) of Gal4 (the bait) and fusion of the other protein to the activation domain (AD) of Gal4 (the prey). Once bait and prey interact, the binding of Gal4 DBD to the upstream activating sequences (UAS) of the reporter gene mediates the recruitment of the basal transcriptional machinery and expression of the reporter. Adapted and modified from Stynen et al 2012.* ***b:*** *The interaction of two proteins can be visualized using the technique of fluorescent resonance energy transfer (FRET). Here, the FRET-based probe consists of two different proteins that are labeled with cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP), respectively. Upon interaction of both proteins, the fluorophores are brought into close proximity and YFP is excited by FRET. Adapted and modified from Zhang et al 2002.*

(B) If all prey-bound proteins are to be identified, other methods of detection are required. Mass spectrometry represents an analytical method which is capable to determine the identity and quantity of individual proteins within thousands by detection of the mass-to-charge ratio of charged particles and can be performed after co-immunoprecipitation followed by high resolution liquid chromatography (David Sparkman, 2000; Hubner et al., 2010a). Furthermore, quantitative or tandem affinity purification can be used in conjunction with co-immunoprecipitation to further enhance specificity and sensitivity (Puig et al., 2001; Rigaut et al., 1999). By expression of tagged proteins and a subsequent generic affinity chromatography, core proteomes of the yeast have been mapped on a genome-scale approach and were found to show strong concordance with genetic interactions (Collins et al., 2007a; Collins et al., 2007b; Gavin et al., 2006; Krogan et al., 2006). Furthermore, interactomes have been established for higher organisms including mammals, at least for designated cell lines. However, the need to identify PPIs in a more physiological setting is still insufficiently met and demands the development of techniques suitable to detect physiological PPIs.

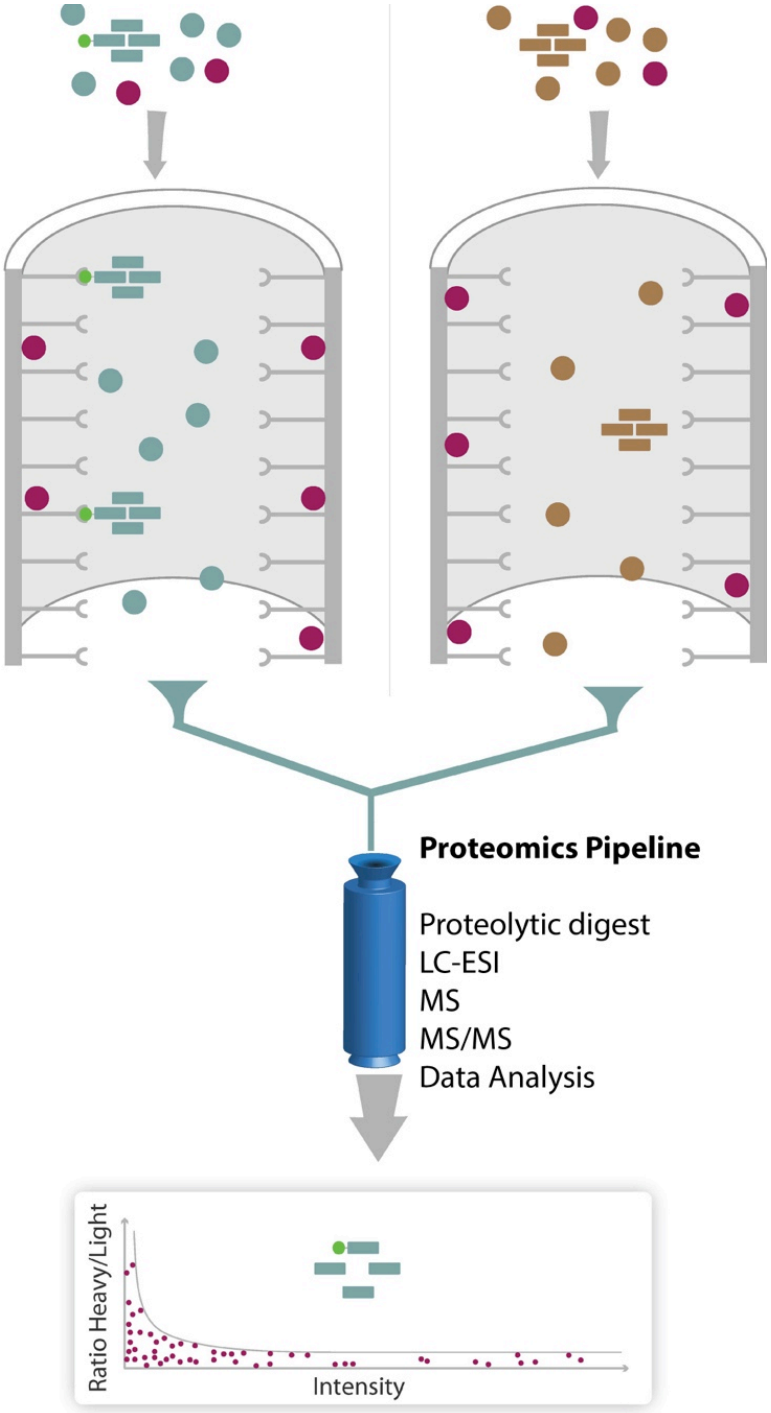


Figure 1.2 LC-MS/MS-based interaction proteomics
For detection of specific binding partners of a tagged (e.g. by GFP, green circles) protein of interest, the abundance of affinity-purified potential interaction partners is compared with the abundance of the same protein from a mock purification under expression of an untagged version of the protein. Protein extracts of SILAC-labeled cells expressing the tagged protein (left, blue) or control populations (right, ochre) are subjected to affinity purification and eluates mixed before analysis by the proteomics pipeline. Following a proteolytic digest, peptides are separated by liquid chromatography (LC) coupled to electro-spray ionization, analyzed by mass spectrometry (MS), fragmentation of selected peptides and analysis of the resulting MS/MS spectra. Subsequent computational analysis is then applied for the identification and quantification of proteins based on the presence of detected peptides. By evaluation of the ratio of proteins labeled with heavy amino acids versus unlabeled proteins, a prediction of interactors (rectangular) and background binders (red circles) can be made. Adapted and modified from Walther et al 2010.

1.1.2. Expression systems for tagged proteins

In order to study protein-protein interactions, the technique of quantitative affinity purification followed by mass spectrometry (AP-MS) was developed over the last years and became a powerful approach for the analysis of PPIs in an unbiased manner (Paul et al., 2011). This technique relies on the biochemical enrichment of the protein of interest together with its interaction partners. Subsequently quantitative mass spectrometry analysis is used to detect proteins that are enriched when compared to controls. While this technique generally yields faithful results and is particularly suitable for singular experiments, such an approach is impractical when aiming for the precise acquisition of large interaction networks encompassing a multitude of different proteins that have to be used as bait. The application of various antibodies requires an individual determination of their epitope detection specificity and non-specific binders or the conduction of laborious validation experiments, rendering this approach unsuitable for high-throughput screening approaches (Schirle et al., 2012). Therefore, streamlined methods for the investigation of PPIs rely on the utilization of a protein tag fused to the protein of interest. Hereby, a small set of specific antibodies with well-known recurring unspecific interactors can be used that proved to be suitable in AP-MS studies conducted in arthropodan or mammalian systems.

In order to introduce an adequately tagged protein into a cell, techniques most commonly used in the field of PPI studies rely on the application of overexpression systems since broadly used strong small promoters provide high amounts of proteins for subsequent AP-MS, thereby increasing the sensitivity (Table 1.1). Classically, expression plasmids containing a cDNA of the respective protein are used for transient overexpression and represent a convenient and straightforward approach to yield large data sets. However, as cells acquire different levels of plasmid or lose it after a few days, results often cannot be reproduced reliably. To overcome this issue, stable cell lines using retro- or lentiviral vectors can be generated in a similarly convenient and efficient manner (Ni et al., 2011). The highly efficient viral transduction of a multitude of recipient cell lines hereby represents a valuable tool for the analysis of cells resistant to efficient transfection by plasmid-based approaches. Whereas such systems unequivocally have significant advantages such as straightforward construction of expression constructs commonly in connection with a high protein expression, the specificity of derived interaction data is inferior to

classical AP-MS. While these approaches have yielded fundamental insights into protein networks and became more relevant as techniques of quantitative proteomics evolved over the last years, they generally suffer from unphysiological high protein levels often leading to false-positive interactome data, for example via misguided localization caused by the abundance of the overexpressed proteins, incorrect protein folding, protein complex assembly and altered downstream regulation (Vermeulen et al., 2008).

The introduction of a tagged gene via homologous recombination offers a perspective to study PPIs at physiological expression levels (Haber et al., 2004; Marcon and Moens, 2005). Here, highly homologous linear DNA constructs are incorporated into chromosomes by homologous recombination, retaining their endogenous regulatory sites. In theory, this process would be perfectly suitable to examine proteins expressed at endogenous levels, as these modified variants possess the same regulatory sites and intron/exon composition than their non-modified analog. However, the efficiency of integration is very low and therefore unsuitable for the establishment of larger numbers of cell lines (Porteus, 2011; Smithies et al., 1984).

For the insertion of a DNA construct into a specific site of the chromosome, a multitude of suitable techniques, termed genome editing, were developed over the last years. The application of zinc finger nucleases is a highly efficient process, capable to precisely introduce double stand breaks at predetermined loci mediated by a fusion protein of the type II restriction endonuclease *FokI* and engineered DNA sequence-specific zinc finger domains. These double strand breaks increase the efficiency of the integration of introduced linear DNA by homologous recombination (Ramalingam et al., 2012). Similarly, the transcription activator-like effector nucleases (TALEN), are composed of a *FokI* endonuclease cleavage domain fused to a TALE DNA binding domain that can be readily manipulated to target any genomic site of interest (Miller et al., 2011). The most recent technique for genome editing, the clustered, regularly interspaced, short palindromic repeats (CRISPR)–CRISPR-associated protein (Cas) system utilizes the CAS endonuclease that is guided to genomic sequences by artificial RNA sequences and thus can be easily customized without changing the protein component (Chang et al., 2013; Cho et al., 2013). However, all methods bear the risk for off-target effects (e.g. non-specific DNA

cleavage) and therefore have to be tightly monitored for inadvertently introduced aberrations (Cheng et al., 2012; Urnov et al., 2010).

An alternative technique for genome editing exploits the guided genomic integration properties at low frequencies of the adeno-associated virus (AAV), which typically integrates into the AAVS1 site on the human chromosome 19 (Kotin et al., 1990; Surosky et al., 1997). The modification of AAV recognition sites enabled the precise targeting of a specific genomic locus and the introduction or manipulation of small sequences (Vasileva and Jessberger, 2005). By exploiting the virus-mediated homologous recombination system, efficiencies for targeting are much higher compared to cell-intrinsic homologous recombination but struggle with size limitations of sequences to be introduced. While homologous recombination typically is conducted in pluripotent cell populations, the AAV system can also be employed to target somatic cells (Bronson et al., 1996; Kohli et al., 2004). Thus, the application of homologous recombination is not easily accessible for streamlining and high-throughput screens – even with modern tools like Zinc finger nucleases, TALENs, CRISPR or AAV vectors -, and require the application of alternative techniques for the expression of engineered proteins at endogenous levels.

When aiming for endogenous expression levels of tagged proteins that are regulated comparably to their unaltered counterparts, the majority of the genomic locus of a gene should be included in constructs encompassing its endogenous promoter as well as naturally occurring regulatory elements. Frequently, such DNA constructs are available as bacterial artificial chromosomes (BAC) and can have a size of more than 300 kb (O'Connor et al., 1989). Traditionally, linearized BAC constructs are employed in studies conducting homologous recombination, providing large homologous regions that facilitate a site-directed and relatively efficient integration to predefined sites in pluripotent cell populations (Thomas and Capecchi, 1987). However, standardized approaches for the generation of transgenic mice employ techniques like pronuclear microinjection or electroporation into embryonic stem cells and result in random rather than site-directed integration of the DNA construct into the genome (Tunster et al., 2011). The alternative integration of DNA constructs typically is mediated by non-homologous end joining (NHEJ), resulting in an integration of the DNA at random loci (Guirouilh-Barbat et al., 2004). However, integration efficiencies are much higher and therefore are applicable for streamlined applications. Hereby, the large size of the BAC transgenes ensures the presence of most, if not all,

regulatory elements and results in an expression that closely resembles that of the endogenous genes (Poser et al., 2008). Unlike small expression constructs stably integrating into a cell's genome, BACs are less likely to be influenced by cis-acting regulatory elements or silencing events caused by chromatin positional effects as they contain their own natural chromatin domains (Blaas et al., 2012). Stable BAC-transgenic cell lines therefore represent an attractive system expressing tagged proteins including different splice variants with expression levels similar to endogenous variants.

Table 1.1: Characteristics of different expression systems

	Plasmid [1,2]	Lentivirus [3,4]	Retrovirus [5,6]	homologous recombination [7,8,9]	genome editing [10,11,12,13,14]	AAV-mediated integration [15,16]	BAC [17,18]
Stable integration	no	yes	yes	yes	yes	yes	yes
Specific integration site	no	no	no	yes	yes	yes	no
Introduction into non-dividing cells	+	+	-	-	+/-	+/-	no
Expression level	++	++	++	endogenous	endogenous	endogenous	endogenous-like
Efficiency	++	++	++	--	+/-	+/-	+
Protein-protein interaction specificity	?	?	?	reference	to be evaluated	to be evaluated	to be evaluated
Typical size of introducable DNA	~ 20 kb	~ 10 kb	~ 10 kb	≥ 300 kb	≥ 300 kb	~ 4 kb	≥ 300 kb

[1]Campeau et al., 2001; [2]Stadler et al., 2004; [3]Kumar et al., 2001; [4]Naldini et al., 1996; [5]Grossmann et al., 1993; [6]Cepko et al., 1984; [7]Folger et al., 1982; [8]Bronson et al., 1996; [9]Sung et al., 2006; [10]Cho et al., 2012; [11]Mali et al., 2013; [12]Miller et al., 2011; [13]Hockemeyer et al., 2011; [14]Urnov et al., 2010; [15]Khan et al., 2010; [16]Kim et al., 2008; [17]Poser et al., 2008; [18]Rostovskaya et al., 2012

1.1.3. Model systems for protein-protein interactions

Protein-protein interaction (PPI) studies are conducted in a large variety of systems and provide results for different scientific questions. Hereby the quality of generated data is dependent on the system itself, the method of measurement and the class of protein investigated. In principle, two modes of protein-protein interaction systems can be distinguished: Systems where prey and bait proteins are known and brought in direct vicinity in order to analyze their interaction, termed *binary PPI* as results can only be "true" or "false" and *quantitative PPI*, as preys are more or less enriched, not yielding a binary result. For the analysis of protein-complex formation or for assessing subtle changes in complex compositions, only quantitative PPI can provide suitable results as in some conditions only fractions of complexes in a subgroup of cells may differ. When screening for yet unknown interactors of a tagged protein, the proteomic context in which PPI measurements are conducted is of significant importance as any interaction candidates originate from this environment. For that reason, human cell lines like HeLa cells were used for PPI studies of tagged human protein variants in multiple studies rather than employing cells derived from different

species (Hubner et al., 2010a; Poser et al., 2008). In these and similar studies the cell lines used were primarily derived either from tumor-tissue or immortalized primary cells to exploit the potential of unlimited self-renewal and lack of senescence (Graham et al., 1977; Ivanković et al., 2007; Tsuchiya et al., 1980). Although the dynamic equilibrium of PPI networks substantially varies between different cell types, tissues and differentiation stages, our current understanding of protein interaction networks is mainly based on overexpression paradigms in non-human or transformed cell lines (Schirle et al., 2012). Therefore, physiological regulation of cell cycle or apoptotic control-associated proteins - amongst others - is not guaranteed and could lead to an erroneous model of cellular processes and conditions in somatic tissue. Hence, it would be desirable to generate protein-specific PPI networks in a proteomic environment similar to physiologically relevant cell types. Although the majority of basic cellular processes are comparable in different cell lines, cell type specificity is of utmost importance for the validation and discovery of disease processes that are directly linked to specific tissues with unique proteomes. Therefore, human tissue samples should ideally be equipped with an endogenously expressed, regulated and spliced tagged protein for AP-MS analysis. Unfortunately, our current techniques allow for the introduction of complex expression cassettes with insufficient efficiencies, demanding a proliferating cell population amenable to genetic manipulation and subsequent enrichment of transgenic cells by chemoselection or other means of purification. Somatic tissue samples largely consist of post-mitotic cells, which have long lost their ability for self-renewal and therefore render themselves unsuitable for such approaches. In particular, this concern appears to be highly relevant for the central nervous system, as postmitotic neurons are highly specialized cells with unique protein compositions, impossible to cultivate as required for the generation of a homogenous cell population facilitating the expression of tagged proteins at physiological levels (Uhlen et al., 2010a).

1.2. Stem cells

Stem cells represent the most competent cell population in the development of an organism. In mammals, two main stem cell types can be distinguished, the cell population able to develop all three germ layers of an organism - embryonic stem (ES) cells - and adult- or somatic stem cells, whose purpose is the homeostasis and replenishment of organs like the liver, hair follicles or the intestinal system (Baddour et al., 2012; Boehnke et al., 2012; Radtke and Clevers, 2005). All stem cell populations share the ability for continuous self-renewal and differentiation into more restricted specialized somatic cells. Thereby, the range of distinct cell types a stem cell is able to generate defines its potency. Adult stem cells are multipotent and therefore have the ability to generate progeny of several distinct cell types like neurons and glial cells derived from neural stem cells or myeloid and lymphoid lineages that are generated from hematopoietic stem cells (Clements and Traver, 2013). Embryonic stem (ES) cell cultures are established by extraction of the inner cell mass of blastocysts during the early phases of embryonic development and can give rise to all three germ layers - ectoderm, mesoderm and endoderm.

The areas of application in research projects using stem cells are widespread. Classical studies examining hypotheses associated with processes during the early development of an organism often use mouse ES cells as a tool to create transgenic mice as model systems. Another key field is the generation of specific tissues or cellular subtypes by establishment of suitable differentiation protocols. Here, adult stem cells serve as a naturally occurring somatic stem cell population and usually can be readily differentiated into mature postmitotic cells of a certain fate. These specialized subtypes then are either used for an *in vitro* modeling of disease-related processes, screening of pharmacological compounds influencing these mechanisms or evaluated for their ability to serve as a cellular replacement in diseases associated with cellular loss. Furthermore, basic cellular mechanisms like the core transcriptional regulatory circuitry or the comparability of different PS cell lines themselves are subject of intense research (Boyer et al., 2005; Initiative et al., 2007). While adult stem cells are residing in the majority of the somatic tissues, only a fraction of these populations is readily accessible for purposes of fundamental research. This circumstance is negligible when investigating such populations in animal models but represents a great challenge in human systems, as here only a limited number of populations are accessible by reasonable clinical intervention. These populations

encompass stem cells derived from bone marrow, adipose tissue and blood yet especially exclude the central nervous system.

1.2.1. The scientific history of human pluripotent stem cells

While in 1981 Evans and Kaufman reported the establishment of mouse pluripotent cell lines directly from in vitro cultures of mouse blastocysts, the inner cell mass from human blastocysts was extracted and cultivated only in 1994 by the group around Bongso (Bongso et al., 1994; Evans and Kaufman, 1981). Four years later, Thomson and colleagues established the first human embryonic stem (hES) cell lines (Thomson et al., 1998). Subsequently, techniques involving the derivation and maintenance of hES cells rapidly evolved, and the field of stem cell biology significantly improved. Highly sophisticated methods, including immunosurgery or mechanical and laser-assisted isolation of the inner cell mass from the morula were developed (Kim et al., 2006; Peura et al., 2008; Strelchenko et al., 2004; Turetsky et al., 2008; Unger et al., 2008). With the perspective of a future application in regenerative medicine, culture conditions were improved to meet the requirements for the regulations for medical products and good manufacturing practice (GMP) guidelines by replacing compounds of animal origin with chemical compounds (Unger et al., 2008). Quickly, protocols for the generation of specific human somatic cells of high purity were developed. These encompass cells of the central nervous system (Perrier et al., 2004; Zhang et al., 2001), cardiomyocytes (Kehat and Gepstein, 2003), hepatocytes (Hay et al., 2008) and pancreatic cells (Wang and Sander, 2012). The analysis of fundamental processes defining pluripotency employing murine and human ESCs led to a deeper understanding of the molecular processes of pluripotency (Cartwright et al., 2005; Chambers et al., 2003) and culminated in the discovery of its basic regulatory and determining machinery when Takahashi and Yamanaka demonstrated that the four transcription factors (TF), Oct3/4, Klf4, Sox2 and c-Myc were sufficient to reprogram somatic cells to the state of induced pluripotent stem cells (iPSC) (Nakagawa et al., 2008; Takahashi et al., 2007). The advent of iPSCs provided a tool for the scientific community enabling the examination of all cell types that can be derived from PS cells in any genetic background. Although some studies reported occasional cases of increased tumorigenesis of iPSC chimeric mice (Okita et al., 2007; Okita et al., 2011) or cultivation-induced chromosomal aberrations (Ronen and Benvenisty, 2012), iPSCs are considered

largely identical to ES cells in terms of their capacity for self-renewal and differentiation potential (Gore et al., 2011; Hussein et al., 2011; Lister et al., 2011). With the perspective of using iPSC-derived cell populations one day as material for cell replacement therapy in regenerative medicine, safer and more sophisticated methods for the generation of iPSCs were demanded (Okita et al., 2011). Emphasis was laid on avoiding integration-prone or viral methods of reprogramming during iPSC generation to avoid genetic alteration of the cells, which was addressed by multiple measures: (I) A reduction of transcription factors (TF) necessary for reprogramming (Nakagawa et al., 2008), (II) their replacement by chemical compounds (Zhu et al., 2010) and (III) development of integration-free methods. These include excision of expression vectors after reprogramming (Sommer et al., 2010), protein transduction of the TF (Grant et al., 2009) transfection of modified mRNA (Warren et al., 2010) and the transfection of mature microRNAs (Anokye-Danso et al., 2011; Miyoshi et al., 2011; Sommer et al., 2010). One of the most efficient and broadest applicable method to date is transient expression of TFs by using non-integrating Sendai virus (Fusaki et al., 2009). However, mastering the complex differentiation of authentic somatic cell types in vitro remains a key challenge. The availability of pluripotent stem cells of non-embryonic origin ameliorates the ethical situation and most legal issues in many countries and facilitates the access of a broader community to this valuable scientific resource.

1.2.2. Differentiation potential of human pluripotent and neural stem cells

Human pluripotent stem cells (hPSCs) as the most immature cell population have the potential to differentiate into all three embryonic germ layers (Thomson et al., 1998). Hereby, the maintenance of a finely balanced signaling network that enables the preservation of a complex functional network of protein-protein interactions together with miRNA-regulation is critical to maintain hPSCs (Dalton, 2012; Lichner et al., 2011; Smith et al., 2011). While hPSC generation methods were improved successfully by employing integration-free techniques that enable comparability and largely avoid the formation of degenerated cell populations, defined differentiation into specific somatic cell types still poses major obstacles (Initiative et al., 2007; Lee et al., 2012; Ohmine et al., 2011; Okita et al., 2011; Park et al., 2012).

While many tissues show a pronounced regeneration potential with the ability to replace lost cells, the human central nervous system has a limited capacity for regeneration. For that reason, the development of biological models aiming to

describe molecular processes during traumata and the generation of cell populations fit to serve as cell replacement or model system for neurodegenerative diseases are focus of many research projects. Although a number of studies showed that upon withdrawal of factors used in standard proliferation conditions murine and human ESCs tend to show a neuroectodermal “default” mechanism for differentiation, the resulting neural populations show distinct variations in their cellular composition along with unsatisfactory neural purities (Hsu et al., 2007; Muotri et al., 2005; Reubinoff et al., 2001; Vallier et al., 2004). In order to provide suitable tools for the generation of mature human cell types, individual differentiation protocols were developed – either relying on directed differentiation or on lineage selection approaches (Barberi et al., 2005; Ladewig et al., 2008; Laflamme et al., 2007; Nistor et al., 2005; Wernig et al., 2002; Yan et al., 2005; Zhang et al., 2001). One major prerequisite for the successful utilization of hPSCs for the study of neurodegenerative diseases or other applications such as systems biological approaches is the efficient and robust differentiation of the required neuronal subtypes as defined cultures with minimized batch-to-batch differences. Initial methods for the derivation of neural cell types were based on ‘run-through’ protocols, where hPSC were directly differentiated into mature neuronal cultures, passing the state of multipotent progenitors and often resulting in decreased homogeneity due to incompletely differentiated cultures with individual features inherent to lengthy differentiation protocols (Kim et al., 2011). The isolation of neural tube-like structures, an intermediate precursor cell population emerging in cultures of plated embryoid bodies (EBs), resulted in the establishment of a highly homogenous long-term neural stem cell population (lt-NES). This cell population can be differentiated into glial or neuronal cells with highly reproducible neuronal differentiation rates, independent from variances in the neural differentiation propensity of the parental hPSC line. Furthermore, lt-NES cells remain susceptible to morphogens such as Sonic hedgehog (SHH), FGF8 and retinoic acid, facilitating the targeted generation of different subtypes such as midbrain dopamine neurons or spinal motor neurons even after long-term cultivation (Falk et al., 2012; Koch et al., 2009). Notably, these neurons express functional K^+ and Na^+ channels, are able to fire action potentials, spontaneously form synaptic networks and receive synaptic input. These characteristics are generally considered as important hallmarks of neuronal maturity and enable these cells to form a functional synaptic network *in vitro* (Koch et al., 2009). Upon cerebral transplantation in laboratory animals, lt-NES cells

give rise to long-living mature neurons with a remarkable potential to generate functional long-range axonal projections to distant areas of the adult mammalian brain (Steinbeck et al., 2011).

1.2.3. iPSC-based disease modeling

While extensive efforts to establish animal models for various diseases in order to address their underlying molecular processes have been made, they do not necessarily provide results applicable to diseases in humans (Elder et al., 2010; Sager et al., 2010). Despite rich contributions to the field that substantially formed our current understanding of key processes in disease initiation and progression, these models ultimately fail to provide non-biased outcomes, which might be attributable to the xenobiotic setting or the late-onset of many diseases studied (Kunkanjanawan et al., 2011; Peters et al., 2010; Pournasr et al., 2011). With the advent of iPSC technology, new models for the analysis of molecular processes underlying disease onset and progression were established and provided a cellular system closely related to affected cell types and genotypic setting. In this context, an increasing number of human iPSC-based disease models was established, exploiting the capacity to generate and investigate the affected somatic cell population, providing valuable information regarding disease-development (Colman and Dreesen, 2009; Han et al., 2009; Koch et al., 2011). For instance, first insights into the pathological molecular mechanisms of the neurodegenerative late-onset Machado-Joseph disease (MJD; syn. spinocerebellar ataxia type 3; SCA3) were gained using afflicted lt-NES cells derived from a patient for the generation of homogenous, mature neuronal cultures. In this disease, an abnormal expansion of trinucleotide CAG repeats within the Ataxin-3 gene (Atxn3) is assumed to lead to a loss of cells. The elongated isoform of ATXN3 leads to an aberrant protein conformation that initiates a pathogenic cascade leading to neuronal dysfunction and degeneration of neurons in the brainstem and other regions (Costa and Paulson, 2012; La Spada and Taylor, 2010). By taking advantage of patient-specific lt-NES cells, a new mechanism for the formation of aggregates suspected to be involved in MJD-development was proposed: Upon activation-dependent Ca^{++} influx via voltage-gated Ca^{++} channels, extended ATXN3 is cleaved in a calpain-dependent process (Koch et al., 2011). These observations indicate that neurons derived from lt-NES cells closely resemble human neurons *in vivo* and are sufficiently mature to show features of late-onset

diseases. While *in vitro* generated neural derivatives certainly represent rather immature counterparts of neurons or glial cells found in the adult human brain, they nevertheless provide a suitable human cell source with respect to availability, homogeneity and genetic accessibility.

1.3. Aim of the study

While biomedical research made substantial progress employing cell-based systems for the investigation of disease processes, the majority of recent discoveries concern general metabolic, regulatory or developmental processes. The analysis of protein complex composition is an important part of studies exploring cell signaling cascades, cell-cell interactions as well as principles of signal transduction mechanisms. However, the majority of human interactome studies conducted was based on overexpression paradigms in tumor cell lines that often lead to common problems like unspecific interactions caused by supra-physiological protein expression levels, the use of transformed cells and non-tissue specific proteomic environments.

The ability to generate human neurons from embryonic stem (ES) cells and induced pluripotent stem cells (iPSC) provides the perspective to conduct studies aimed at discovering the underlying molecular principles of disease processes in previously inaccessible cell populations. The objective of this study is to establish and evaluate techniques capable to perform protein-protein interaction analyses in vital human neurons generated from human PSC.

In the first part, bacterial artificial chromosomes (BACs) will be used for the introduction of tagged proteins into PSC-derived long-term neuroepithelial-like stem (lt-NES) cells to facilitate their expression at endogenous levels. Following validation procedures, interactors of the introduced proteins can then be determined by affinity purification of proteins in combination with quantitative mass spectrometry. Data generated in this context will be used to comparatively characterize the composition of protein complexes in lt-NES cells and their neuronal progeny. Furthermore, the applicability of BAC transgenesis for iPSC-derived lt-NES cells will be assessed in order to evaluate the feasibility of quantitative BAC–green fluorescent protein interactomics (QUBIC) analysis in a diseased background.

Although BAC-mediated protein expression is well suitable for most - if not all - proteins, the size of a GFP protein tag together with the probability of a mild overexpression due to the additional genomic locus caused by the BAC integration might be insufficient for some studies. Therefore, a further objective is to employ an AAV-based technique for epitope tagging of endogenous genes in order to generate lt-NES cell populations without modification of promoter-associated regulatory regions or exonic/intronic sequences.

2. Materials

2.1. Technical equipment

Device	Name	Manufacturer	Registered office
Autoclave	D-150	Systec	Wettenberg, Germany
Balance	BL610	Sartorius	Göttingen, Germany
Balance	LA310S	Sartorius	Göttingen, Germany
Block heater	Thermomixer compact	Eppendorf	Hamburg, Germany
Bone drill (0.7-mm burrs)	Micro Drill	Fine Science Tools	Hamburg, Germany
Centrifuge (cell culture)	Megafuge 1.0R	Sorvall	Hanau, Germany
Centrifuge (table top)	5415D	Eppendorf	Hamburg, Germany
Chemidoc	Chemidoc 2000	BioRad	Hercules, USA
Concentrator	Concentrator 5301	Eppendorf	Hamburg, Germany
Counting chamber	Fuchs-Rosenthal	Faust	Halle, Germany
Cryostat	Cryostat HM 560	Microm Laborgeräte	Walldorf, Germany
Digital camera	C 5050 Zoom	Olympus Optical	Hamburg, Germany
Digital camera	Canon Power Shot G5	Canon	Krefeld, Germany
Electroporator	Gene Pulser Xcell	BioRad	Hercules, USA
FACS® analyzer	FACS® Calibur	BD Biosciences	San Jose, USA
FACS® sorter	FACS® DiVa	BD Biosciences	San Jose, USA
Freezer -80°C	HERAfreeze	Kendro	Hanau, Germany
Glass-Microelectrode Puller	PE-21	Tritech Research	Los Angeles, USA
Gel chamber	Agagel	Biometra	Göttingen, Germany
Geldoc	GelDoc EZ	BioRad	Hercules, USA
Incubator	HERAcell	Kendro	Hanau, Germany
Inverse light microscope	Axiovert 25	Carl Zeiss	Jena, Germany
Liquid nitrogen store	MVE 611	Chart Industries	Burnsville, USA
LED light source	Colibri 2	Carls Zeiss	Jena, Germany
Microliterpipet	Microliterpipet 1710N	Hamilton	Bonaduz, Switzerland
Microscope	Axiovert 40 CFL	Carl Zeiss	Jena, Germany
Microscope	Axiovert 200M	Carl Zeiss	Jena, Germany
Microscope	Axioskop 2	Carl Zeiss	Jena, Germany
Microscope	IX 81 Confocal	Olympus	Hamburg, Germany
Microscope camera	Axiocam MRM	Carl Zeiss	Jena, Germany
Microscope camera	ProgRes C14	Jena Optic	Jena, Germany
Microscope laser	Laser Mells Griot	Griot Lasergroup	Carlsbad, Germany
Microscope slides	Superfrost plus	Menzel-Gläser	Braunschweig, Germany
Nucleofector	Nucleofector 2b	Lonza	Basel, Switzerland
Operating microscope	OpMi	Carl Zeiss	Jena, Germany
PAGE/Blot equipment	Mini-PROTEAN 3	BioRad	Hercules, USA

PAGE gel power supply	PowerPac 200	Biorad	Hercules, USA
Plastic coverslips	Thermanox®plastic coverslips	Nunc	Wiesbaden, Germany
PCR cycler	T3000 Termocycler	Biometra	Göttingen, Germany
pH-meter	CG840	Schott	Mainz, Germany
Pipetteboy	Accu-Jet	Brand	Wertheim, Germany
Platereader	Infinite® 200 PRO	Tecan	Männedorf, Switzerland
Polyester membrane	Transwell-Clear	Corning	Bodenheim, Germany
Real-time qPCR mashine	Mastercycler realplex	Eppendorf	Hamburg, Germany
Refrigerator	G 2013 Comfort	Liebherr	Lindau, Germany
Shaker	Bühler Schüttler WS 10	Johanna Otto	Hechingen, Germany
Spectrophotometer	Nanodrop ND-1000	Peqlab	Erlangen, Germany
Sterile laminar flow hood	HERAsafe	Kendro	Hanau, Germany
Stereo microscope	STEMI 2000-C	Carl Zeiss	Göttingen, Germany
Stereotactic Frame	Stereotactic Frame	Stoelting	Illinois, USA
Sterilizer	Hot Bead Sterilizer	Fine Science Tools	Heidelberg, Germany
Thermocycler	T3 Thermocycler	Biometra	Göttingen, Germany
Table centrifuge	Centrifuge 5415R	Eppendorf	Hamburg, Germany
Transplantation tools	Transplantation Tool Set	Fine Science Tools	Heidelberg, Germany
Ultracentrifuge	Sorvall	Discovery 90SE	Hanau, Germany
UV-Vis Spectrophotometer	BioPhotometer	Eppendorf	Hamburg, Germany
Vacuum pump	Vacuubrand	Brand	Wertheim, Germany
Vortex	Vortex Genie 2	Scientific Industries	New York, USA
Water bath	1008	GFL	Burgwedel, Germany
Water bath	WB7	Memmert	Schwabach, Germany
Water filter	Millipak® 40	Millipore	Eschborn, Germany

2.2. Cell culture consumables

Consumables	**Manufacturer**	**Registered Office**
4 well culture dishes	Nunc	Wiesbaden, Germany
6 well culture dishes	PAA	Pasching, Austria
12 well culture dishes	Nunc	Wiesbaden, Germany
96 well culture dishes	PAA	Pasching, Austria
Cellstar ^{TM} 175 cm^2 flasks	Greiner Bio-One	Solingen, Germany
Cellsieve Nylon 40 µm	BD Falcon	Bedford, USA
Petri dishes ∅ 3.5 cm	BD Biosciences	Heidelberg, Germany
Petri dishes ∅ 6 cm	BD Biosciences	Heidelberg, Germany
Petri dishes ∅ 10 cm	BD Biosciences	Heidelberg, Germany
Round bottom tubes - 12x75 mm	BD Biosciences	Heidelberg, Germany

Serological pipettes	BD Biosciences	Heidelberg, Germany
Syringes 20 ml	BD Biosciences	Heidelberg, Germany
Syringe filter 0.2 μm	PALL	Dreieich, Germany
Syringe filter 0.45 μm	Sarstedt	Nuembrecht, Germany
TC dishes ∅ 3.5 cm	PAA	Pasching, Austria
TC dishes ∅ 6 cm	PAA	Pasching, Austria
TC dishes ∅ 10 cm	PAA	Pasching, Austria
TC dishes ∅ 15 cm	SPL	Eumhyeon-ri, Korea
Tubes 0.5 ml	Greiner Bio-One	Solingen, Germany
Tubes 1.5 ml	Greiner Bio-One	Solingen, Germany
Tubes 15 ml	Greiner Bio-One	Solingen, Germany
Tubes 50 ml	Greiner Bio-One	Solingen, Germany
Cryovials 1 ml	Nunc	Wiesbaden, Germany
Cryovials 1.8 ml	Nunc	Wiesbaden, Germany

2.3. Chemicals and reagents

Substance	**Manufacturer**	**Registered office**
Acetic Acid	Sigma Aldrich	Deisenhofen, Germany
30% Bis/Acrylamide	Carl Roth	Karsruhe, Germany
Agar	Sigma Aldrich	Deisenhofen, Germany
Agarose	PeqLab	Erlangen, Germany
Ampiciline	Sigma Aldrich	Deisenhofen, Germany
Ascorbic acid	Sigma Aldrich	Deisenhofen, Germany
B27 supplement	Invitrogen	Karlsruhe, Germany
Benzoase	Sigma Aldrich	Deisenhofen, Germany
Bone wax	Fine Scientific Tools	Heidelberg, Germany
Bromphenol blue	Sigma Aldrich	Deisenhofen, Germany
BSA solution (7.5%)	Invitrogen	Karlsruhe, Germany
$CaCl_2$	Sigma Aldrich	Deisenhofen, Germany
cAMP	Sigma Aldrich	Deisenhofen, Germany
Chemiluminescent Substrates	Thermo Scientific	Frankfurt, Germany
Chloroquin	Sigma Aldrich	Deisenhofen, Germany
Collagenase IV	Sigma Aldrich	Deisenhofen, Germany
cOmplete ULTRA™ tablets	Roche Diagnostics	Basel, Swizerland
CytocoonTM Buffer II	Evotec Technologies	Hamburg, Germany
DABCO	Sigma Aldrich	Deisenhofen, Germany
DAPI	Sigma Aldrich	Deisenhofen, Germany
DAPT	Sigma Aldrich	Deisenhofen, Germany
DMEM high glucose	Invitrogen	Karlsruhe, Germany

DMEM/F12 (1:1)	Invitrogen	Karlsruhe, Germany
DMSO	Sigma Aldrich	Deisenhofen, Germany
DNA ladder (100bp / 1kbp)	PeqLab	Erlangen, Germany
DNase (cell culture)	Cell Systems	St. Katharinen, Germany
DNaseI (mol. bio.)	Invitrogen	Karlsruhe, Germany
dNTPs	PeqLab	Erlangen, Germany
Doxycycline	Sigma Aldrich	Deisenhofen, Germany
EDTA	Sigma Aldrich	Deisenhofen, Germany
EGF	R&D Systems	Wiesbaden, Germany
Ethanol	Sigma Aldrich	Deisenhofen, Germany
Ethidium bromide	Sigma Aldrich	Deisenhofen, Germany
FCS	Invitrogen	Karlsruhe, Germany
Fentanyl	Janssen-Cilag	Beerse, Belgium
FGF2	R&D Systems	Wiesbaden, Germany
Flumazenil	Hikma-Pharmaceuticals	London, UK
Forene®	Abbott	Baar, Switzerland
G418 solution	PAA	Pasching, Austria
Gelatine	Sigma Aldrich	Deisenhofen, Germany
Glucose	Sigma Aldrich	Deisenhofen, Germany
Glycerol	Sigma Aldrich	Deisenhofen, Germany
Glycin	Sigma Aldrich	Deisenhofen, Germany
H_2O_2	Sigma Aldrich	Deisenhofen, Germany
HCl	Sigma Aldrich	Deisenhofen, Germany
HEPES	Sigma Aldrich	Deisenhofen, Germany
Human brain RNA	Aligent Technologies	Santa Clara, USA
Insulin	Sigma Aldrich	Deisenhofen, Germany
Isopropanol	Sigma Aldrich	Deisenhofen, Germany
IPTG	Sigma Aldrich	Deisenhofen, Germany
Knockout-DMEM	Invitrogen	Karlsruhe, Germany
L-glutamine (10x)	Invitrogen	Karlsruhe, Germany
Laminin (Ln)	Sigma Aldrich	Deisenhofen, Germany
Lysozyme	Sigma Aldrich	Deisenhofen, Germany
Matrigel (MG)	BD Biosciences	Heidelberg, Germany
2-Mercaptoethanol	Invitrogen	Karsruhe, Germany
Mowiol	Sigma Aldrich	Deisenhofen, Germany
N2 supplement (100x)	Invitrogen	Karlsruhe, Germany
N2 supplement (100x)	PAA	Pasching, Austria
NaHCO3	Sigma Aldrich	Deisenhofen, Germany
Naloxon	Ratiopharm	Ulm, Germany

NaCl	Sigma Aldrich	Deisenhofen, Germany
NaOH	Sigma Aldrich	Deisenhofen, Germany
Neurobasal medium	Invitrogen	Karlsruhe, Germany
Non-essential amino acids (10x)	Invitrogen	Karlsruhe, Germany
Paraformaldehyde	Sigma Aldrich	Deisenhofen, Germany
PBS	Invitrogen	Karlsruhe, Germany
Penicillin Streptomycin	Invitrogen	Karlsruhe, Germany
PFA	Sigma Aldrich	Deisenhofen, Germany
PhosSTOP™ tablets	Roche Diagnostics	Basel, Swizerland
Poly-L-ornithine	Sigma Aldrich	Deisenhofen, Germany
Puromycin	PAA	Pasching, Austria
RIPA	Sigma Aldrich	Deisenhofen, Germany
SDS	Sigma Aldrich	Deisenhofen, Germany
Serum Replacement	Invitrogen	Karlsruhe, Germany
Sodium azide	Sigma Aldrich	Deisenhofen, Germany
Sodium pyruvate (10x)	Invitrogen	Karlsruhe, Germany
Sucrose	Sigma Aldrich	Deisenhofen, Germany
Taq DNA polymerase	Invitrogen	Karlsruhe, Germany
TEMED	Sigma Aldrich	Deisenhofen, Germany
Tissue-Tek®	Weckert Labortechnik	Kitzingen, Germany
Tris	Sigma Aldrich	Deisenhofen, Germany
Triton-X-100	Sigma Aldrich	Deisenhofen, Germany
Trypanblue	Invitrogen	Karlsruhe, Germany
Trypsin inhibitor (TI)	Invitrogen	Karlsruhe, Germany
Trypsin-EDTA	Invitrogen	Karlsruhe, Germany
Tryptone	Sigma Aldrich	Deisenhofen, Germany
Vectashield® mounting medium	Vector Labs	Burlingame, USA
Xylene cyanol	Sigma Aldrich	Deisenhofen, Germany
Yeast extract	Sigma Aldrich	Deisenhofen, Germany

2.4. Cell lines and animals

Cell line or mouse strain	Source
E. coli DH5a	Invitrogen, Deisenhofen, Germany
E. coli DH10b	Invitrogen, Deisenhofen, Germany
E. coli Stbl3	Invitrogen, Deisenhofen, Germany
HEK-293FT	Leiden, Netherlands, Dr. Alex Van der Eb
AAV-293	Agilent, Palo Alto, California, USA
hESC line H9.2	Haifa, Israel (Itskovitz-Eldor et al., 2000)
hESC line I3	Haifa, Israel (Itskovitz-Eldor et al., 2000)
iPSC line MJD	Bonn, Germany (Koch et al., 2011)
iPSC line WT	Bonn, Germany (Koch et al., 2011)
iPSC line 4247#2	Bonn, Germany
iPSC line 4247#8	Bonn, Germany
Mouse strain C57BL/6	Charles River, Wilmington, USA
Mouse strain C57BL/6 / $Rag2^{-/-}$	Charles River, Wilmington, USA

2.5. Cell culture media

All cell culture reagents are prepared under sterile conditions and stored at 4°C; % = v/v.

Human pluripotent stem (hPS) cell medium			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Knockout-DMEM	79%	Invitrogen	Karlsruhe, Germany
Serum replacement	20%	Invitrogen	Karlsruhe, Germany
Non-essential amino acids	1x	Invitrogen	Karlsruhe, Germany
L-glutamine	1x	Invitrogen	Karlsruhe, Germany
2-Mercaptoethanol	0.1 mM	Invitrogen	Karlsruhe, Germany
FGF2	4 ng/ml	R&D Systems	Wiesbaden, Germany
Penicillin / Streptomycin	1x	Invitrogen	Karlsruhe, Germany

Embryoid body (EB) medium			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Knockout-DMEM	79%	Invitrogen	Karlsruhe, Germany
Serum replacement	20%	Invitrogen	Karlsruhe, Germany
Non-essential amino acids	1x	Invitrogen	Karlsruhe, Germany
L-glutamine	1x	Invitrogen	Karlsruhe, Germany
Penicillin / Streptomycin	1x	Invitrogen	Karlsruhe, Germany

Neural stem cell (N2) medium			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
DMEM:F12 (1:1)	98%	Invitrogen	Karlsruhe, Germany
N2 supplement	1x	PAA	Karlsruhe, Germany
Glucose	1.5 mg/ml	Sigma Aldrich	Deisenhofen, Germany
Penicillin / Streptomycin	1x	Invitrogen	Karlsruhe, Germany

Neuronal generation (NG) medium			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Neural stem cell (N2) medium	50%		
Neurobasal medium	48%	Invitrogen	Karlsruhe, Germany
B27 supplement	0.5x	Invitrogen	Karlsruhe, Germany
cAMP	100 ng/ml	Sigma Aldrich	Deisenhofen, Germany
Penicillin / Streptomycin	0.5x	Invitrogen	Karlsruhe, Germany

Mouse embryonic fibroblast (MEF) medium			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
DMEM high glucose	87%	Invitrogen	Karlsruhe, Germany
FCS, heat inactivated	10%	Invitrogen	Karlsruhe, Germany
Sodium pyruvate	1x	Invitrogen	Karlsruhe, Germany
Non-essential amino acids	1x	Invitrogen	Karlsruhe, Germany
L-glutamine	1x	Invitrogen	Karlsruhe, Germany
Penicillin Streptomycin	1x	Invitrogen	Karlsruhe, Germany

induced pluripotent stem cell (iPSC) medium			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
DMEM/F-12	1x	Invitrogen	Karlsruhe, Germany
KnockOut™ Serum Replacement	20%	Invitrogen	Karlsruhe, Germany
GlutaMAX™-I Supplement	1x	Invitrogen	Karlsruhe, Germany
Non-essential amino acids	100 µM	Invitrogen	Karlsruhe, Germany
β-mercaptoethanol	100 µM	Invitrogen	Karlsruhe, Germany
Penicillin Streptomycin	1x	Invitrogen	Karlsruhe, Germany
Basic-FGF	4 ng/ml	Sigma Aldrich	Deisenhofen, Germany

Neural stem cell freezing medium			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Serum replacement	70%	Invitrogen	Karlsruhe, Germany
Cytocoon™ Buffer II	20%	Evotec	Hamburg, Germany
DMSO	10%	Sigma Aldrich	Deisenhofen, Germany

FCS-based freezing medium			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
FCS, heat inactivated	90%	Invitrogen	Karlsruhe, Germany
DMSO	10%	Sigma Aldrich	Deisenhofen, Germany

Virus freezing medium			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Hepes buffered solution (HBS)	92%		
BSA solution (7.5%)	8%	Invitrogen	Karlsruhe, Germany

2.6. Cell culture additives

Reagent	Concentration	Solvent	Manufacturer	Registered office
Ascorbic acid	200 mM	H_2O	Sigma Aldrich	Deisenhofen, Germany
B27	pure supplement mix	pure supplement mix	Invitrogen	Karlsruhe, Germany
BrdU	10 mg/ml	H_2O	Invitrogen	Karlsruhe, Germany
Chloroquin	50 mM	H_2O	Sigma Aldrich	Deisenhofen, Germany
DAPT	10 mM	DMSO/Ethanol (1:4)	Sigma Aldrich	Deisenhofen, Germany
DNAse	0.1%	PBS	Cell Systems	St. Katharinen, Germany
Doxycycline	1 mg/ml	H_2O	Sigma Aldrich	Deisenhofen, Germany
EGF	10 µg/ml	0.1% BSA in H_2O	R&D Systems	Wiesbaden, Germany
FGF2	10 µg/ml	0.1% BSA in H_2O	R&D Systems	Wiesbaden, Germany
G418	50 mg/ml	H_2O	PAA	Pasching, Austria
Puromycin	1 mg/ml	H_2O	PAA	Pasching, Austria

2.7. Cell culture solutions

HEPES Buffered Solution (HBS)			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
NaCl	16 mg/ml	Sigma Aldrich	Deisenhofen, Germany
HEPES	10 mg/ml	Sigma Aldrich	Deisenhofen, Germany
KCl	0.76 mg/ml	Sigma Aldrich	Deisenhofen, Germany
Na_2HPO_4	0.25 mg/ml	Sigma Aldrich	Deisenhofen, Germany
Glucose	0.25 mg/ml	Sigma Aldrich	Deisenhofen, Germany
H_2O; adjust pH = 7.05			

Poly-L-ornithine (PO) coating solution			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
H_2O	100%		
Poly-L-ornithine (PO)	10 mg/ml	Sigma Aldrich	Deisenhofen, Germany

Laminin (Ln) coating solution			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
H_2O	100%		
Laminin (Ln)	1 µg/ml	Sigma Aldrich	Deisenhofen, Germany

Matrigel (MG) coating solution			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
DMEM:F12	97%	Invitrogen	Karlsruhe, Germany
Matrigel (MG)	3%	BD Biosciences	Heidelberg, Germany

Trypsin/EDTA (TE) 1x			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
PBS	90%	Invitrogen	Karlsruhe, Germany
Trypsin/EDTA (TE) 10x	10%	Invitrogen	Karlsruhe, Germany

Trypsin inhibitor (TI)			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
PBS	100%	Invitrogen	Karlsruhe, Germany
Trypsin inhibitor (> 700 units/mg)	0.25 mg/ml	Invitrogen	Karlsruhe, Germany

2.8. Plasmids

Plasmid	Source or parent DNA sequence
pMD2.G	Gift from Didier Trono, Lausanne, Switzerland
psPAX2	Gift from Didier Trono, Lausanne, Switzerland
pAAV-SEPT-Acceptor	Gift grom Todd Waldman, Washington, USA
pAAV-RC	Stratagene
pHELPER	Stratagene
pLVX-EF1α	Modified from pLVX-Tight-Puro (Clontech)
pLVX-EF1α RFP.NLS	RFP cDNA from pCMVBrainbow 1.1 (Gift from Daniel Trageser)
pLVX EF1α Cre	Cre cDNA from pCMV-HTN.Cre (Gift from Michael Peitz)

2.9. Bacterial Artificial Chromosomes

BAC	Tag	MCB	Source or parent DNA sequence
AURKA	LAP	MCB_0002476	Gift from Ina Poser, Dresden, Germany*
CDK2AP1	LAP	MCB_0003855	Gift from Ina Poser, Dresden, Germany*
FGD1	LAP	MCB_0003318	Gift from Ina Poser, Dresden, Germany*
JARID1C	LAP	MCB_0003325	Gift from Ina Poser, Dresden, Germany*
JUN	LAP	MCB_0005073	Gift from Ina Poser, Dresden, Germany*
MAP1LC3B	NFLAP	MCB_0005264	Gift from Ina Poser, Dresden, Germany*
MARK2	LAP	MCB_0002195	Gift from Ina Poser, Dresden, Germany*
MECP2	LAP	MCB_0003305	Gift from Ina Poser, Dresden, Germany*
N-PAC	LAP	MCB_0003407	Gift from Ina Poser, Dresden, Germany*
NCL	LAP	MCB_0002472	Gift from Ina Poser, Dresden, Germany*
NCSTN	LAP	MCB_0003576	Gift from Ina Poser, Dresden, Germany*
PCNA murine	LAP	MCB_0002066	Gift from Ina Poser, Dresden, Germany*
PCNA	LAP	MCB_0005616	Gift from Ina Poser, Dresden, Germany*
PSEN1	NFLAP	MCB_0003584	Gift from Ina Poser, Dresden, Germany*
PSEN2	NFLAP	MCB_0003586	Gift from Ina Poser, Dresden, Germany*
RBP-J	LAP	MCB_0004087	Gift from Ina Poser, Dresden, Germany*
RPS27A	LAP	MCB_0003652	Gift from Ina Poser, Dresden, Germany*
RUVBL1	LAP	MCB_0002791	Gift from Ina Poser, Dresden, Germany*
SETD1B	LAP	MCB_0004774	Gift from Ina Poser, Dresden, Germany*
UHRF1	LAP	MCB_0003868	Gift from Ina Poser, Dresden, Germany*
VCP	LAP	MCB_0004449	Gift from Ina Poser, Dresden, Germany**

*Max Planck Institute for Molecular Cell Biology and Genetics, Hyman Group, Dresden, Germany

2.10. Primers

Cloning primers*		
Target	**Identification**	**Sequence (5' → 3')**
EF1α	Ef1α-BamHI-fw	GCGGCGGGATCCGGGTCGAAATTCCTCACGACACC
	Ef1a-ClaI-rev	GCGGCGATCGATGCGCCCGACGATAAGCTTTGC
Cre	Cre-NotI-fw	GCAGCAGCGGCCGCGCCGCCACCATGGGCCATCACCATCACCATC
	Cre-MluI-rev	GCAGCAACGCGTTCATCAAATCGCCATCTTCCAGCAG
dTomatoe.NLS	dTom-NotI-fw	GCAGCAGCGGCCGCGCCGCCACCATGGTGAGCAAGGGCGAGG
	dTom.NLS MluI-rev	TGCTGCACGCGTTTACTACACCTTCCTCTTCTTCTTCAACATGCCTCCGCCCTTGTACAGCTCG TCCATGCCG
Cre mRNA	Cre-T7 Kozak-FLAG-fw	GCAGCATAATACGACTCACTATAGGGGCAGCAGCCGCCACCATGGATTACAAGGATGACGAC GATAAGGGCCATCACCATCACCATCACGG
	Cre-PolyA-rev	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAATCGCCATCTTCCAGCAGGC
Atxn3 N-Term. LHA	ATXN3-201-LHA (EcoRINotI)-fw	GCAGCAGAATTCGCGGCCGCGCCGTTGGCTCCAGACAAATAAACATG
	ATXN3-201-LHA (KpnI)-rv	GCAGCAGGTACCGTGCCAAAGCTGTCCTGGGGAAGGC
Atxn3 N Term. RHA	ATXN3-201-RHA (BamHI)-fw	GCAGCAGGATCCATGGTATCATGTTTCATAATAGTTTC
	ATXN3-201-RHA (SalI,NotI)-rv	GCAGCAGTCGACGCGGCCGCACTAAGAGGGATTTCTTTCGGGTAAG
Selection Cassette Atxn3 N-Term	FIRESNEO (SmaI)-fw	GCAGCACCCGGGATAACTTCGTATAGCATACATTATACGAAGTTATGGAATTAATTCGC
	FIRESNEO (XhoI)rv	GCAGCACTCGAGATAACTTCGTATAATGTATGCTATACGAAGTTATTCCCAACTCATC
Atxn3 N-Term mutagenesis primer	ATXN3-201 (LHA-ATG-FLAG)fw	CTCCAGACAAATAAACATGGATTACAAGGACGACGATGACAAGGAGTCCATCTTCCACGAGA
	ATXN3-201 (LHA-ATG-FLAG)rv	TCTCGTGGAAGATGGACTCCTTGTCATCGTCGTCCTTGTAATCCATGTTTATTTGTCTGGAG

Sequencing primers*		
Target (direction)	**Identification**	**Sequence (5' → 3')**
Upstream of Tet response element (→ 3')	seqTet_fw	ATCGATGAGGCCCTTTCGTCTTCACTCGAG
Downstream of PGK promoter (→ 5')	seqPGK_rv	CCTACCGGTGGATGTGGAATG
Upstream of T7 sequence (→ 3')	seqT7_fw	TAATACGACTCACTATAGG

* All primers were obtained from Life Technologies, Karlsruhe, Germany.

2.11. siRNA

Target	Identification	Sequence (5' → 3')
IRF3	Irf3.1_sense	GGAGGAUUUCGGAAUCUUC-dTdT
	Irf3.1_antisense	GAAGAUUCCGAAAUCCUCC-dTdG

siRNA was obtained from Biomers, Ulm, Germany.

2.12. Antibodies

Primary antibody	Dilution	Source
Actin	1:2000	Sigma Aldrich
β-III-tubulin (TuJ1; ms)	1:1000	Covance
β-III-tubulin (rb)	1:3000	Covance
BACE1 (AB7520)	1:1000	Walter et al. 2001
BrdU	1:50	Becton Dickinson
cleaved Notch1 (Val1744)	1:1000	Cell Signaling
Cre (ms, MMS-106P)	1:500	Covance
DACH1	1:50	ProtenTech
GFP (rb)	1:3000	Abcam
GFP (ms)	1:500	Roche
FLAG-tag	1:800	Sigma Aldrich
GABA	1:500	Sigma Aldrich
GFAP	1:1000	DAKO
MAP2ab	1:500	Chemicon
Nestin	1:500	R&D Systems
NeuN	1:100	Millipore
Neurofilament (HO14)	1:100	gift from Virginia Lee, Philadelphia, USA
Nicastrin	1:2000	Sigma Aldrich
Nurr1	1:50	Santa Cruz
Otx2	1:200	R&D Systems
PLZF	1:50	Calbiochem
PS1 (AB3109)	1:500	Prager et al. 2007
PS1 (APS18)	1:200	GeneTex
Sox2	1:300	R&D Systems
TH	1:500	Chemicon
ZO1	1:50	Zymed

Secondary antibody	Dilution	Source
Cy3 gt-anti-ms	1:250	Jackson / Dianova
Cy3-gt-anti-rb	1:250	Jackson / Dianova
Cy3-dk-anti-gt	1:250	Jackson / Dianova
Cy3-gt-anti-rat	1:250	Jackson / Dianova
Cy5-gt-anti-rat	1:250	Jackson / Dianova
Cy5-gt-anti-ms	1:250	Jackson / Dianova
FITC-gt-anti-ms	1:250	Jackson / Dianova
FITC-gt-anti-rb	1:250	Jackson / Dianova
HRP-gt-anti-ms	1:10000	Thermo Scientific
HRP-gt-anti-rb	1:10000	Thermo Scientific

ms = mouse; rb = rabbit; gt = goat; dk = donkey

2.13. Molecular biology reagents

Affinity purification lysis buffer (AP-lysis)			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
NaCl	150 mM	Sigma Aldrich	Deisenhofen, Germany
Tris pH 7,5	50 mM	Sigma Aldrich	Deisenhofen, Germany
Glycerol	5%	Sigma Aldrich	Deisenhofen, Germany
IGEPAL-CA-630	1%	Sigma Aldrich	Deisenhofen, Germany
$MgCl_2$	1 mM	Sigma Aldrich	Deisenhofen, Germany
Benzoase	200 U	Merck	Darmstadt, Germany
EDTA-free complete protease inhibitor	1	Roche	Basel, Schweiz

Affinity purification wash buffer I (AP-wash-I)			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
NaCl	150 mM	Sigma Aldrich	Deisenhofen, Germany
Tris pH 7,5	50 mM	Sigma Aldrich	Deisenhofen, Germany
Glycerol	5%	Sigma Aldrich	Deisenhofen, Germany
IGEPAL-CA-630	0,05%	Sigma Aldrich	Deisenhofen, Germany

Affinity purification wash buffer II (AP-wash-II)			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
NaCl	150 mM	Sigma Aldrich	Deisenhofen, Germany
Tris pH 7,5	50 mM	Sigma Aldrich	Deisenhofen, Germany
Glycerol	5%	Sigma Aldrich	Deisenhofen, Germany

Affinity purification elution buffer (AP-elution)			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Urea	2 M	Sigma Aldrich	Deisenhofen, Germany
Tris pH 7,5	50 mM	Sigma Aldrich	Deisenhofen, Germany
Chloroacetamide	5 mM	Sigma Aldrich	Deisenhofen, Germany

Blotting buffer (10x)			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Tris	30 g	Sigma Aldrich	Deisenhofen, Germany
Glycin	144 g	Sigma Aldrich	Deisenhofen, Germany
H_2O to 1000 ml; add 20% methanol before use to 1x blotting buffer			

Borate Buffer			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Sodium borate	3.8 g	Sigma Aldrich	Deisenhofen, Germany
H_2O to 100 ml (pH 8.5)	add		

Blocking solution			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
PBS	90%	Invitrogen	Karlsruhe, Germany
FCS	10%	Invitrogen	Karlsruhe, Germany
Triton-X-100 (intracelluar epitopes)	0.1%	Sigma Aldrich	Deisenhofen, Germany

DNA loading buffer			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Tris solution (0.1 M; pH 8.0)	70%	Sigma Aldrich	Deisenhofen, Germany
Glycerol	29.2%	Sigma Aldrich	Deisenhofen, Germany
Bromphenol blue	0.4%	Sigma Aldrich	Deisenhofen, Germany
Xylene cyanol	0.4%	Sigma Aldrich	Deisenhofen, Germany

LB medium			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Tryptone	10 g	Sigma Aldrich	Deisenhofen, Germany
Yeast extract	5 g	Sigma Aldrich	Deisenhofen, Germany
NaCl	5 g	Sigma Aldrich	Deisenhofen, Germany
NaOH solution (1 M)	1 ml	Sigma Aldrich	Deisenhofen, Germany
H_2O to 1000 ml; autoclave and store at 4°C	add		

LB agar plates			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Agar	7 g	Sigma Aldrich	Deisenhofen, Germany
LB medium to 1000 ml; autoclave and store at 4°C	add		

Mowiol/DABCO			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Tris solution (0.2 M; pH 8.5)	12 ml	Sigma Aldrich	Deisenhofen, Germany
H_2O	6 ml		
Glycerol	6 g	Sigma Aldrich	Deisenhofen, Germany
Mowiol	2.6 g	Sigma Aldrich	Deisenhofen, Germany
DABCO	0.1 g	Sigma Aldrich	Deisenhofen, Germany

Paraformaldehyde (PFA) fixation solution (4%)			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Paraformaldehyde	40 g	Sigma Aldrich	Deisenhofen, Germany
H_2O to 1000 ml (heat to dissolve; filter; pH 7.4)	add		

Protein loading buffer			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Tris solution (0.1 M; pH 6.8)	76%	Sigma Aldrich	Deisenhofen, Germany
Glycerol	20%	Sigma Aldrich	Deisenhofen, Germany
SDS	4%	Sigma Aldrich	Deisenhofen, Germany
Bromphenol blue	0.25%	Sigma Aldrich	Deisenhofen, Germany

Resolving gel buffer			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
1.5 M Tris solution (pH 8.8)	99.6%	Sigma Aldrich	Deisenhofen, Germany
SDS	0.4%	Sigma Aldrich	Deisenhofen, Germany

Stacking gel buffer			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
1.5 M Tris solution (pH 6.8)	99.6%	Sigma Aldrich	Deisenhofen, Germany
SDS	0.4%	Sigma Aldrich	Deisenhofen, Germany

SDS-PAGE running buffer			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Tris	30 g	Sigma Aldrich	Deisenhofen, Germany
Glycin	144 g	Sigma Aldrich	Deisenhofen, Germany
SDS	10 g	Sigma Aldrich	Deisenhofen, Germany
H_2O to 1000 ml (pH 8.0)	add		

Saline-sodium citrate (SaSoC) buffer (20x)			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
NaCl	3 M	Sigma Aldrich	Deisenhofen, Germany
Trisodium citrate	300 mM	Sigma Aldrich	Deisenhofen, Germany
pH 7.0			

TBST (10x)			
Reagent	**Concentration**	**Manufacturer**	**Registered office**
Tris	121 g	Sigma Aldrich	Deisenhofen, Germany
NaCl	87 g	Sigma Aldrich	Deisenhofen, Germany
Tween-20	1%	Sigma Aldrich	Deisenhofen, Germany
H_2O to 1000 ml	add		

2.14. Molecular biology enzymes, kits and compounds

Enzyme / Kit	Manufacturer
5-Methylcytidine-5'-Triphosphate	Trilink
Antarctic Phosphatase	New England Biolabs
FastAP™ Phosphatase	Thermo Scientific
LongAmp® Taq DNA Polymerase	New England Biolabs
MEGAScript® T7 Kit	Ambion
mini Quick Spin RNA Columns	Roche
NucleoBond® BAC 100	Macherey-Nagel
peqGOLD™ Gel Extraction Kit	PeqLab
peqGOLD™ Plasmid Miniprep Kit I	PeqLab
Phusion® High-Fidelity DNA Polymerase	New England Biolabs
Poly(A) Polymerase Tailing Kit	Epicentre
Pseudouridine-5'-Triphosphate	Trilink
PureLink™ HiPure Plasmid Filter Maxiprep Kit	Invitrogen
QuikChange®	Agilent
Restriction Endonucleases	New England Biolabs
ScriptCap 2'-O-Methyltransferase	Epicentre
ScriptCap m7G Capping System	Epicentre
T4 DNA Ligase	New England Biolabs
Taq DNA Polymerase	Invitogen
Zymoclean™ DNA Clean & Concentrator-25	Hiss Diagnostics

2.15. Software

Computer program	Manufacturer
ApE - A Plasmid Editor 2.0.39	M. Wayne Davis
AxioVision 40 4.5.0.0	Carl Zeiss
Brain Explorer 2	Allen Reference Atlas
Cytoscape 2.7	The Cytoscape Collaboration
Excel 2008	Microsoft
FlowJo 6.8	Tree Star
Fiji 1.47n	http://fiji.sc/
GenomeStudio 2011	Illumina
Image J 1.42q	NIH
InDesign CS5	Adobe
MaxQuant 1.2.2.5	MPI for Biochemistry, Munich
Quantity One 4.6.8	Bio Rad
Perseus 1.1.0.17	MPI for Biochemistry, Munich
Photoshop CS3	Adobe
Prism 5	Graphpad
QuikChange® Primer Design	Agilent
UniProt	UniProt Consortium
Word 2011	Microsoft

3. Methods

3.1. Generation of iPS cells

All cell culture experiments were performed in a sterile vertical/horizontal laminar flow hood with sterile media, plastic and glass instruments. Prepared buffers and solutions were autoclaved, where applicable, or filtered using a 0.2 µm sterile filter. Cells were cultivated in an incubator at 37°C, 5% CO_2 and saturated air humidity.

Adult human fibroblasts were cultivated in mouse embryonic fibroblast (MEF) medium on gelatin-coated dishes. For infection, 70 000 cells were transferred to a gelatin-coated well of a 12-well dish and cultivated to a confluency of 80-90%. Four Sendai viral vectors each expressing Klf4, Oct3/4, Sox2 or c-Myc were used for transduction at a multiplicity of infection (MOI) of 4. 24 h later, the medium was changed and cells were cultivated for additional 7 days. At day 8, transduced adult human fibroblasts were separated using TrypLE Select and 50 000 cells were transferred to irradiated MEFs and cultivated in iPSC medium for the following 3-4 weeks. Emerging colonies were mechanically isolated using a scalpel, transferred to MEF covered dishes and cultivated for ≥5 passages. Generated iPSC lines were characterized by analysis for the expression of the pluripotency markers Tra1-60 and Tra1-81. Genome-wide single nucleotide polymorphism analysis was performed to validate genomic integrity of derived iPSC lines and teratoma assays to validate pluripotency (Wesselschmidt, 2011).

3.2. In vitro differentiation of hPS cells into lt-NES cells

HPS cells were cultivated on a layer of irradiated mouse fibroblasts in hPSC medium, which was changed daily. Cells were passaged using collagenase IV (1 mg/ml). Differentiation of hPSC (hESCs and iPSCs) into lt-NES cells was performed as described previously (Koch et al., 2009). Embryoid bodies were generateded by cultivation of hPSCs in EB medium as floating aggregates for four days. EBs were transferred to poly-l-ornithine (PO) coated tissue culture dishes and propagated in N2 (PAA) medium containing FGF2 (10 ng / ml; R&D Systems). Within 10 days, neural tube-like structures developed in the EB outgrowth and were mechanically isolated and propagated for another three days as free-floating neurospheres in N2 medium containing FGF2 (10 ng/ml). Neurospheres were dissociated by incubation with trypsin/EDTA (TE, Invitrogen) and trypsin inhibitor (TI, Invitrogen), mechanically triturated into single cells and plated on poly-l-ornithine (3h, 37°C) Laminin (≥ 12h, 4°C; PO/Ln; Sigma Aldrich) coated tissue culture dishes. Generated lt-NES cell lines were cultivated in N2 medium containing FGF2, EGF (each 10 ng / ml) and B27 supplement (1 µl/ml; Invitrogen) on PO/Ln coated tissue culture dishes, passaged using TE/TI and seeded at densities of 0.5 - 0.8 x 10^6 cells / cm^2. Lt-NES cell lines show a highly clonogenic potential, and can be passaged for more than 100 passages and were monitored daily regarding proliferation rate and morphological integrity. lt-NES cells were stored after TE/TI treatment in neural stem cell freezing medium initially at -80°C (<7 days), for long-term storage in liquid nitrogen storage tanks.

3.3. In vitro differentiation of lt-NES cells into neuronal cultures

For initiation of terminal differentiation, lt-NES cells were transferred to matrigel (MG)-coated tissue culture dishes and cells were grown to confluence. The culture medium was changed to NGc medium, which was exchanged every other day. Usually, cells were analyzed four weeks after induction of differentiation but could be matured for more than three months.

3.4. BAC transfection of lt-NES cells

Bacterial artificial chromosomes (BACs) have the ability to carry large amounts of genetic information (up to 350 kb) and can easily be modified by recombination methods. Most genomic loci can be found on established mapped BAC libraries, originating from human genome sequencing projects. Tagged variants of these BACs were provided as stab culture by Ina Poser (Max Planck Institute for Molecular Cell Biology and Genetics, Dresden, Germany). BACs were prepped according to manufacturers instructions (BAC-100, Macherey-Nagel). lt-NES cells from iPSCs or hESCs were pelleted (250 g, 5 min), resuspended and incubated in Cytocoon™ Buffer II (30 min, RT). 2 µg circular BAC DNA (≤6 µl) were transferred to a nucleofection cuvette, cells were resuspended in Nucleofection Buffer V (100 µl, Amaxa) and added directly to the DNA. Immediately after nucleofection (program A-33), cells were replated. Chemoselection was initiated 5 days after nucleofection by addition of G418 (50 µg / ml). Concentration of G418 was increased in 50 µg steps every other day until increased cell death could be detected. Cells were cultivated without selection for two passages and chemoselection was reapplied at last-used concentrations, increasing each day until concentrations of 75 µg were reached. Generated BAC-transgenic cell lines were continuously cultivated in proliferation medium supplemented with G418 (≥ 50 µg / ml).

3.5. Immunocytochemical analysis

Cells were fixed in 4% PFA (10 min, RT), washed twice with PBS and blocked with 10% FCS. For staining of intracellular epitopes 0,1% Triton™ X-100 was added. Respective primary antibodies (listed) were applied in blocking solution (2h, RT), and cells were washed thrice with PBS (10 min) before secondary antibodies were applied (1h, RT). DAPI (5 min, RT) or TO-PRO®3 (15 min, RT; Life technologies) solution was used for nuclear counterstaining. Cells were rinsed thrice with PBS before mounting with Moviol/DABCO and covering with a glass coverslip.

3.6. Fluorescence activated cell sorting (FACS) analysis

For FACS analysis, single cells were prepared by trypsinization with subsequent inhibition of trypsin (TE/TI), centrifuged and resuspended in PBS ($3x10^5$ cells / 100 µl). Cells were analyzed with the FACSCalibur® cytometer equipped with an argon-ion laser (488 nm) and the following values were recorded: forward scatter (FSC), side scatter (SSC) and GFP fluorescence. Further data analysis was performed using the FlowJo software.

3.7. Western immunoblotting

Cells were scraped from dishes in the presence of PBS, pelletized and incubated for 1 h on ice in RIPA buffer (Sigma Aldrich) containing protease inhibitors (Roche c0mplete ULTRA^{TM}). Lysates were clarified by centrifugation (>20.000 g, 15 min, 4 °C) and stored in aliquots (-80 °C) or used directly. Protein lysates were mixed with 6x Lämmli buffer and boiled (10 min, 95 °C) before separation on a SDS-PAGE gel (10%). Briefly, the SDS-PAGE gel was composed of a stacking gel (5%) and a resolving gel, consisting of 30% acrylamide/bisacrylamide (37,5:1, Carl Roth), TEMED, 10% APS and the respective gel buffers. Prepared protein lysates were loaded on SDS-PAGE gels and run in MINI-protean chambers (Bio-Rad) at 150V for up to 1,5 h. Separated proteins were blotted from the gel to a previously with MeOH activated PVDF membrane (0,45 µm, Millipore) at 30V o/n. Resulting membranes were blocked with milk powder (1%) solved in PBS-T for 30 min, incubated with an appropriate primary antibody (o/n at 4°C on a rotator) and washed thoroughly the next day. The membranes were incubated with an HRP-linked secondary antibody (1h, RT) and developed using chemoluminiscence (SuperSignal West Dura, Thermo Scientific).

3.8. Fluorescence in situ hybridization (FISH)

I3 ES cell derived lt-NES cell lines transgenic for BACs containing the AURKA, Jun or PCNA gene were subjected to fluorescence in situ hybridization (FISH). Probes were generated by nick translation of respective complete BACs (AURKA.GFP, Jun.GFP or PCNA.GFP) using biotin-dUTP. Individual chromosomes were stained using digoxigenin-dUTP labeled whole chromosome paint probes generated by degenerative oligonucleotide priming PCR (DOP-PCR) (Solovei et al., 2002). For FISH, cells were fixed in 4% PFA, washed in PBS (3x5 min), incubated in 0.5% Triton^{TM} X-100 in PBS for 10 min, and equilibrated in 20% glycerol in PBS (30 min). Cells were frozen in liquid nitrogen and thawed to RT (5 times). Following a PBS washing step (5min, RT) they were incubated for 5 min in 0,1 N HCl, rinsed in 2xSaSoC and stored in 50% formamide/2xSaSoC for ≥ 1 h. The labeled probes were dissolved in a mixture of 50% formamide, 10% dextran sulfate and 1xSaSoC, applied to the cells and sealed. Following a 75°C denaturation step for 2 min, hybridization was carried out in a humid atmosphere at 37°C for 48 h. Cell nuclei were counterstained using DAPI and signals acquired using standard fluorescence microscopy.

Dr. Irina Solovei (University of Munich (LMU) Biozentrum, Department of Biology II Anthropology & Human Genetics, Martinsried) performed FISH analysis and probe generation.

3.9. Design of lentiviral vectors

As a conditional overexpression system, a modified variant of the Lenti-X^{TM} Tet-On Advanced system (Clonetech) was used (the CMV promoter had been exchanged by an EF1α promoter in the pLVX-Tet-On plasmid to regulate expression of $rtTA_{Adv.}$ protein). The cDNA for the respective protein was amplified by PCR (cloning primers listed) from image plasmids (BioCat) or total lt-NES cell cDNA. PCR products were separated by agarose gel electrophoresis, extracted by manual excision of the respective band, isolated by gel purification (peqGOLD™ Gel Extraction Kit, Peqlab), digested using appropriate restriction enzymes (listed) and ligated (T4 DNA Ligase, NEB) into linearized, dephosphorylated (FastAP^{TM}, Fermentas) pLVX-Tight-Puro vector under control of the inducible TRE_{tight} promoter, resulting in $pLVX_{TP}$-gene vectors (listed).

3.10. Production and concentration of lentiviral particles

For production of lentiviral particles, HEK293-FT cells were grown on a poly-ornithine-coated 15 cm dish in MEF medium. 80-90% confluent cells were co-transfected with lentiviral plasmids of the second generation by calcium-phosphate precipitation as described previously (Kutner et al, 2009). Briefly, 7 µg of the envelope plasmid pMD2.G and 15 µg of the packaging plasmid psPAX2 together with 30 µg of the transfer vector were mixed, and H_2O was added to a final volume of 1400 µl. 178 µl of $CaCl_2$ solution (2,5 M) was added, and the solution was slowly mixed with 1400 µl of 2xHBS buffer (pH=7,05). The suspension was incubated for 40 min at RT and added drop wise to the 293FT cells that were pre-incubated with with chloroquin (25 µM, 5 min, 37°C in MEF). The next day, transfected cells were washed with PBS (37°C) once. The supernatants of day two and three were pooled, filtered through a 0,45 µm filter and concentrated by centrifugation (44000 g; 4°C, 2h). Viral particles were resuspended in virus freezing medium and stored in aliquots at -80°C.

3.11. Lentiviral transgenesis of lt-NES cells

To generate a lt-NES cell line for conditional overexpression, lt-NES cells from hES cell line I3 were first transduced with lentiviral particles containing the EF1α regulated $rtTA_{Adv}$ protein. Following chemoselection with G418 (100 µg / ml, PAA), cells were transduced with the respective $pLVX_{TP}$-gene virus to obtain inducible cell lines for the respective gene. Cell lines were continuously cultured in G418 and puromycin containing culture media (Puromycin: 1 µg / ml, Sigma Aldrich; G418: 100 µg / ml, PAA).

3.12. Live cell imaging of lt-NES cell populations

In order to discriminate individual cells within lt-NES cell populations naturally growing at high densities, a fraction the cells was transduced with lentiviral particles containing the red fluorescent protein (RFP) cDNA fused to a nuclear localization sequence (NLS; RFP.NLS) under transcriptional control of the eukaryotic translation elongation factor 1 alpha (EF1α) at a multiplicity of infection (MOI) between 1 and 2 (corresponding to 63-83% of affected cells). Cells were thawed and seeded at densities of 50.000 cells / cm^2 one day before image acquisition, which was performed using a microscope attached to a humidified chamber at 37°C containing 5% CO_2.

3.13. Affinity-based protein-protein interaction analysis

Cell pellets were lyzed in AP-lysis buffer (on ice; 30 min), incubated with magnetic antibody-covered beads (anti-GFP, mouse, Miltenyi), washed 3x with buffer AP-wash-I, 2x with buffer AP-wash-II, loaded to an equilibrated magnetic column, digested at RT for 30 min by addition of 150 ng trypsin (Promega), eluted twice using AP-elution buffer, and stored on two C18 Stage Tips at 4°C. Pulldown was performed on an automated liquid-handling platform (Freedom EVO 200; Tecan). For LC-MS/MS analysis, peptides were eluted from tips and separated on line to the mass spectrometer using an easy nano-LC system (Proxeon Biosystems), coupled to a mass spectrometer (LTQ-Orbitrap, Thermo Fischer Scientific) via a nanoscale LC interface (Proxeon Biosystems). Data was processed using MaxQuant software, searching against the human database concatenated with reversed copies of all sequences and supplemented with frequently observed contaminants using

MASCOT. Volcano plots were generated by combining t test p-values with ratio information, while significance lines in the volcano plot were chosen to correspond to a given false discovery rate (FDR) by a permutation-based method using the software R and the provided script QUBIC-LABELFREE.R (Hubner et al., 2010a). Marco Hein (Max Planck Institute of Biochemistry, Martinsried, Germany) performed affinity purification, measurement and statistical processing and provided normalized data and volcano plots.

3.14. Design of AAV targeting constructs

AAV constructs were designed to target specific sites within a human cell. For that, respective genomic regions were analyzed for interspersed repeats and low complexity DNA sequences (RepeatMasker.org). Homology arms up- and downstream of a synthetic exon promoter trap (SEPT) containing a G418 resistance gene were chosen to each contain less than 50% of masked repeats and span a region of 1000 BP while primers were designed to exclusively bind in non-masked regions (Figure 3.1).

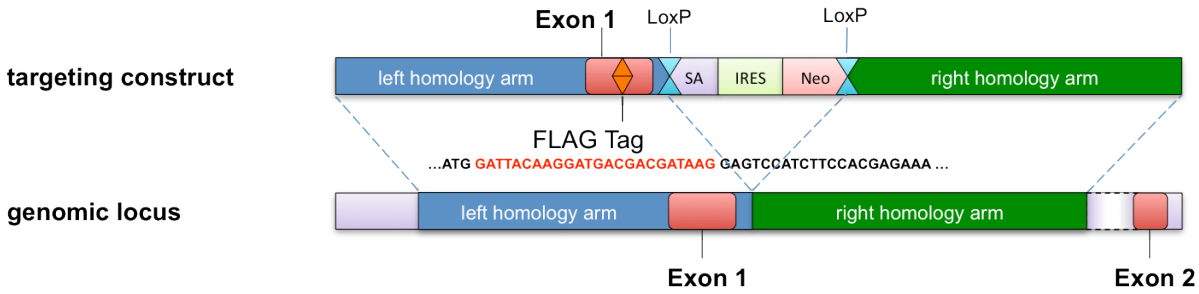


Figure 3.1: Design of AAV targeting constructs
The targeting construct for adeno-associated virus (AAV) mediated gene targeting consists of two homology arms (≥1000 bp) with less than 50% of interspersed repeats and low complexity DNA sequences containing the modified exon. The synthetic exon promoter trap flanked by loxP Cre recombination sites is introduced between both homology arms. The neomycin resistance gene is linked by an internal ribosomal entry site (IRES) to the transcriptional control of the promoter controlling the targeted locus.

Homology arms were amplified with cloning primers (LongAmp® Taq DNA Polymerase, NEB), purified by agarose gel electopuration, extracted (peqGOLD™ Gel Extraction Kit, Peqlab), digested (appropriate restriction enzymes, NEB) and ligated in a linearized, dephosphorylated (FastAP™, Fermentas) pAAV-SEPT-Acceptor together with a purified SEPT cassette to result in a pAAV-*gene*-SEPT vector. Subsequently, a FLAG tag was introduced by mutagenesis PCR

(QuikChange®, Agilent) directly after the initiating ATG of the first exon, resulting in a N-terminally tagged pAAV-FLAG-*gene*-SEPT vector (Fig. 3.1).

3.15. Production of AAV vectors

For production of AAV particles, 293-AAV cells were grown on poly-ornithine-coated dishes in MEF medium. 80-90% confluent cells were co-transfected with plasmids pHELPER and pAAV-RC (Stratagene) by calcium-phosphate precipitation as described previously (Kim et al., 2008). The medium was exchanged the next day, and cell monolayers were scraped into 1 ml PBS and subjected to 4-5 freeze/thaw cycles (-80°C/37°C). The lysate was cleared by centrifugation (20000 g, 10 min, RT). The virus-containing supernatant was aliquoted and stored at -80°C.

3.16. Production of synthetic pseudotyped mRNA

As co-transcriptional incorporation of a cap analogue results in incomplete capping, enzymatic capping and additional 2'O-methylation of in vitro transcription products was conducted supplementary to the introduction of stabilizing nucleotides, resulting in a modified mRNA (Fig. 3.17). cDNA templates for in vitro RNA transcription were created by PCR-based introduction of T7 promoter and 3-prime poly-adenylation sites. RNA was produced according to the manufacturers manual (Ambion). Briefly, ribonucleic acids (ATP, GTP, Kit) together with stable, artificial analogues of UTP (Pseudouridine, Trilink) and CTP (5'-Methylcytidine, Trilink) were used to assemble stable RNA constructs. The resulting RNAs were purified using mini Quick Spin RNA columns (Roche, according to manual) and modified at the 5-prime end using the ScriptCap m7G capping system together with the ScriptCap 2'-O-methyltransferase kit (Epicentre, according to manual). To further enhance stability within cells, the C1-capped mRNAs were polyadenylated (Poly(A) Polymerase Kit, desired A-repeats: 200, Epicentre). To eliminate any residual triphosphates, phosphatase was added for 30 min at 37°C after purification (FastAP, Fermentas) and heat inactivated by incubation at 65°C for 5 min. Resulting C1-capped and polyadenylated mRNA was aliquoted and stored at -80°C.

3.17. mRNA transfection of cultured cells

Two days before transfection of artificial mRNA into a cell line, these cells were transfected with a siRNA targeting the interferon regulatory factor 3 (IRF3) to avoid RIG-I and MDA-5 based transduction of innate immune recognition signaling.
Briefly: 0,25 µg IRF3 siRNA were diluted in 200 µl Opti-MEM I and 6 µl of RNAi-MAX were diluted in a second vial containing 200 µl Opi-MEM I. Both vials were mixed and incubated for 15 min at RT before addition to the cells. Cell media was changed 4h later. Two days later, 1 µg Cre recombinase mRNA was diluted in 200 µl Opti-MEM I while 6 µl RNAi-MAX was diluted in a second vial under same conditions. The growth medium of the cells was replaced with Optifect medium (Opti-MEM I, 0,5% BSA). After mixture of both vials the suspension was incubated for 15 min at RT and added to the cells hereafter. The medium was exchanged after 4h with normal growth medium and cells were recovered for 2 days under normal culture conditions.

3.18. Statistical analysis

Quantitative data was generated in ≥two biological and ≥three technical replicates unless otherwise mentioned. All results show means ±standard deviation (SD). Means and SD were computed using Microsoft Excel 2008 and GraphPad Prism software. A student's t-test was performed to determine whether a significant difference between groups exits. A one-way ANOVA in conjunction with a Dunnett's multiple comparison test was performed to determine whether a significant difference exists between control and test group. For volcano plots, results were plotted using the open source statistical software R and the script QUBIC-LABELFREE.R (Hubner et al., 2010a).

4. Results

4.1. Human pluripotent stem cell-derived neural stem cells for the study of protein-protein interactions

Human pluripotent stem (hPS) cells provide a virtually unlimited source for the generation of a stable neural precursor population and authentic human neurons derived thereof. The first part of the presented work is aimed at harnessing human neural stem cells and their neuronal derivatives for the exploration of protein-protein interactions (PPI) of several proteins involved in diseases (APH1A, APH1B, AURKA, BACE1, CDK2AP1, FGD1, JARID1C, JUN, MAP1LC3B, MARK2, MECP2, MGA, N-PAC, NCL, NCSTN, P53, PCNA, PSEN1, PSEN2, RBPJ, RUVBL2, SETD1B, Sin3A, SmcHD1, RPS27A, Tau, TMEM132A, UHRF1, VCP) introduced by transfection of bacterial artificial chromosomes (BAC) harboring their GFP-tagged isoforms.

4.1.1. Transfection of BACs is suitable for the generation of stable long-term neuroepithelial-like stem (lt-NES) cell line

To permit endogenous expression and regulation of tagged proteins, bacterial artificial chromosome (BAC) vectors carrying C- or N-terminally GFP-tagged candidate proteins containing most, if not all, regulatory elements (Fig. 4.1 a and b, Poser et al., 2008) were used for the generation of transgenic lt-NES cell lines (Fig. 4.1 c). Nucleofection of lt-NES cells with circular BACs (5 µg, $1x10^6$ cells) was followed by a multistep chemoselection process that was developed to permit the derivation of transgenic polyclonal cultures (pools) of homogenous lt-NES cells (Fig. 4.1 c). Routinely, the percentage of GFP-positive cells was quantified by flow cytometric analysis (FACS), the size of the fusion protein assessed by Western immunoblotting (GFP epitope) and the intracellular protein localization examined by fluorescence microscopy (Fig. 4.1 c).

By design, insoluble membrane-associated complexes are less likely to be detected using standard sample processing for AP-MS/MS approaches (Havugimana et al., 2012). Proteins involved in such complexes are less amenable to affinity enrichment and therefore were systematically excluded from experiments (Dubinsky et al., 2012; Schirle et al., 2012). Hence, soluble-, cytoskeleton- or nucleus-associated proteins were primarily used for the generation of BAC-transgenic lt-NES cell lines.

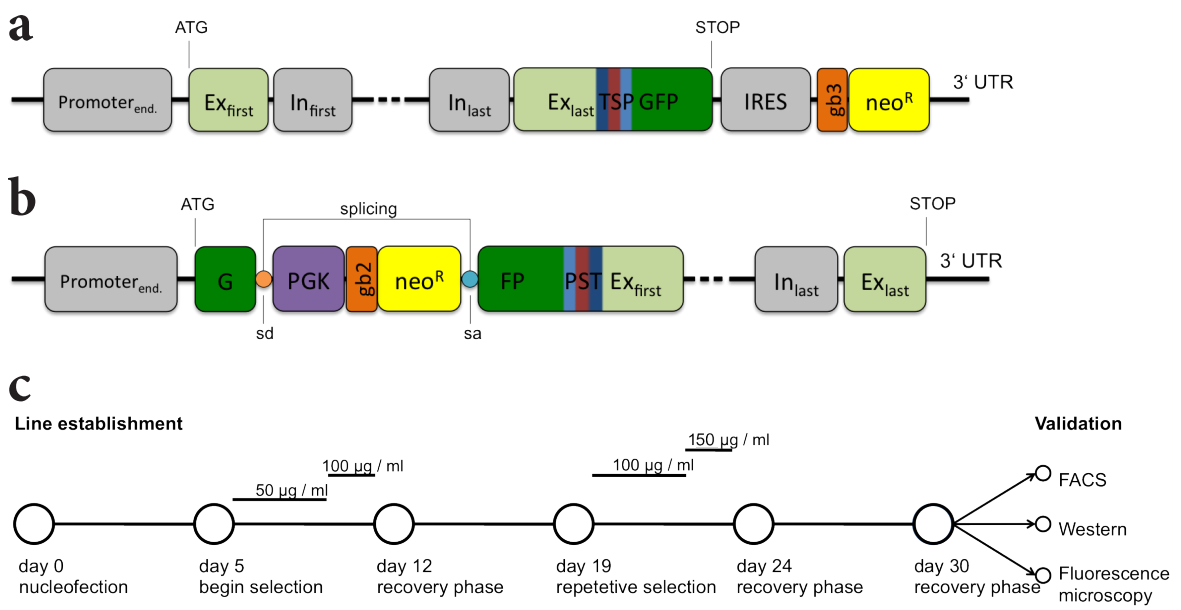
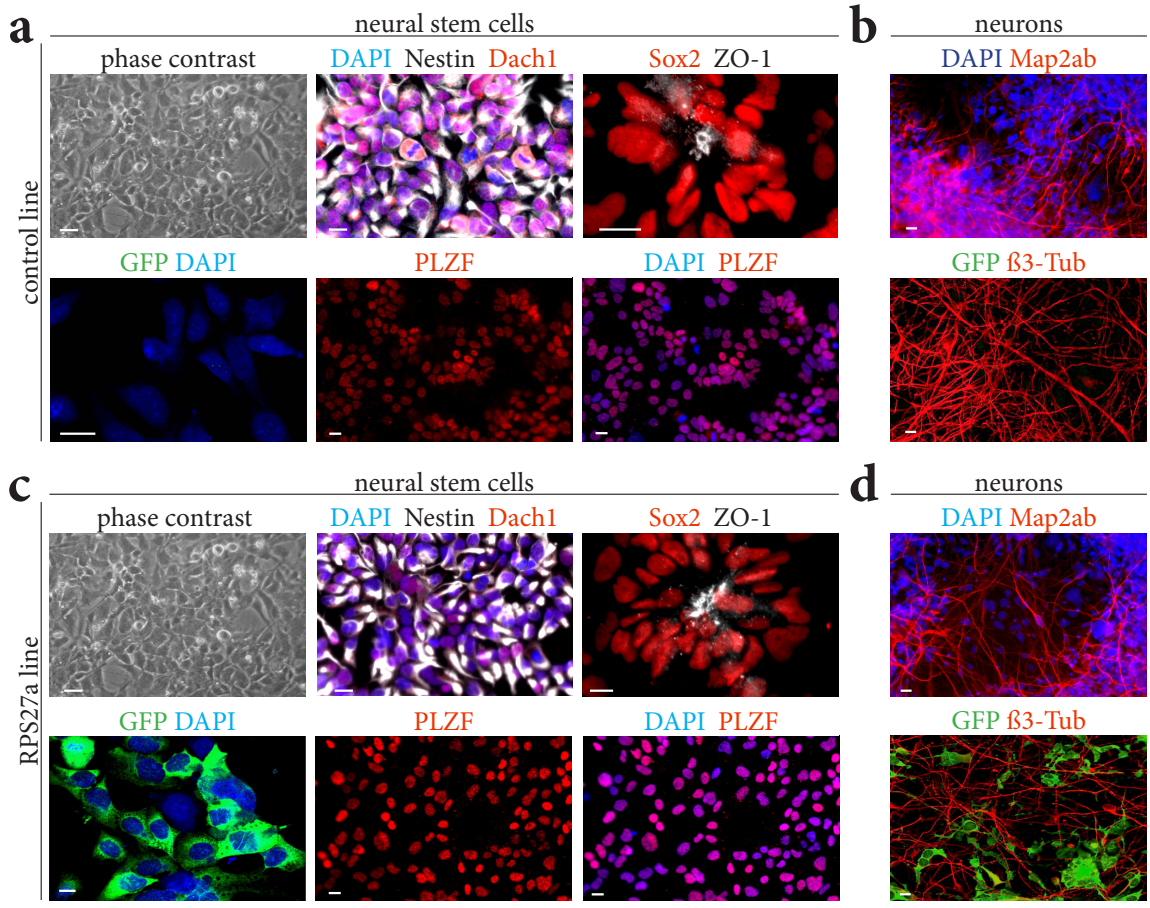


Figure 4.1: BAC-mediated protein tagging
***a and b**: Schematic composition of bacterial artificial chromosomes used for introduction of tagged proteins: Whole-genomic regions of the gene of interest encompasses 5'- and 3'-UTRs, Exons (Ex) and introns (In). Tag composition: TEV cleavage site (T), S-peptide (S) PreScission cleavage site (P), enhanced green fluorescent protein (GFP). **a**: The C-terminal tag is introduced within the last used exon just before the stop codon. An internal ribosomal entry site (IRES), enables the simultaneous expression of the neomycin resistance (neo); a gb3 bacterial promoter allows bacterial amplification of the construct. **b**:* The *N-terminal tag variant is composed of GFP containing an artificial intronic site with phosphoglycerate kinase (PGK) promoter enabling constitutive expression of the neomycin resistance gene fused to the initial exon (Ex_{first}); sd = splice donor, sa = splice acceptor. **c**: Timeline for generation of BAC transgenic cell lines based on lt-NES cells. Nucleofection of lt-NES cells was followed by recovery for 5 days; G418 initially was applied at 50 µg/ml and increased to 100 µg/ml at day 3 of the first selection interval. After a 7 days recovery period, a second selection interval with 100 µg/ml (2 days) and 150 µg/ml for 1 day was conducted, followed by recovery (6 days) and propagation of the cell line under continuous application of antibiotics (100 µg/ml).*

4.1.2. BAC-transgenic lt-NES cell lines retain their differentiation potential

Selected BAC-transgenic pools were analyzed for the maintenance of neural stem cell characteristics and their differentiation potential. Immunocytochemical analysis for the expression of the specific rosette-type neural stem cell markers PLZF, Nestin, Dach1, Sox2 as well as apically accentuated ZO-1, a marker typically found in rosette-stage neural stem cells, was positive for all lines investigated (Fig. 4.2 a, c and e). For analysis of preserved neuronal differentiation potential, growth factors were omitted from the culture. Expression of the neuronal markers MAP2ab and β3-tubulin together with a typical change of the cellular morphology could be observed after 4 weeks (Fig. 4.2 b, d and e).



e

Marker expression

	Nestin	Dach1	Sox2	ZO-1	PLZF	ß3-Tub
AURKA	+	+	+	+	+	+
CDK2AP1	+	+	+	+	+	+
FGD1	+	+	+	+	+	+
JUN	+	+	+	+	+	+
NCL	+	+	+	+	+	+
MAP1LC3B	+	+	+	+	+	+
MECP2	+	+	+	+	+	+
PCNA	+	+	+	+	+	+
RBP-J	+	+	+	+	+	+
RPS27A	+	+	+	+	+	+
RuvBL2	+	+	+	+	+	+
SetD1B	+	+	+	+	+	+
UHRF1	+	+	+	+	+	+
VCP	+	+	+	+	+	+

Figure 4.2: Characterization of BAC transgenic cell lines
***a**: Characterization of control and BAC transgenic proliferating lt-NES cells: Phase contrast image shows a confluent culture of lt-NES cells, immunocytochemical analysis for nestin, PLZF, Sox2 and ZO-1 indicate preserved neural stem cell identity **b**: Immunocytochemical analysis for β3-tubulin and microtubuli-associated protein 2a+b (MAP2ab) demonstrates efficient neurogenic potential of cells. **c**: BAC transgenic lt-NES cells (exemplary shown for RPS27A) demonstrate preserved lt-NES cell marker expression and neurogenic potential (**d**). Scale bar = 10 µm. **e:** The expression of neural stem cell markers and neurogenic potential of BAC transgenic cell lines were confirmed for all lines analyzed.*

4.1.3. Characteristic protein localization is maintained in GFP-tagged proteins of BAC-transgenic lt-NES cells

The cultivation of BAC-transgenic lt-NES pools under continuous application of antibiotics together with the tag design of the predominantly used C-terminally fused GFP (Fig. 4.1 a) ensures the simultaneous expression of both the tagged protein and the resistance gene, while N-terminal tagging (Fig. 4.1 b) may also yield resistant non-expressing cells. Proteins involved in cancer, mental retardation or other developmental-associated diseases were selected and grouped in proteins involved in (I) protein modification/signaling, (II) histone modification and (III) other functions. Among group (I) were (a) aurora kinase A (AURKA), a cell cycle-regulated mitotic serine/threonine kinase localized to mitotic spindles that is overexpressed in many cancers (Xia et al., 2013), (b) the faciogenital dysplasia 1 protein (FGD1), an activator of CDC42, a member of the Ras-like family of Rho- and Rac proteins involved in Aarskog-Scott syndrome and predominantly detected at microtubules (Lebel et al., 2002), (c) the ribosomal protein S27A (RPS27A), a cytoplasmic fusion protein of ubiquitin and protein S27A (Kirschner and Stratakis, 2000) and (d) the nuclear ubiquitin-like with PHD and ring finger domains 1 (UHRF1) protein, a member of a subfamily of RING-finger type E3 ubiquitin ligases, which was reported to be overexpressed in diverse forms of human cancer (Sabatino et al., 2012). The group of proteins involved in histone modification (II) predominantly localize to the nucleus and consisted of (a) cyclin-dependent kinase 2-associated protein 1 (CDK2AP1), a protein forming a core subunit of the nucleosome remodeling and histone deacetylation (NuRD) complex reported to be deleted in oral cancers (Peng et al., 2006), (b) the recombination signal binding protein for immunoglobulin kappa J region (RBPJ), a transcriptional regulator that is involved in Adams-Oliver syndrome 3 and recruits chromatin remodeling complexes containing histone deacetylase or histone acetylase proteins to Notch signaling pathway genes (Hassed et al., 2012), (c) the RuvB-like protein 2 (RUVBL2), a DNA helicase and component of the NuA4 histone acetyltransferase complex reported to be overexpressed in some tumors (Rousseau et al., 2007) and (d) the SET domain containing protein 1B (SETD1B), a histone methyltransferase that specifically methylates 'Lys-4' of histone H3 (Lee and Skalnik, 2012). Members of the last group of proteins used for the generation of BAC-transgenic lt-NES cells (III) comprise (a) the nuclear JUN proto-oncogene (JUN), a transcription factor binding to the enhancer heptamer motif 5'-TGA[CG]TCA-

3', that was reported to induce cell transformation (Hartl et al., 2003), (b) the microtubule-associated protein 1 light chain 3 beta (MAP1LC3B), a protein involved in microtubule assembly that has been reported to be involved in autophagy-related mechanisms (Mikhaylova et al., 2012), (c) the nuclear methyl CpG binding protein 2 (MECP2), a protein binding to methylated DNA that is associated with for the X-linked mental retardation of the Rett syndrome (Amir et al., 1999) and (d) the proliferating cell nuclear antigen (PCNA), a cofactor of DNA polymerase delta often used as diagnostic marker for tumors (Cappello et al., 2006). For assessment of naive fluorescence of GFP-fusion proteins and an evaluation of the percentage of cells expressing it, flow cytometric analysis (FACS) was performed and revealed robust and homogenous expression of GFP-tagged candidate proteins in the majority of pools (Figs. 4.3, 4.4 and 4.5; FACS (b; e; h; k)). In order to validate the expression of fully intact fusion proteins, expected protein sizes were correlated with signals observed in Western immunoblots. To that end, protein sizes were calculated for the GFP tag (C-terminal 29, 9 kDa, N-terminal 35 kDa), added to the protein weights (Table 4.1) and compared to sizes detected by Western immunoblotting using an anti-GFP antibody (Fig. 4.3, 4.4 and 4.5; Western Blot (c; f; i; l)). Proteins of the expected size could be detected in most cases (AURKA, CDK2AP1, JUN, MECP2, PCNA, RBPJ, RPS27A, RUVBL2, UHRF1) but were close to or beneath the detection threshold for SETD1B, MAP1LC3B and FGD1.

Table 4.1: Calculated sizes of GFP fusion proteins and tag position

	Protein size (kDa)	Tag size (kDa)	Fusion protein (kDa)	Tag	Western Blot
AURKA	45,8	29,9	75,7	C-terminal	double band
CDK2AP1	12,4	29,9	42,3	C-terminal	positive
FGD1	106,6	29,9	136,5	C-terminal	weak signal
JUN	35,7	29,9	65,6	C-terminal	weak signal
MAP1LC3B	14,7	35,0	49,7	N-terminal	weak signal
MECP2	52,4	29,9	82,3	C-terminal	positive
PCNA	28,8	29,9	58,7	C-terminal	positive
RBPJ	55,6	29,9	85,5	C-terminal	positive
RPS27A	18,0	29,9	47,9	C-terminal	positive
RUVBL2	51,2	29,9	81,1	C-terminal	positive
SETD1B	208,7	29,9	238,6	C-terminal	no signal
UHRF1	89,8	29,9	119,7	C-terminal	positive

The size of proteins (protein size) present on employed BAC constructs is shown for their predominant isoform. Dependent on the localization of the GFP tag to the N- or the C-terminus, the linking element has a different size (Tag size), resulting in the calculated size of the fusion protein. Ratings of signal intensities for protein analyses (Western Blot) are stated.

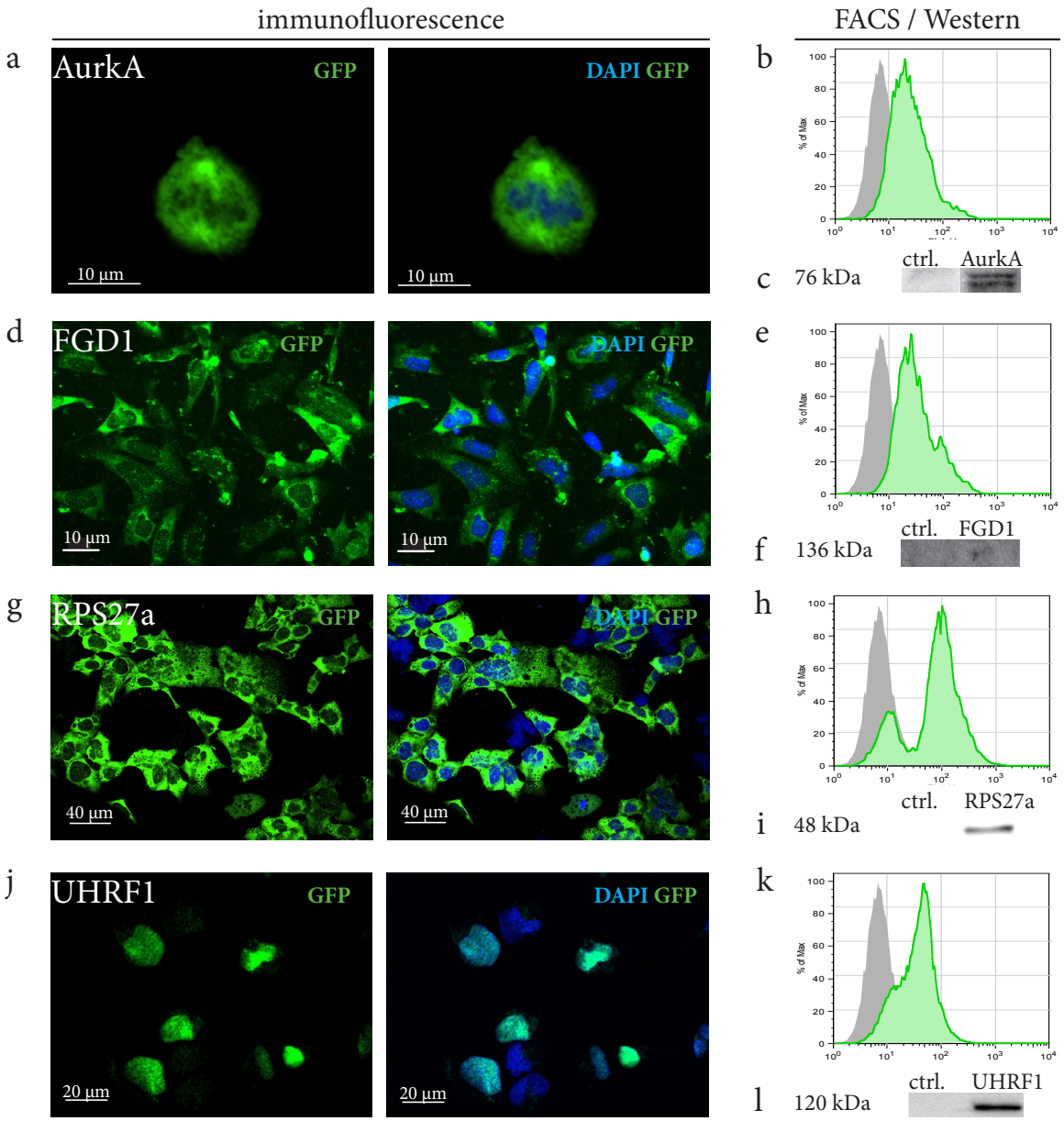


Figure 4.3: Characterization of BAC transgenic cell lines involved in protein modification or signaling, group (I)

Confocal immunofluorescence microscopy: *BAC transgenic cell lines (AURKA.GFP (**a**), FGD1.GFP (**d**), RPS27A.GFP (**g**) and UHRF1.GFP (**j**)) were stained with an anti-GFP antibody and localization of fluorescent signals was analyzed. Confocal images were reconstructed from a z-stack of ≥8 1 µm optical sections.* ***FACS:*** *The percentage of cells expressing naïve-fluorescent GFP-fusion proteins was assessed using flow cytometry (**b**, **e**, **h** and **k**).* ***Western:*** *The size of fusion proteins was analyzed by Western immunoblotting (**c**, **f**, **i** and **l**).*

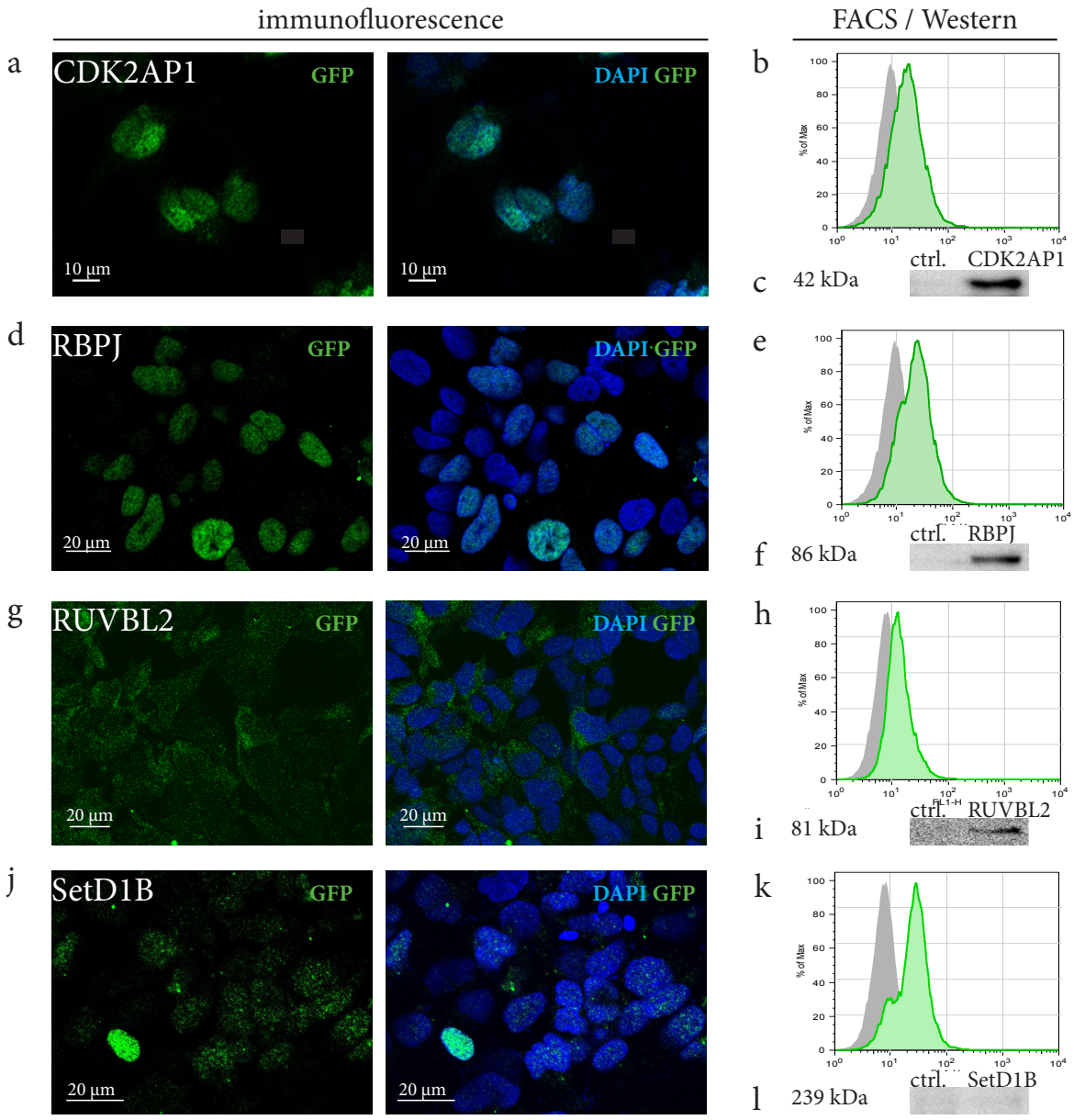


Figure 4.4: Analysis of cells harboring GFP-tagged proteins involved in histone modification, group (II)

Confocal immunofluorescence microscopy: *Lt-NES cell lines harboring CDK2AP1.GFP (**a**), RBPJ.GFP (**d**), RUVBL2.GFP (**g**) and SETD1B.GFP (**j**) were analyzed for the localization of fluorescent signals using an anti-GFP antibody. Confocal images were reconstructed from a z-stack of ≥8 1 µm optical sections.* ***FACS:*** *The fraction of cells expressing naïve-fluorescent GFP-fusion proteins was evaluated using flow cytometric analysis (**b**, **e**, **h**, **k**).* ***Western:*** *Western immunoblotting was used for analysis of fusion protein size (**c**, **f**, **I** and **l**).*

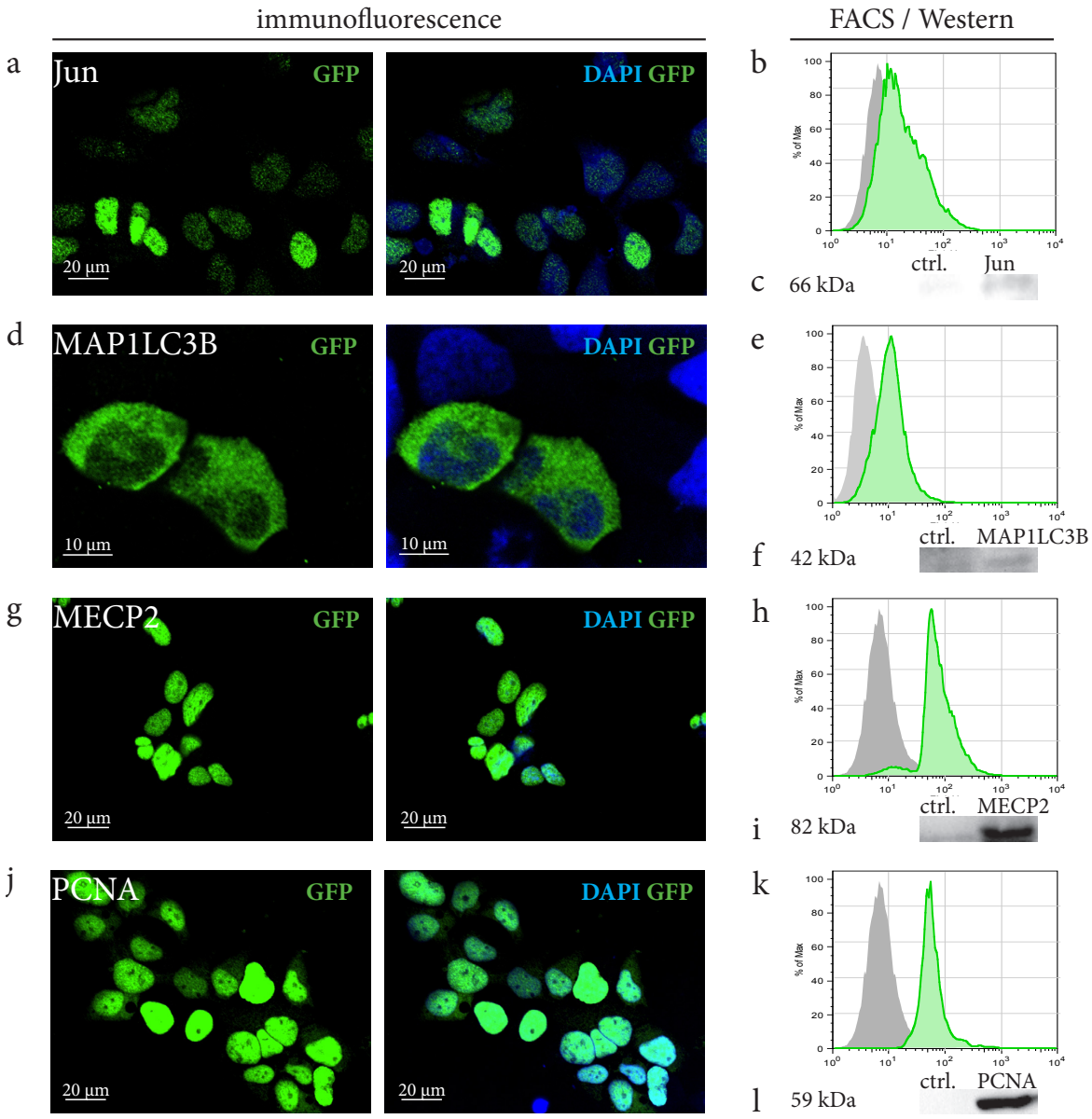


Figure 4.5: Characterization of BAC transgenic cell lines, group III
Confocal immunofluorescence microscopy: *Lt-NES cell lines transgenic for BACs encoding JUN.GFP (**a**), MAP1LC3B.GFP (**d**), MECP2.GFP (**g**) and PCNA.GFP (**j**) were analyzed by confocal micoscropy using an anti-GFP antibody. Confocal images were reconstructed from a z-stack of ≥8 1 µm optical sections.* ***FACS:*** *Flow cytometric analysis was used for evaluation of percentage of cells expressing a naïve-fluorescent GFP-fusion protein (**b**, **e**, **h**, **k**).* ***Western:*** *The size of GFP-fusion proteins was analyzed by Western immunoblotting (**c**, **f**, **I** and **l**).*

Most importantly, all transgenic cell lines were subjected to confocal fluorescence microscopic analysis, and recruitment of GFP-tagged proteins into expected subcellular compartments (uniprot.org) could be confirmed for almost all proteins (*immunofluorescence*, Fig. 4.3, 4.4 and 4.5). The only exception was RUVBL2, reported to mainly localize to the nucleus (Sigala et al., 2005). However, a scattered non-nuclear signal within the corresponding lt-NES cell lines was observed (Fig. 4.4 g).

4.1.4. BAC transgenesis results in cell lines with low copy number integrations

To assess the number of integration sites, a set of BAC transgenic lt-NES cell lines including Aurora kinase A (AURKA Fig. 4.6 a), the Jun proto-oncogene (JUN, Fig. 4.6 b) and proliferation-associated cellular nuclear antigen (PCNA, Fig. 4.6 c) was subjected to fluorescence in situ hybridization (FISH). To that end, chromosome painting probes suitable for the endogenous locus of AURKA (chromosome 20), JUN (chromosome 1) and PCNA (chromosome 20) were applied together with corresponding FISH probes that were generated using the respective BAC (conducted by Dr. Irina Solovei). For quantification, metaphase spreads comprising a diploid set of the respective endogenous chromosome were analyzed. Unaltered healthy cell populations exhibit two sites in their genome, corresponding to their alleles, while e.g. a duplication of an allele or the random integration of a BAC construct increases the number of observed signals. Here, an arithmetically averaged signal of BAC probes in AURKA.GFP (n=3) and PCNA.GFP cells (n=15) of 3,0 ± 0,0 loci per cell was observed, while 3,18 ± 0,4 sites in JUN.GFP lt-NES cells were detected (Fig. 4.6 d). Previous studies investigating BAC integration frequencies reported a repetitive integration of multiple BAC copies per locus (Sparwasser and Eberl, 2007). While FISH analysis predominantly is used for detection of translocations or amplifications of large chromosomal fragments, approaches for the quantitative assessment of FISH signals are established (Iourov et al., 2005). Hereby, the signal intensity of known sites can be correlated with that of unknown sites. Based on this ratio, the size of the unknown section can be estimated. To evaluate the number of copies per site, the signal intensity of each site was compared to averaged signal intensities of endogenous sites and resulted in similar signal intensities for AURKA.GFP (n=9) and Jun.GFP (n=19) (≥ 75% of endogenous sites, about equivalent to a single copy integration) while PCNA.GFP (n=16) cells had increased signal intensities (> 187% of endogenous sites, corresponding to double or multiple copy integrations), indicating integrations of several copies of the construct per locus (Fig. 4.6 e). Notably, in every analyzed line a fraction of cells showed decreased signal strengths of the newly introduced loci, possibly originating from an incompletely integrated BAC fragment (Fig. 4.6 e).

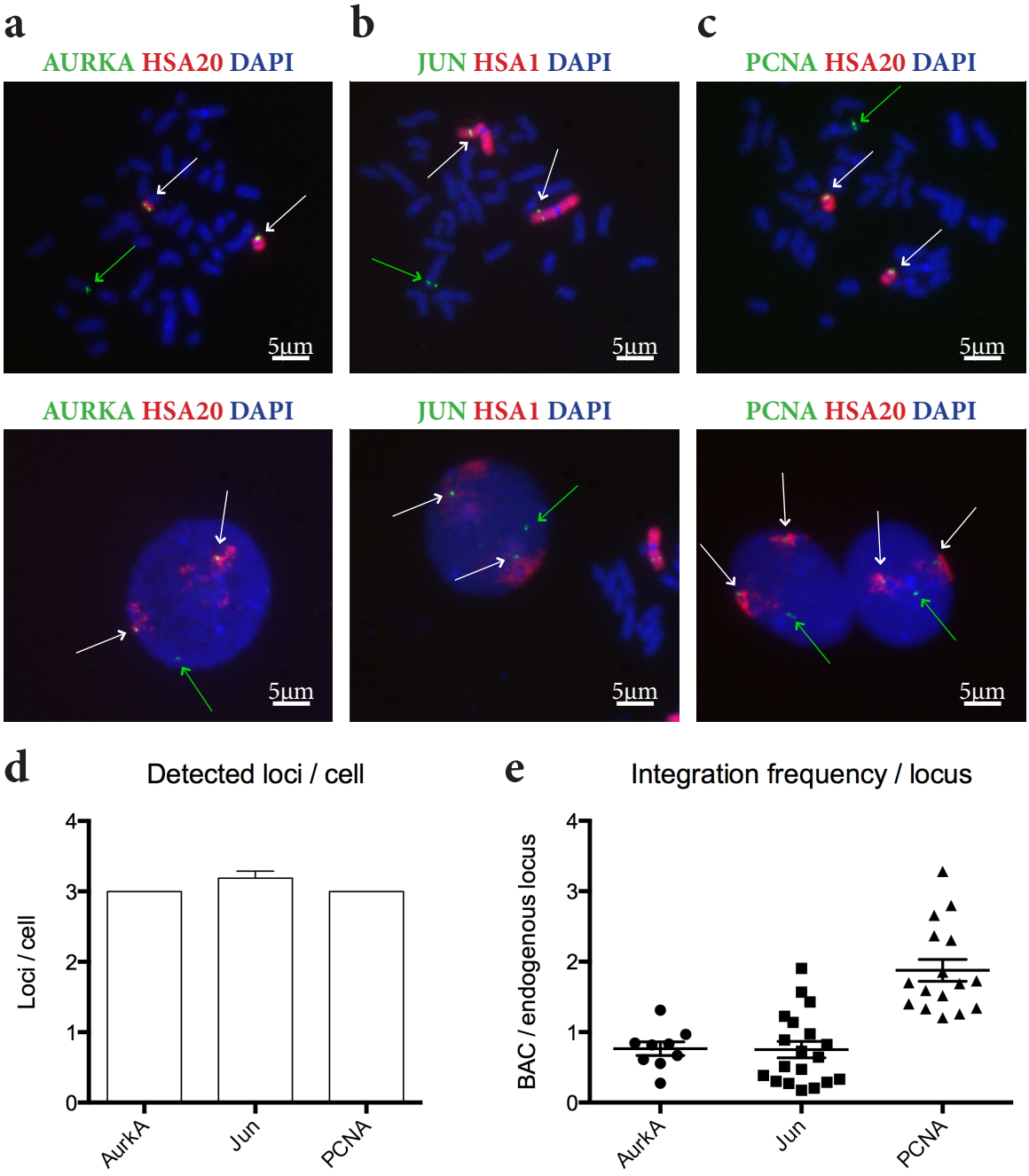


Figure 4.6: Characterization of BAC transgenic cell lines
Fluorescence in situ hybridization (FISH) analysis using BAC-specific probes was applied to determine the number of genomic sites with high analogy after stable integration of the BAC transgene. ***a:*** The AURKA-targeting probe detects two signals (green) on chromosome 20 (HSA20) harboring the endogenous loci of the AURKA gene as well as one further locus on another chromosome *(upper panel metaphase spread, lower panel prophase analysis).* ***b:*** *The endogenous copies of JUN are located on chromosome 1 (HSA1), detectable by the flourescent probe (green) with one additional signal on another chromosome.* ***c:*** *The endogenous sites of PCNA, located on chromosome 20 (HSA20), as well as one additional locus on a different chromosome can be detected by the probe (green).* ***d:*** *The column diagram shows the number of detected loci per cell for 3 different lines: AURKA.GFP (3,0 ± 0,0; n=3), JUN.GFP (3,18 ± 0,4; n=16) and PCNA.GFP (3,0 ± 0,0; n=15).* ***e:*** *Fluorescence signal intensity analysis for additional BAC copies using ImageJ. BAC-transgenic lt-NES lines AURKA.GFP and JUN.GFP show an average signal intensity corresponding to one copy/locus (0,76 ± 0,28 respectively 0,75 ± 0,51) while PCNA.GFP signals correspond to two copies/locus (1,87 ± 0,62). Graphs show mean ± SD. FISH images were kindly generated by Dr. Irina Solovei.*

4.1.5. Proteins expressed by integrated BACs are localized and regulated in response to cell cycle and differentiation

In previous reports, the influence of protein tags and their positional impact by fusion to N- or C-terminal sequences was evaluated (Poser et al., 2008; Riesen et al., 2002). Hereby, functional examinations of tagged proteins were conducted and confirmed in most cases that correctly localized proteins were able to fulfill their physiological functions.

In correspondence to these studies, two GFP-tagged candidate proteins were chosen and the corresponding lt-NES cell lines investigated. One candidate used was the tagged Aurora kinase A (AURKA.GFP). This protein has previously been reported to localize to mitotic spindles and centrosomes during mitosis while being evenly distributed throughout the cell otherwise (Ding et al., 2011). In AURKA.GFP transgenic lt-NES pools, the GFP signal during the long-lasting (30 h) interphases was weak and evenly distributed. During metaphase, an increasing intensity could be observed, presumably localizing to centrosomes and mitotic spindles, dissolving after completion of mitosis (6 h, Fig. 4.7 a). The analysis of the protein size via Western immunoblotting failed due to insufficient protein levels present in proliferating AURKA.GFP lt-NES cell lines. In order to enrich cells in mitosis, the spindle degradation was blocked by application of Taxol, a compound that binds to ß-tubulin and thereby inhibits the degradation of the microtubules involved in mitosis (24 h, 0,3 and 1 µM, Fig. 4.8 b, c). Flow cytometry analysis was used to examine the proportion of cells with enriched AURKA.GFP expression compared to controls and revealed that a concentration of 1 µM Taxol was sufficient for an increased AURKA.GFP expression in comparison to untreated AURKA.GFP cells (Fig. 4.7 b). The upregulation of the AURKA.GFP protein level in Taxol-treated cells was also reflected in Western immunoblotting analysis and revealed proteins of two distinct sizes, consistently with previous observations (Fig. 4.8 c, Seki et al., 2008).

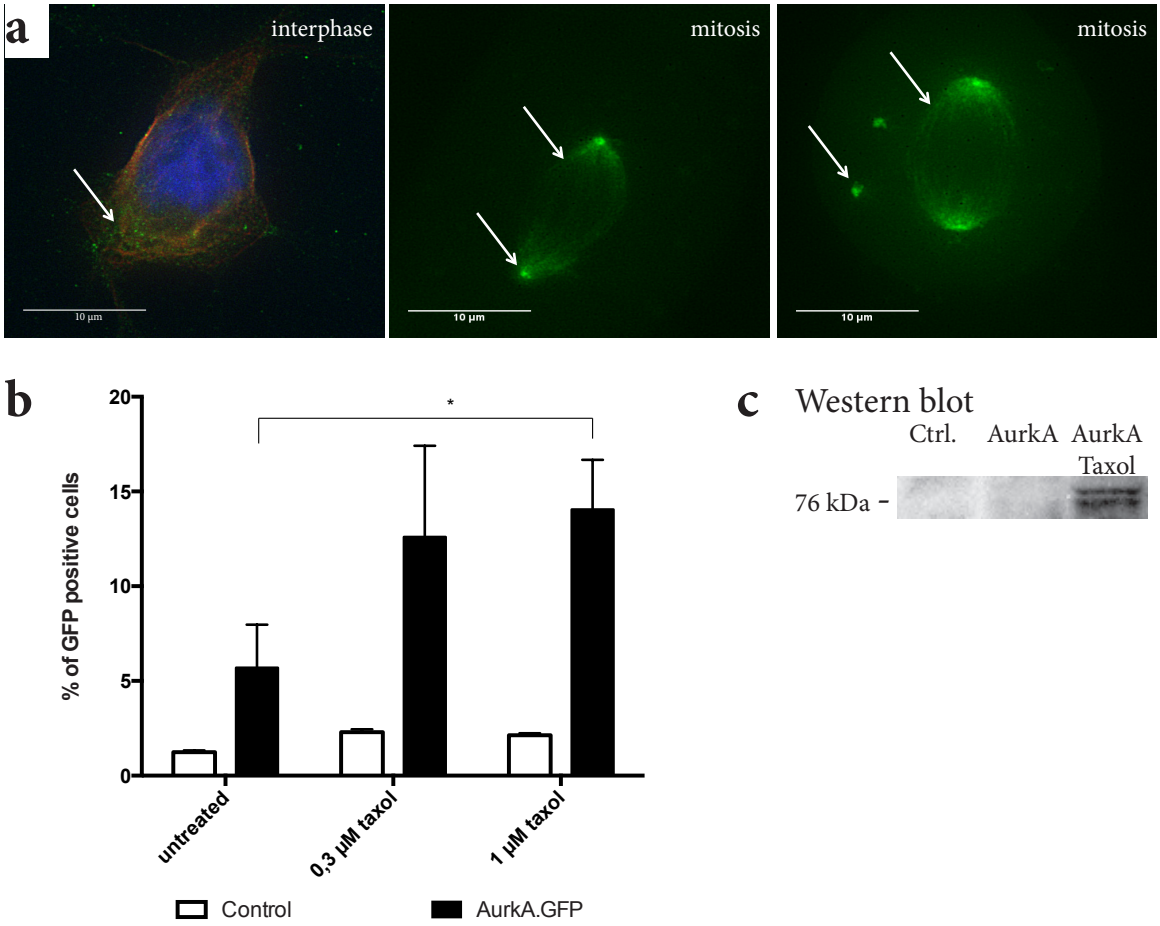


Figure 4.7: Analysis of cell cycle dependent expression of AURKA.GFP
***a**: Aurora kinase A (AURKA) GFP lt-NES cells show scattered signals for GFP during interphase (DAPI/actin/AURKA.GFP) and concentration to subcellular structures during mitosis, presumably representing centrosomes and spindle microtubules (arrows). Scale bars = 10 µm. Images were recorded in cooperation with Nicolas Berger, MPI-CBG Dresden. **b**: Following taxol treatment for 24 h, (0,3 and 1 µM), flow cytometry analysis demonstrated a significant enrichment of cells with increased AURKA.GFP expression (1 µM). Graphs show mean ± SD. **c**: Western blot analysis for AURKA.GFP lt-NES cells under standard proliferation conditions resulted in no detectable signal for the calculated size of the fusion protein of AURKA and GFP (76 kDa) while AURKA.GFP was detectable after taxol treatment (1 µM, 24 h). Two different sizes of AURKA have been described previously (Seki et al., 2008).*

To address the question whether BAC-mediated expression of proteins is regulated analogously to their endogenous equivalents, a protein reported to be exclusively expressed in proliferating cells, the proliferating cell nuclear antigen (PCNA), was chosen for further analysis (Dietrich, 1993). PCNA.GFP transgenic lt-NES cells were differentiated under neuronal induction conditions that give rise to increasing numbers of postmitotic neurons in the course of differentiation (Koch et al., 2009). In order to observe this process, cells were analyzed before, 7, 14 and 21 days after induction of differentiation. With progressing differentiation and concomitant emergence of post-mitotic neurons (MAP2ab$^+$), the proportion of PCNA.GFP positive cells declined from more than 90% under proliferative conditions to less than 40%

three weeks later (Fig. 4.8 a, b). These immunofluorescence-based observations were supported by Western immunoblotting analysis, since increasing amounts of β3-tubulin, an early cytoskeletal neuron-specific marker, correlated with decreasing PCNA.GFP levels (Fig. 4.8 d).

Although minor differences that could be caused by the GFP tag in binding affinities or protein half-life might exist, both exemplary investigations recapitulate physiological alterations in expression levels and subcellular localization.

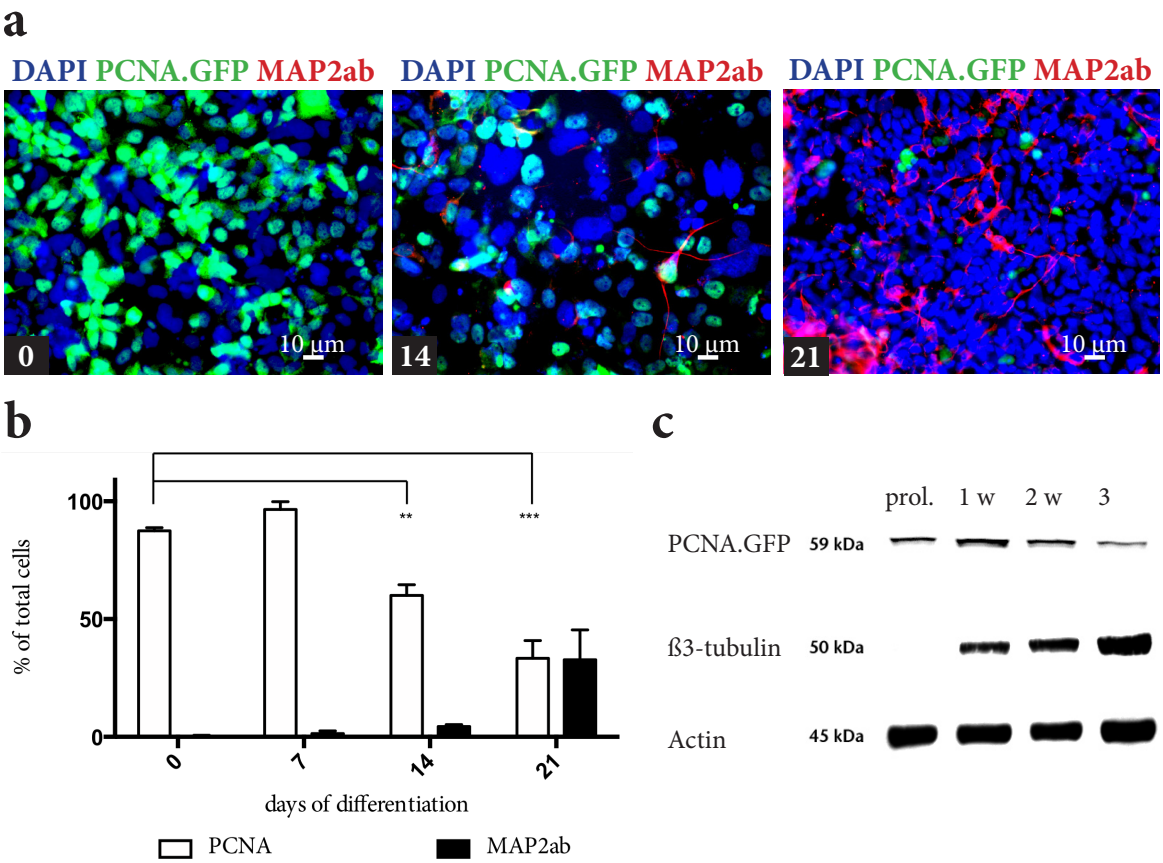


Figure 4.8: Analysis of PCNA.GFP regulation during differentiation

***a**: A time course (0, 7, 14 and 21 days after initiation of differentiation) of proliferating cell nuclear antigen (PCNA) GFP cells reveals an increase of mature neurons (MAP2ab; red), and a decrease of PCNA.GFP (green) cells over time. Scale bars = 10 µm. **b:** A diagram shows the quantity of PCNA.GFP and MAP2ab positive cells during a 3-week period. Graphs shown mean ± SD; n=400. **c:** Western immunoblotting analysis for PCNA.GFP and β3-Tubulin shows a decline of PCNA.GFP and an increase of early neuronal marker β3-Tubulin over time. Actin was analyzed as loading control.*

4.1.6. Endogenously expressed GFP-tagged proteins can be used to visualize dynamics of protein trafficking in live cell imaging

For live cell imaging, the AURKA.GFP expressing lt-NES cell line was chosen based on its GFP-fusion protein fluorescence intensity and specific GFP localization pattern. AURKA is a serine/threonine kinase involved in the regulation of centrosome maturation, spindle assembly, centrosome separation and mitotic checkpoint control (Ding et al., 2011). In order to discriminate individual cells within lt-NES cell populations naturally growing at high densities, a fraction of the cells was labeled using a red fluorescent protein (RFP) with a nuclear localization signal. During live cell imaging, the recruitment of AURKA.GFP to mitotic spindles and centrosomes during mitosis could be visualized (Figure 4.9).

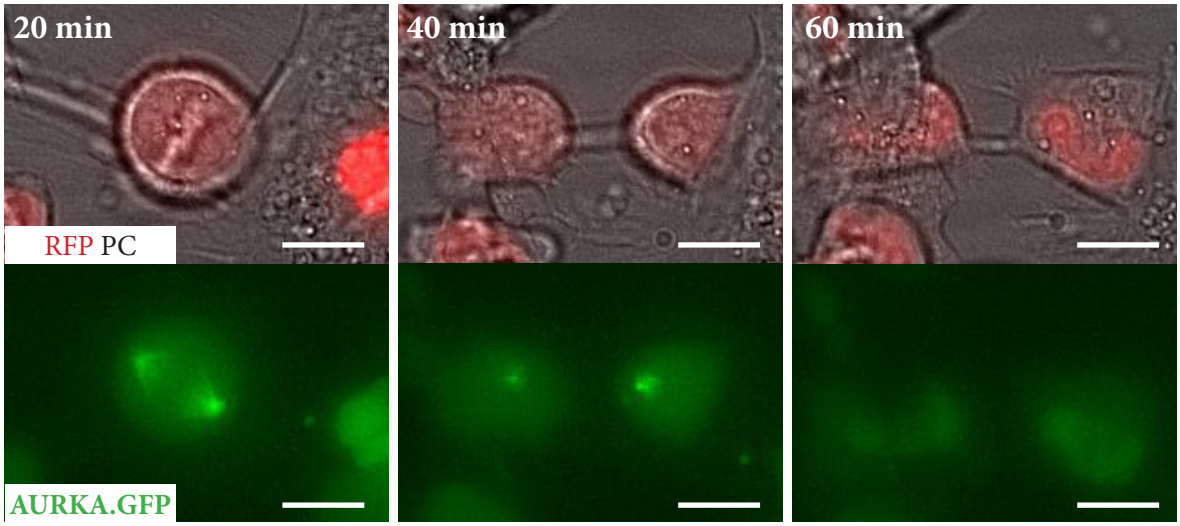


Figure 4.9: Live cell imaging of BAC transgenic lt-NES cell lines
Life cell imaging of AURKA.GFP lt-NES cells under proliferative conditions. The displayed cell is imaged during its mitotic phase, showing AURKA.GFP recruitment to mitotic spindles and centrosomes during mitosis (20 min) and its dissociation following cytokinesis (40 and 60 min). Live cell microscopy was performed in cooperation with Nicolas Berger, MPI-CBG Dresden. Scale bars = 10 µm.

This data demonstrates that BAC-mediated expression of a GFP-tagged protein is sufficient for live-cell visualization of subcellular localization.

4.1.7. Interactors of GFP-tagged proteins expressed by integrated BACs can be detected by mass spectrometry

Protein-protein interaction analyses can be conducted in multiple systems by using a variety of techniques. For the discovery of novel interactors possibly associated with multiple complexes, co-immunoprecipitation-based methods are employed predominantly. Traditionally, quantitative proteomic approaches rely on stable isotope labeling by amino acids in cell culture (SILAC) with a subsequent high-resolution liquid chromatography (LC) tandem MS (LC-MS/MS). Here, the unaltered cell line is cultured in medium containing the $C^{12}N^{14}$ light form of lysine while the tagged cell line is cultured in a medium containing the $C^{13}N^{15}$ heavy form of lysine. Subsequently, separate pull-downs are performed, eluates merged, peptides identified by LC-MS/MS and quantified by comparison of relative intensities of the light and heavy forms of each peptide present in the mass spectrum (Hubner et al., 2010a). In this work, label-free quantitative interaction proteomics by LC-MS/MS was used in conjunction BAC transgenesis (Fig. 4.10 a, Vermeulen et al., 2008). This approach avoids the application of heavy amino acids by employing a label-free technique where eluates from pull-downs of tagged and control cells were not combined but analyzed separately by LC-MS/MS with a subsequent software-based quantification of proteins (MaxQuant, label-free algorithm, Cox and Mann, 2008).

For the analysis of interactors of tagged proteins within lt-NES cells, transgene-free and transgenic lt-NES cells ($1{,}5 \times 10^8$) were produced. Cell pellets were frozen (-80°C) as technical triplicates, stored and processed in large batches to avoid variations caused by measurements at different days. Quantitative BAC-green fluorescent protein interactomics (QUBIC) was performed by Marco Hein, MPI Martinsried as described previously (Fig. 4.4 a Hubner et al., 2010a). Data was obtained either as statistically processed volcano plots (Fig. 4.10 b) or as pre-filtered and normalized raw data (Fig. 4.10 c). Further analysis of raw data was performed by using hierarchical clustering with Euclidean distance in the Perseus software to exclude concealment of significant interactors by statistical filters used for the generation of the volcano plots (Fig. 4.10 c). Previous studies successfully applied QUBIC for the analysis of protein-protein interactions and complex composition in human cell lines (Hubner et al., 2010a; Smits et al., 2013; Spruijt et al., 2010). Taken

together, this approach provides a both sensitive and selective technique suitable to identify protein interactors with good confidence.

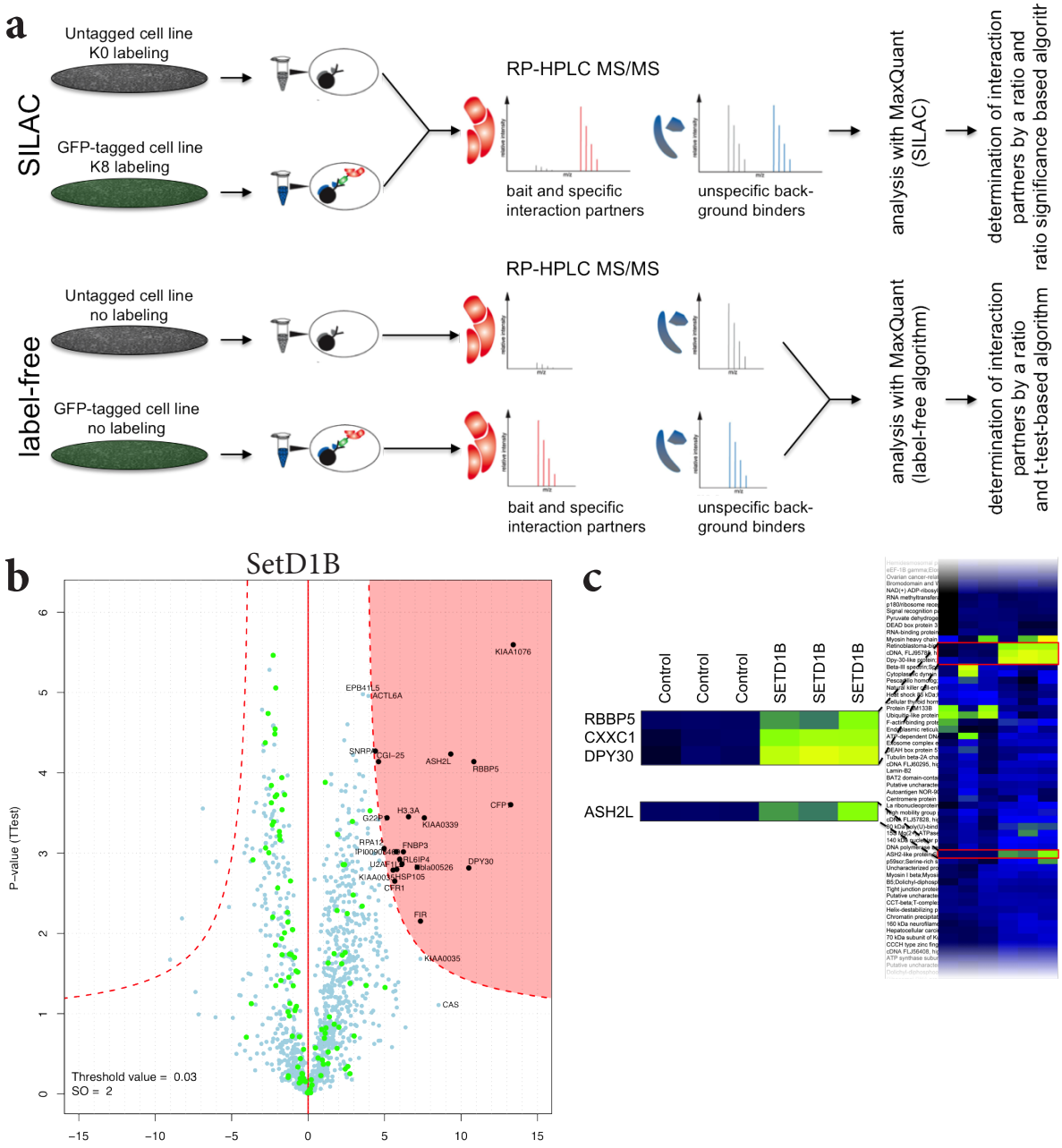


Figure 4.10: Detection of enriched interactors by mass spectrometry
***a**: Schematic procedure for stable isotope labeling by amino acids in cell culture (SILAC) opposing to label-free quantitative proteomics performed with LC-MS/MS. In SILAC-based quantitative proteomics, separate pulldowns are performed and eluates mixed directly after elution by in-column digestion whereas eluates in label-free approaches are analyzed separately by LC-MS/MS. SILAC probes are quantified by direct comparison of the relative intensities of the light and heavy forms of peptides present in the mass spectrum. Proteins in the label-free approach are quantified using the label-free algorithm in MaxQuant. Scheme adapted from Hubner et al., 2010. **b**: In volcano plots, the logarithmic ratio of protein intensities is plotted against the negative logarithm of the p-values of the t-test performed from triplicates. A hyperbolic curve separates significantly enriched proteins (with a p-value of ratio significance <0.03) marked in black (red dotted line) from background (blue dots; green dots are linked to ribosomal proteins). **c**: Heat plot view from technical triplicates of control and SETD1B.GFP transgenic lt-NES cells generated with the software Perseus. Known interactors of SETD1B that are significantly enriched are magnified.*

4.1.8. Comparison of generated protein-protein interaction (PPI) data with bioinformatics databases comprising known and predicted PPIs demonstrates new and established interactors

Analysis of protein-protein interaction data has to be performed very carefully, especially when accessing latest methods in measurement and data processing. All proteins marked as significantly enriched were counter-checked with public protein or gene databases (genecards.org, uniprot.org, string-db.org or proteinatlas.org; access 11/2013) and classified corresponding to findings previously reported. Principally, every generated BAC-transgenic lt-NES cell line with markedly detectable signals in immunoblots for the introduced tagged protein was used for quantitative BAC-green fluorescent protein interactomics (QUBIC) based analysis of interacting proteins. However, it was often difficult to interpret these data when few proteins that were expected to be interactors were present among significantly enriched proteins. This chapter is focused on the detailed description of PPI data generated from GFP tagged proteins where proteins usually reported to interact were significantly enriched in coprecipitates of the respective transgenic cell lines. For analysis of proteomic mapping of interactors, four lt-NES cell lines expressing PCNA.GFP, CDK2AP1.GFP, SETD1B.GFP and RUVBL2.GFP were chosen.

PCNA, a gene highly expressed in proliferating cells, is involved in the control of eukaryotic DNA replication by increasing the processibility of DNA polymerase delta during elongation of the leading strand. In QUBIC analysis, several significantly enriched proteins were detected (Fig. 4.11 a). Amongst them, members of the replication factor C (RFC1, RFC2 and RFC4), a heteropentameric AAA+ protein clamp loader of PCNA, were enriched (Bowman et al., 2004b; Yao et al., 2003). Another interactor previously also detected in a yeast two-hybrid screen was p15/PAF1, which has been reported to compete with p21 for binding to PCNA (Chen and Scheller, 2001b). The enriched protein DNA (Cytosine-5)-methyltransferase 1 (DNMT1) is another known interactor of PCNA that maintains the methylation pattern in the newly synthesized strand and is essential for epigenetic inheritance (Hervouet et al., 2010; Viré et al., 2006). Several proteins with physiological functions that can be associated with auxiliary tasks of PCNA (23) were also enriched. Amongst them are proteins that may be indirectly linked like the cell division cycle and apoptosis regulator 1 (CCAR1) via non-detected cyclin-dependent kinase inhibitor 1A (p21/CDKN1A), or large fractions of the p15/PAF1-linked cleavage and

polyadenylation specific factor complex (CPSF1, CPSF2, CPSF4 Nagaike et al., 2011) and the herewith interconnected CSTF2 (Fig. 4.11 a).

Another bait that allowed the copurification of the majority of the members from its complex, the Mi-2/nucleosome remodeling deacetylase (NuRD) complex, was cyclin-dependent kinase 2 associated protein 1 (CDK2AP1) GFP fusion protein. Elements of this complex were previously described as interacting elements of CDK2AP1 using BAC transgenesis together with SILAC based quantitative proteomics in HeLa cells (Spruijt et al. (2010). This complex is functionally associated with transcriptional repression and consists of the chromodomain helicase DNA binding protein 4 (CHD4/Mi2), the histone deacetylase 2 (HDAC2), the metastasis-associated proteins 1, 2 and 3 (MTA1, MTA2, MTA3), the GATA zinc finger domain containing 2B (GATAD2B), the Methyl-CpG binding domain protein 3 (MBD3) and the retinoblastoma-binding protein P46 and P48 (RBAP46, RBAP48) and dissociates in active open chromatin loci (Denslow and Wade (2007a). A comparison of interactors previously reported with the set of proteins detected in this work revealed, that its components all co-purify with the tagged variant of CDK2AP1.GFP (Fig. 4.11 b). Furthermore, four additional proteins could be observed: H3.3, a histone variant found in nucleosomes of active genes (Tagami et al. (2004a), BCAR1, a kinase involved in cell migration (Modzelewska et al., 2006), g22p1, a single stranded DNA-dependent ATP-dependent helicase (Reeves and Sthoeger, 1989a) and ZNF219, a transcriptional repressor (Sakai et al., 2003a).

Using SETD1B.GFP as bait, the Set1-like multiprotein methyltransferase complex (COMPASS complex) could be detected with its components ASH2L, RBBP5, SetD1A, DPY30 and CXXC1 (Fig. 4.8 c). In this context, histone H3.3 (H3.3A) was co-purified, which is a target of the WD repeat-containing protein 5 (WDR5) involved in positioning the N-terminus of histone H3 for efficient trimethylation by the human Set1A-B complex (Garapaty et al., 2009). In addition, a large set of candidates with functions that are yet unknown in connection with this complex was enriched: KIAA0035, Nbla00526, RL6IP4, EPB41L5, ACTL6A, SNRPA, Cgi-25, G22P, RPA12, IPI00908469, RPA12, FNBP3, HSP105, FIR and CFR1.

When analyzing a GFP-tagged variant of the Holliday junction ATP-dependent DNA helicase RuvB-like 2 (RUVBL2.GFP) by QUBIC, members of several complexes interacting with RUVBL2 were enriched. A total of 23 proteins were detected with 14 interactors that were already known to interact with RUVBL2. RUVBL2 acts as a

subunit or catalytic unit in the multisubunit ATP-dependent chromatin remodeling complex INO80, the nutrient sensing complex URI/Prefoldin, the chromatin remodeling complex SRCAP and the NuA4 histone acetyltransferase complex (Conaway and Conaway, 2009a; Doyon et al., 2004b; Sardiu et al., 2008a; Wong et al., 2007). In QUBIC measurements, representatives from all complexes could be detected: From the INO80 complex, the members INO80H/RuvB-like 1, INO80C/C18orf37, INO80N/ACTR8, INO80A/INO80, ACTR5 and INO80K/ACTL6A/BAF53 were detected. As members of the SRCAP complex, RUVBL1, INO80K/ACTL6A/BAF53 and DMAP1 were found. Copurified members of the NuA4 complex were RUVBL1, EP400/CAGH32, TRRAP/PAF400, DMAP1, EPC2, SRCAP/KIAA0309 and INO80K/ACTL6A/BAF53. Of the URI complex, PFDN2, URI and RuvBl1 were enriched. The generated data confirmed a previous report that described the association of PDRG1 with RUVBL2 but could not assign a biological role to their interaction (Jeronimo et al., 2007a). Furthermore, 9 novel candidate interactors RPN1, RPN2, NEF3, HZGJ, FNDP3, CAS, CKAP4, G22P1 and FIP1 were enriched significantly.

All detected interactors of the three proteins involved in gene regulation – SETD1B as a histone methyltransferase, CDK2AP1 as member of the gene regulating NuRD complex (Denslow and Wade, 2007a) and RUVBL2 as member of the Nu4A histone acetyltransferase complex – were clustered using Cytoscape (Fig. 4.12). The only common interactor of all three proteins was XRCC6, a single stranded DNA-dependent ATP dependent helicase. RUVBL2 and SETD1B are linked via CSE1L and CSTF3, while CDK2AP1 and SETD1B are linked via SPATS2, CORO1C and H3F3A, and RUVBL2 and SETD1B are linked via PRPF40A.

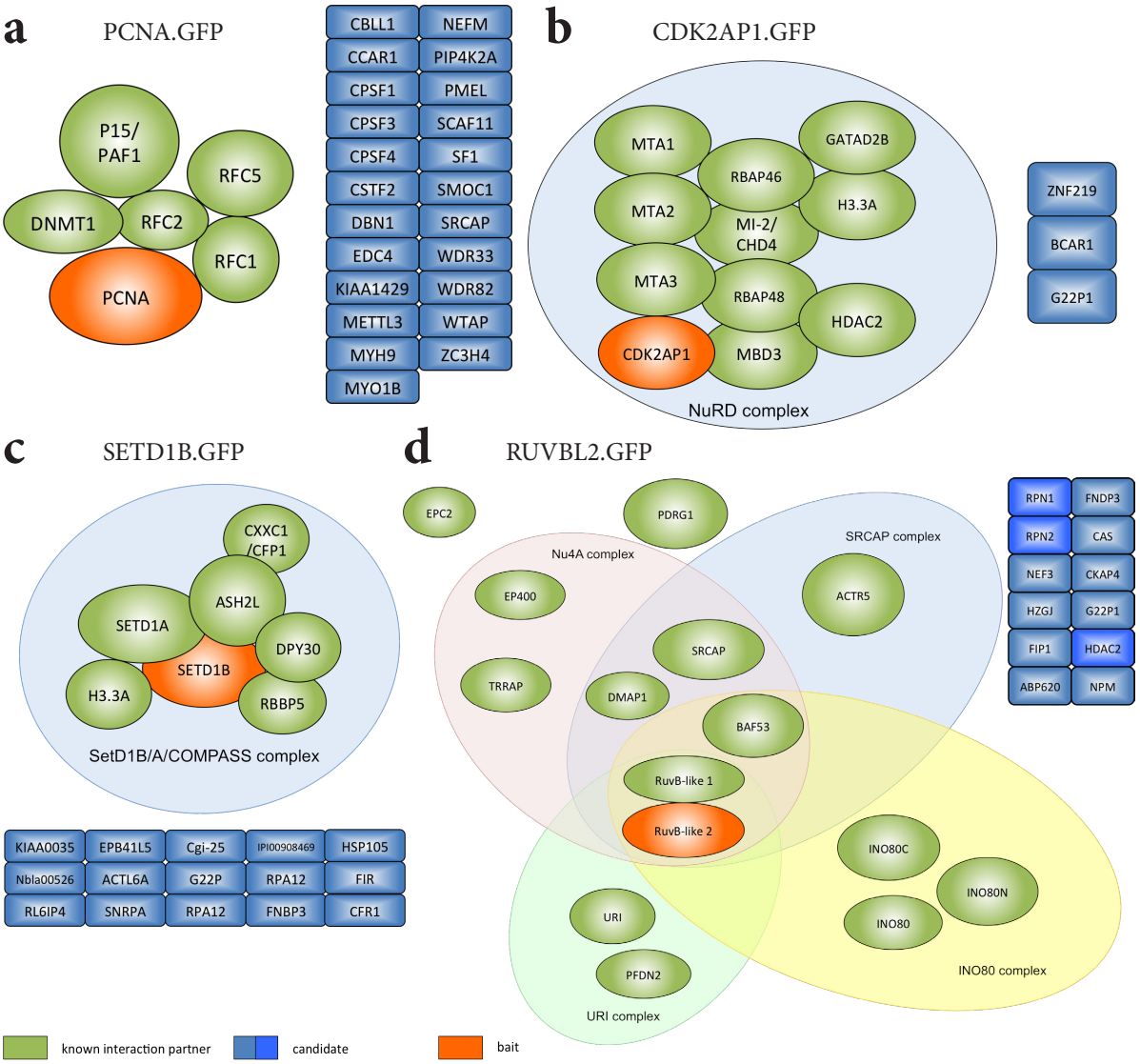


Figure 4.11: Analysis of mass spectrometry data

*Proteins shown in these illustrations were all detected as significantly enriched. **a**: PCNA.GFP used as bait copurified together with established interaction partners DNMT1, P15, RFC1, RFC2, RFC5 and 23 novel candidates. **b**: CDK2AP1.GFP as part of the NuRD complex copurified together with the majority of its members including MI-2, HDAC2, RBAP46, RBAP48 MBD3, GATAD2B, MTA1, MTA2, MTA3 and the three candidates H3.3A, BCAR1 and G22P1. **c**: SETD1B as part of the COMPASS complex copurified with its members H3.3A, SETD1A, ASH2L, CXXC1, DPY30 and RBBP5 as well as 15 candidate interactors. **d**: RUVBL2 is a member of multiple complexes, resulting in enrichment of representatives from all of them: The NuA4-, the SRCAP-, the INO80- and URI complex together encompass EP400, TRRAP, DMAP1, SRCAP BAF53, ACTR5, INO80C, INO80N, INO80A, URI, PFDN2 and RUVBL1. Further proteins copurified with RUVBL2 were the previously reported interactors EPC2 and PDRG1 as well as 9 candidates.*

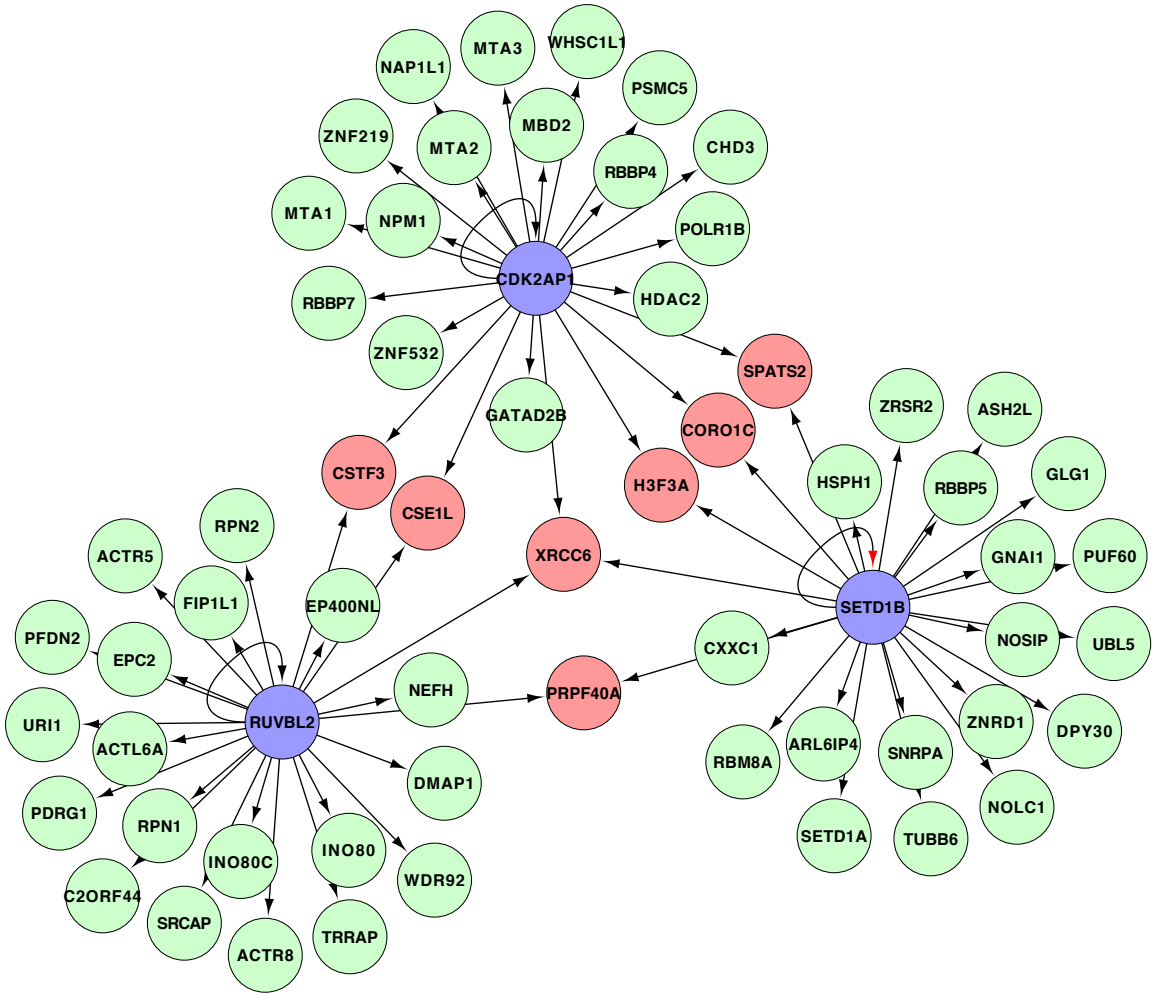


Figure 4.12: Network view of enriched interactors
Cytoscape network view for interactors of RUVBL2 (left), CDK2AP1 (up) and SETD1B (right). Common interactors are illustrated in red, interactors in green and baits are displayed blue. RUVBL2 and SETD1B are linked via CSE1L and CSTF3, CDK2AP1 and SETD1B are linked via SPATS2, CORO1C and H3F3A, RUVBL2 and SETD1B are linked via PRPF40A while XRCC6, a single stranded DNA-dependent ATP dependent helicase, was observed as a common interactor of all three proteins.

4.1.9. The composition of protein complexes is dependent on cellular fate

The observation that protein complexes are highly dynamic systems with the ability to change their protein composition dependent on cell cycle, cell fate decisions and environmental influences, increasingly challenges the classical model of defined core complexes (Colucci-D'amato et al., 2011; Havugimana et al., 2012). To investigate differentiation-associated differences in complex composition, BAC transgenic lt-NES cells were differentiated for 4 weeks before cells were harvested and analyzed by QUBIC. In order to detect differences in complex composition, cell lines harboring tagged proteins that maintained expression in postmitotic neurons were chosen. A comparison of each BAC-transgenic lt-NES cell line in its proliferative and differentiated state revealed variations of interactors caused by the differentiation process.

CDK2AP1 as part of the nucleosome remodeling and histone deacetylation (NuRD) complex presumably contributes to epigenetic regulation during differentiation (Spruijt et al., 2010). In BAC-transgenic CDK2AP1.GFP$^+$ neuronal cell populations, significant differences in the composition of the NuRD complex were observed when compared to CDK2AP1.GFP$^+$ lt-NES cells: While the H3.3A variant of histone 3, discovered only at active genomic loci and both retinoblastoma-binding proteins P46 and P48 and involved in histone acetylation and deacetylation, were significantly enriched in lt-NES cells, they were not detectable in postmitotic cell copurifications (Chow et al., 2005). Furthermore, GATAD2A, a transcriptional repressor, and HDAC1, a histone deacetylase, both copurified with CDK2AP1 only in differentiated cultures (Fig. 4.13 a).

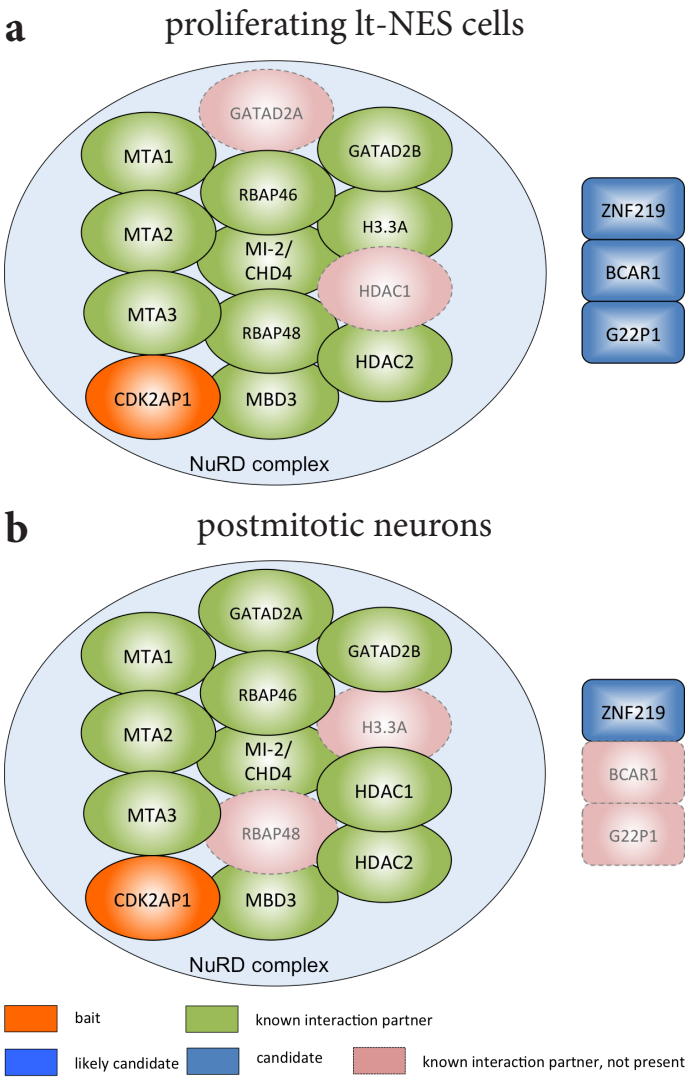


Figure 4.13: Differentiation-dependent variations in protein interactor levels of CDK2AP1
Illustrated proteins were all annotated as significantly enriched in respective QUBIC analyses. ***a****: In contrast to CDK2AP1 transgenic lt-NES cells,* ***b:*** *H3.3A, RBAP46 and RBAP48 were not detectable in neuronal cells while HDAC1 and GATAD2A were enriched.*

In SETD1B transgenic lt-NES cells, the majority of the reported members of the COMPASS complex were present including RBBP5, H3.3A, ASH2L, CXXC1, DPY-30 and SETD1A. SETD1B is a component of the SET1 histone methyltransferase complex, specifically methylating 'Lys-4' of histone H3 under the prerequisite that the neighboring 'Lys-9' is not already methylated (Fig. 4.14 a). This targeted methylation represents a specific tag for epigenetic transcriptional activation (Lee et al., 2007). Following differentiation of lt-NES cells, three components of the COMPASS complex could not be copurified in differentiated SETD1B.GFP cells: DPY-30, a core component of several methyltransferase-containing complexes (Wang et al., 2009b), the active variant of Histone H3 (H3.3A; Chow et al 2005) and SETD1A, a Set1/Ash2 histone methyltransferase complex subunit (Fig. 4.14 b). SETD1A recognizes a

different set of target sequences than SETD1B and contributes in a non-redundant manner to the epigenetic control of chromatin structure and gene expression (Lee et al., 2007).

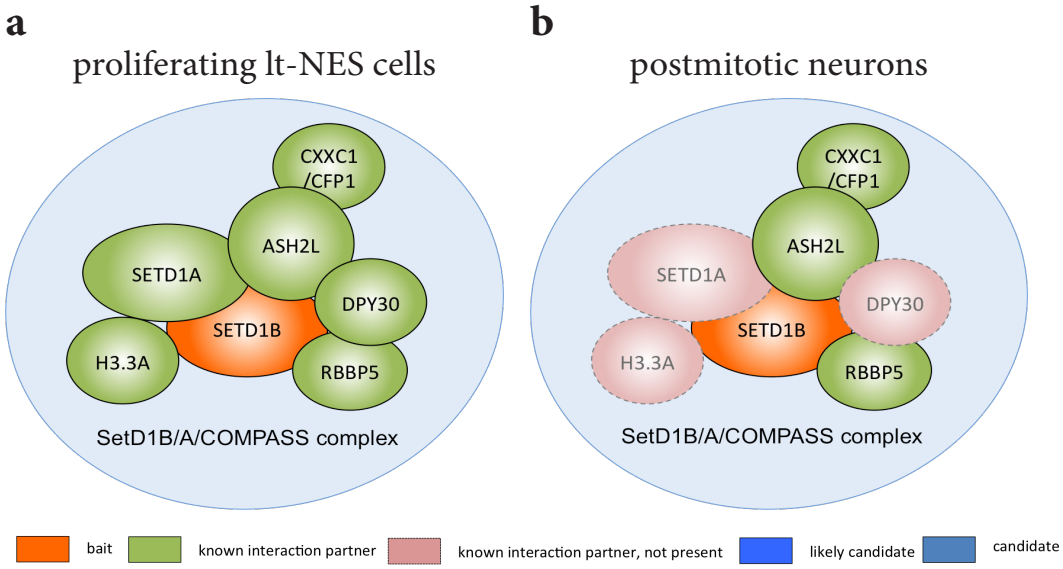


Figure 4.14: Differentiation-dependent variations in protein interactor levels of SETD1B
Illustrated proteins were all annotated as significantly enriched in respective QUBIC analyses. ***a****: SETD1B complex composition of lt-NES cells.* ***b****: In SETD1B.GFP transgenic differentiated cells, SETD1A, H3.3A and DPY30 of the COMPASS complex were not detectable.*

The analysis of RUVBL2 interactors in both conditions (Fig. 4.15 a, b) showed a massive rearrangement of interactors. Here, EPC2 (enhancer of polycomb homolog 2) was bound by RUVBL2 exclusively in proliferating lt-NES cultures. Along with the dissociation of the nuclear precursor complex core component BAF53 (53 kDa BRG1-associated factor A) in neurons, which facilitates transcriptional activation of specific genes by antagonizing chromatin-mediated transcriptional repression, several members of all major complexes could not be detected by QUBIC in differentiated neuronal cultures. These included TRRAP, an adaptor protein of the histone acetyltransferase (HAT) complexes from the NuA4 complex (Cai et al., 2003), PFDN2 (Prefoldin subunit 2) from the URI complex, INO80 and INO80N/ACTR8 (Actin-related protein 8) from the INO80 complex and ACTR5 (Actin-related protein 5) from the SRCAP (Snf2-related CBP activator) complex. This shift in composition was accompanied by the recruitment of INO80D, a member of the INO80 complex and ACTL6B/BAF53B (BRG1-associated factor 53B), a core component of the neuron-specific chromatin remodeling complex nBAF, which is required for postmitotic neuronal development and dendritic outgrowth (Yoo et al., 2011b). Together, these data indicate that the differentiation state of the cell type deployed for QUBIC notably influenced the composition of the protein complexes.

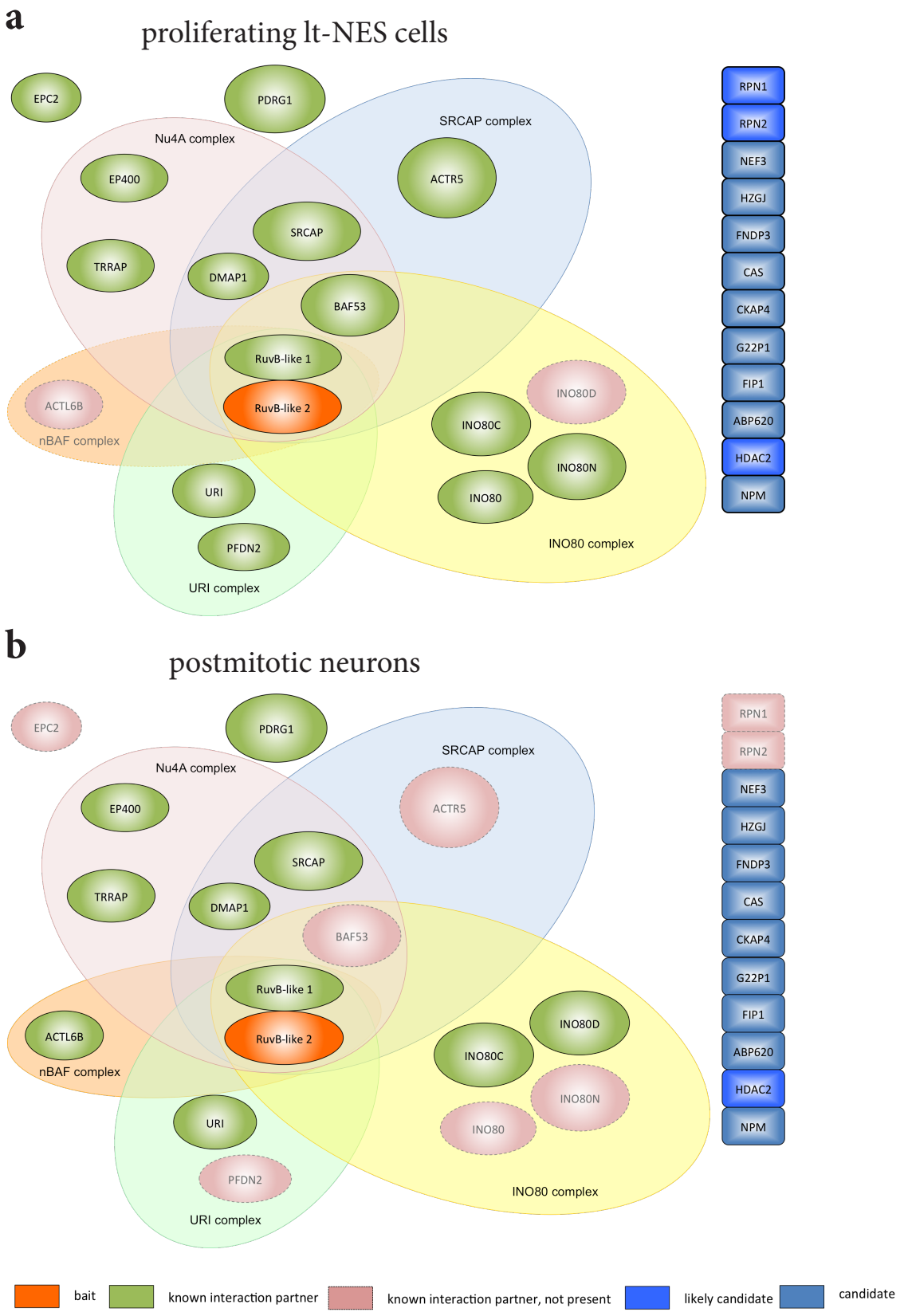


Figure 4.15: Differentiation-dependent variations in protein interactor levels of RUVBL2
Illustrated proteins were all annotated as significantly enriched in respective QUBIC analyses. ***a****: In RUVBL2.GFP transgenic lt-NES cells representatives of the SRCAP-, the NuA4-, the INO- and the URI complex were detected.* ***b:*** *In differentiated cells, several of these proteins were absent, including TRRAP of the Nu4A complex, ACTR5 of the SRCAP complex, INO80N, INO80A and INO80K of the INO80 complex and PFDN2 of the URI complex. In RUVBL2.GFP neurons, a member of the neuron-specific nBAF complex, ACTL6B was enriched. Both candidates linked with TRRAP complexes, ribophorin I/III (RPN1/2), were absent in neuronal cultures.*

4.2. Translation of BAC transgenesis to iPSC-derived lt-NES cells

Protein-protein interaction data generated in ES cell-derived neural stem cells and their neuronal progeny can help to characterize a variety of cellular processes including aspects of metabolism, signaling and differentiation. In order to analyze influences of specific mutations on binding affinities, mutated bait proteins could be used. While these approaches are likely to raise essential data and may lead to a more comprehensive understanding of the concept of diseases, multifactorial disorders or diseases with spontaneous onset remain difficult to study. In order to investigate differences in protein binding affinities in complex diseases or in diseases with unknown genetical cause, QUBIC analysis has to be transferred into a relevant cell population derived from affected individuals. Therefore, data should be generated in cell populations derived from patients with clinical manifestation of respective diseases, providing not only one aberrant protein isoform but the whole complex proteomic environment leading to the clinical picture. To explore this approach, induced pluripotent stem (iPS) cells from patients and healthy controls were generated, differentiated to lt-NES cells, and protocols adapted to permit the generation of stable BAC transgenic donor-specific cell lines that were subsequently used for QUBIC analysis.

4.2.1. Generation and validation of iPSCs

To generate human iPSC lines from adult dermal fibroblasts, cells were transduced with the non-integrating Sendai virus particles for the transient expression of the transcription factors KLF4, c-Myc, Oct3/4 and Sox2 (Fig. 4.16 a). Seven days following transduction and cultivation in iPSC medium containg 20% serum replacement and bFGF, transduced fibroblasts were transferred to a layer of irradiated murine fibroblasts and cultured in iPS medium until clonal growth of arising colonies was observed (between week 3 and 4 after transduction). By mechanical isolation of single iPSC colonies, clonal cell lines were generated that showed a clear-zoned colony-shaped growth (Fig. 4.16 b) and were positive for the pluripotency markers Tra 1-60, Tra 1-81 and SSEA4 and negative for the viral hemagglutinin-neuraminidase protein (HN, Kajiwara et al., 2012), indicating the absence of active virus (Fig. 4.16 c). In order to select iPSC lines that had not acquired overt chromosomal aberrations during the process of reprogramming, each

line was tested by genome-spanning single nucleotide-polymorphism (SNP) analysis (Fig. 4.16 d Mayshar et al., 2010). From 23 analyzed lines, 2 lines acquired duplications on chromosome 14 while no further abnormalities could be observed using standard settings of the CNV partition plugin v3.2.0 that highlights chromosomal copy number changes of regions ≥ 1 Mb containing ≥ 50 continuous SNP probes (Fig. 4.16 e). Using this method, a sufficient number of lines could be generated from each individual, resulting in one or more intact, fully functional iPS cell line.

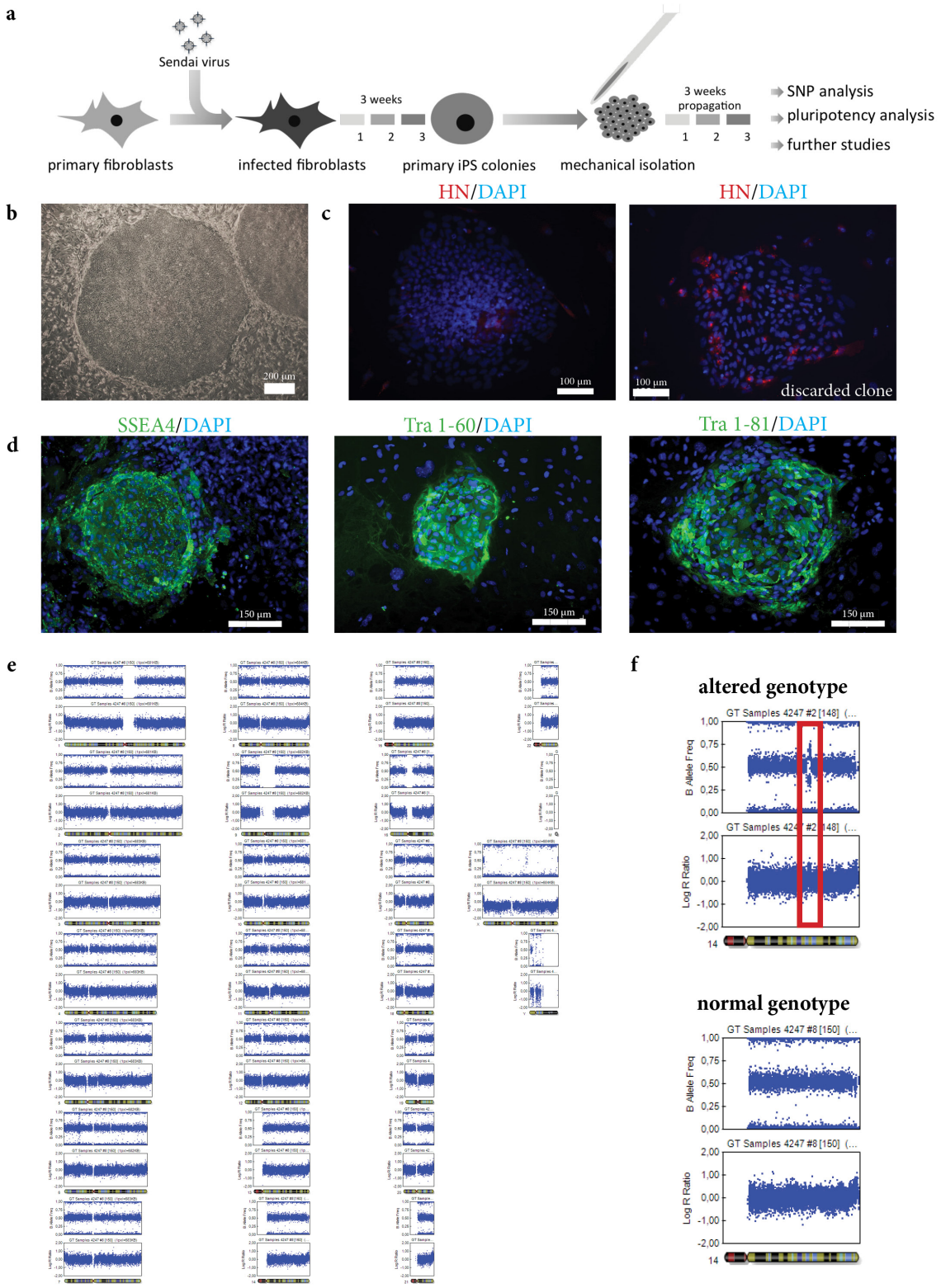


Figure 4.16: Generation and quality control of induced pluripotent stem cells
***a**: Schematic illustration of experimental workflow for the generation of human induced pluripotent stem cells (iPSCs). Healthy control and patient-derived fibroblasts were transduced with Sendai virus coding for c-Myc, Oct3/4, Sox2 and Klf4 and cultivated for 3 weeks until colonies formed. Colonies were isolated mechanically, propagated for 3 weeks and validated. **b**: Phase contrast image of iPSC colony before mechanical isolation. Scale bar = 200 µm. **c**: Immunocytochemical analysis for viral HN protein (red) of iPSC clones with silenced (left) and maintained (right) viral HN expression (discarded clone). Scale bar = 100 µm. **d**: Immunocytochemical analysis for pluripotency-associated proteins Tra 1-60, Tra 1-81 and SSEA4. Scale bar = 150 µm. **e:** Representative genome-spanning single nucleotide polymorphism analysis of a iPSC line with no detectable abnormalities. **f**: Exemplary chromosomal aberration on chromosome 14 (shown lines 4247 #8 (normal), 4247 #2 (altered)).*

4.2.2. Derivation of stable lt-NES cell lines from integration-free iPSCs

The stable and reproducible generation of well-defined neural stem cells and neuronal cultures from different hESC and iPSC lines is an important prerequisite for the reproducibility and comparability of quantitative PPI analysis. Commonly applied 'run-through' protocols, where hPSCs are directly differentiated into mature neuronal cultures often suffer from non-neural cell contaminants due to incomplete differentiation, varying neurogenic potential of different hPSC lines and batch-to-batch variations, which are inherent to lengthy differentiation protocols (Kim et al., 2011). Lt-NES cells represent a stable intermediate multipotent stem cell population between pluripotent stem cells and differentiated neuronal cultures. Once established, lt-NES cells can self-renew as a homogenous population across many passages, stably give rise to mature cultures of postmitotic human neurons and are readily accessible to genetic modification. In fact, lt-NES cells can be propagated for more than 60 passages while maintaining doubling times and avoiding senescence (Falk et al., 2012). In order to generate lt-NES cells, human ES- or iPS cells were detached, differentiated as free floating embryoid bodies (EBs) and allowed to differentiate for 5 days (Fig. 4.17 a). Following plating, neural tube-like structures emerged (Fig. 4.17 b), which were mechanically isolated, grown as neurospheres for 7 days (Fig. 4.17 c), dissociated and further cultured as a homogenous lt-NES cell line in the presence of FGF2, EGF and B27 (Fig. 4.17 d).

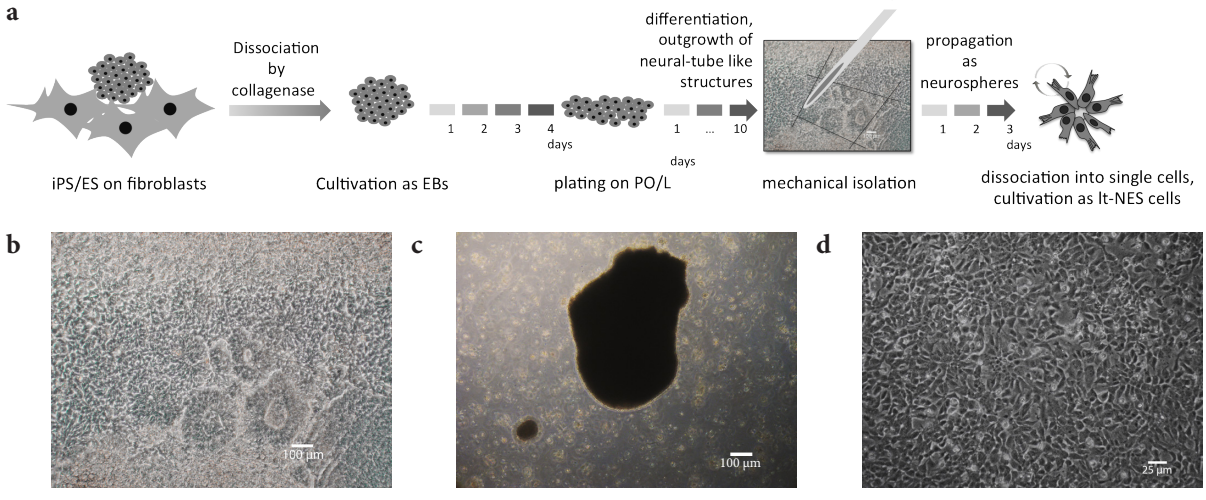


Figure 4.17: Generation of long-term neuroepithelial like stem cells (lt-NES cells)
***a**: Schematic illustration of the lt-NES generation workflow. iPS/ES colonies cultured on fibroblasts were dissociated by collagenase treatment, cultured as free-floating embryoid bodies (EB) for 5 days, plated on PO/L TC dishes and differentiated for 10 days in N2 medium containing FGF2. Neural tube-like structures were mechanically isolated, expanded as free-floating aggregates until neurospheres emerged after 3 days, which were then triturated to single cells serving as starting population for lt-NES cell cultures. **b**: Neural tube-like structures in plated EBs. **c**: Floating neurosphere. Scale bar = 100 µm. **d**: Phase contrast image of established lt-NES cells. Scale bar = 25 µm.*

Human iPS-derived lt-NES cells grow as a homogenous population expressing PLZF, Sox2 and DACH1 in the nucleus, nestin as an intermediate filament and ZO1, a tight junction protein typically found in an apical location in the center of the rosette structures (Fig. 4.18, Koch et al., 2009). Analogous to hESC-derived lt-NES cells, this cell population can be cultured in the presence of growth factors for >60 passages as stable stem cell population and, upon growth factor withdrawal, differentiates to major fractions of β3-tubulin- and MAP2ab- positive postmitotic neurons and a minor fraction of glial fibrillary acidic protein (GFAP)-positive astrocytes (Fig. 4.18 a, c).

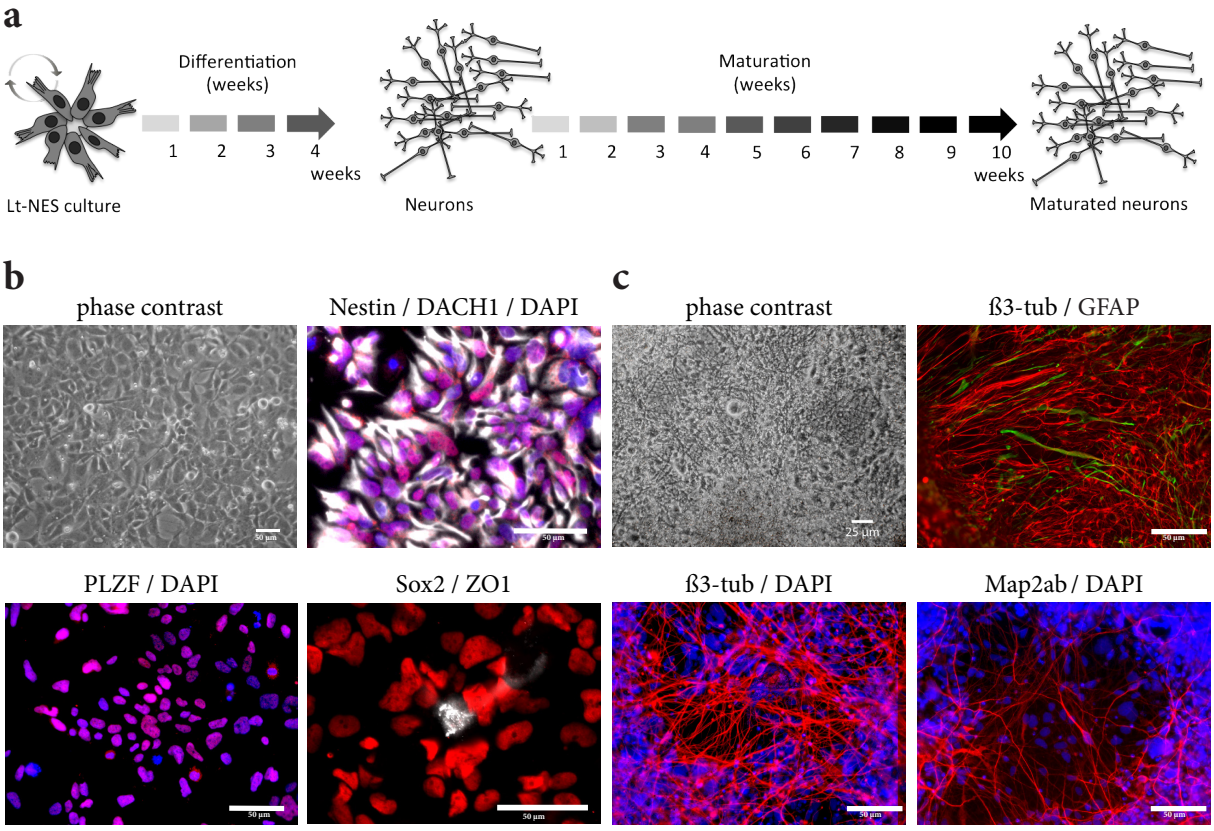


Figure 4.18: Characterization of proliferative and differentiated lt-NES cells
***a**: Schematic illustration of differentiation timeline. Differentiation was induced by withdrawal of growth factors. 4 weeks of differentiation generated immature neurons, which could be matured for more than 10 additional weeks. **b**: Phase contrast image and immunoflourescent images for DAPI (blue), nestin (white), DACH1 (red), PLZF (red), Sox2 (red) and ZO1 (white) of lt-NES cell line 4247#8. **c**: Neuronal cultures after 4 weeks of differentiation were stained for neuronal markers β3-tubulin, MAP2ab and the astrocytic marker GFAP. Scale bars = 50 µm.*

4.2.3. The interactome of human iPS-derived lt-NES cells is similar to hES-derived lt-NES cells

To validate the amenability of iPS cell-derived lt-NES cells for PPI studies, protocols for transfection and enrichment were adapted using the BAC harboring PCNA.GFP. For generation of transgenic cell lines, transfection protocols were adopted but G418 concentrations were generally reduced by 50%. A complete validation of lt-NES-specific marker expression for iPS cell-derived lt-NES cells showed no significant difference while PCNA.GFP signal distribution (Fig. 4.19 a) and FACS analysis revealed similarly high transgene incorporation (Fig. 4.19 b). QUBIC analysis resulted in an analogous pattern of known interaction partners as detected in ES-cell derived cells but generally showed an increase of more significantly enriched interaction candidates (46 versus 28, data not shown; Fig. 4.19 c). The results demonstrate that the established technique of BAC-mediated protein expression can be applied to iPS-derived lt-NES cells and will should thus allow further studies of protein-protein interactions in neural and neuronal cell populations from patients e.g. with neurological diseases.

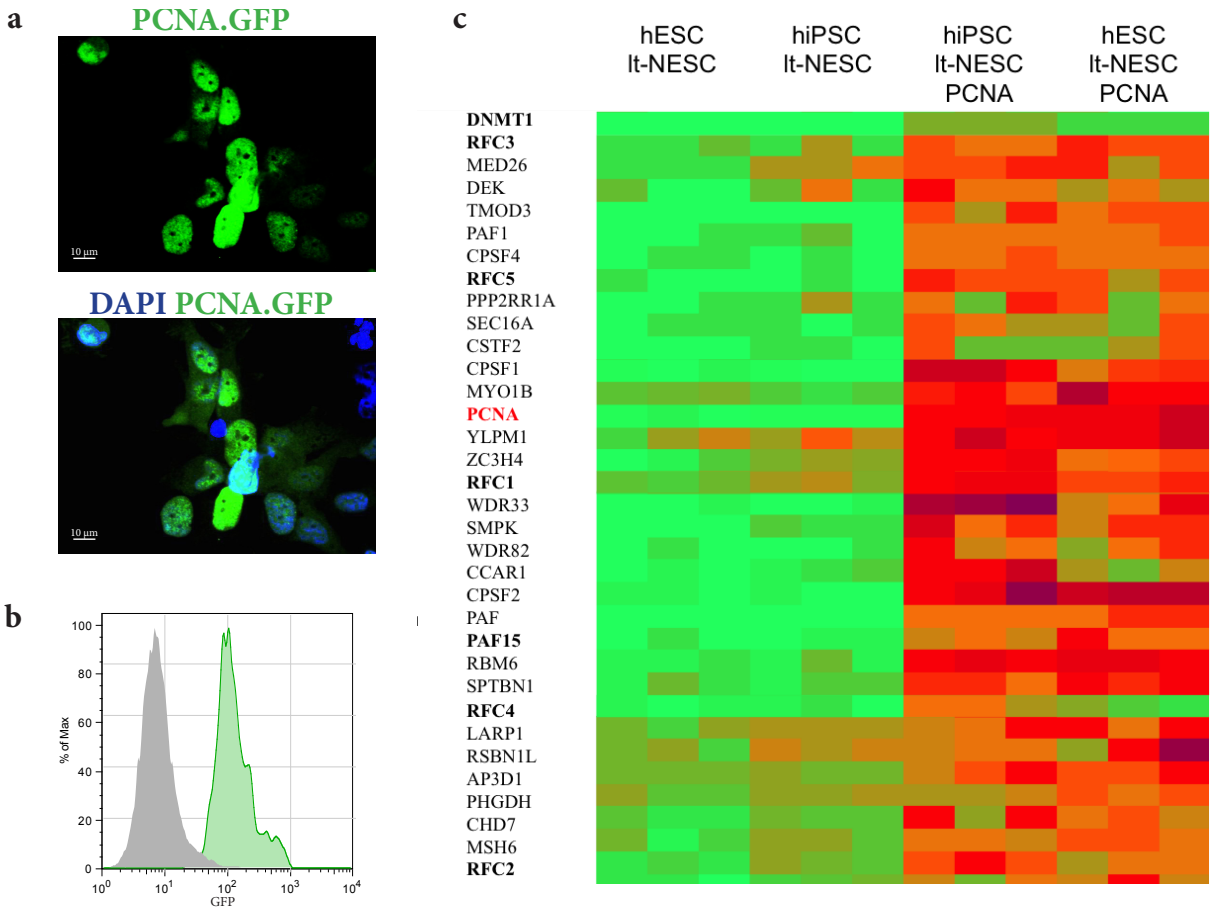


Figure 4.19: BAC-based protein-protein interaction analysis in iPSC-derived lt-NES cells
a: *PCNA.GFP-BAC transgenic iPS-derived (control#8) lt-NES cells. GFP (green), DAPI (blue). Scale bar = 10 µm.* ***b:*** *FACS analysis of PCNA.GFP-BAC transgenic iPS-derived lt-NES cells (green) opposed to non-transgenic lt-NES cells (grey).* ***c:*** *QUBIC analysis of PCNA interactors in iPSC-derived lt-NES cells. The heat plot shows enrichment of multiple proteins in both iPSC and ESC-derived lt-NES cells positive for PCNA.GFP. Both populations show enrichment for RFC1, RFC2, RFC3, RFC4, PAF15 and DNMT1, key components of the PCNA complex. Heat plot rendering using the software Perseus.*

4.3. Evaluation of Adeno-associated virus (AAV)-mediated gene targeting as tool for the integration of epitope tags into endogenous genes

Although BAC-mediated protein expression is well suitable for most - if not all - proteins, the size of a GFP protein as tag together with the probability of a slight overexpression due to the additional copy caused by the integrated BAC might be insufficient for some studies. Therefore, it would be preferable for some applications to perform an epitope tagging of endogenous genes rather than inserting an additional copy (Kim et al., 2008). To explore the feasibility of such an approach, recombinant AAV particles were generated to introduce a FLAG tag to the N-terminus of Ataxin-3 (rAAV-Atxn3N), a deubiquitinating enzyme that causes an aggregation of its C-terminal fragments in patients with Machado-Joseph disease (MJD) when containing pathologically enriched numbers of CAG repeats. Subsequently, a N-FLAG-Atxn3 tagged lt-NES cell line derived from a patient with MJD was generated in a proof-of-principle study.

4.3.1. AAV-mediated gene targeting permits the labelling of endogenous proteins by introduction of epitope tags

AAV-based targeting systems rely on the modification of the viral targeting sequences that usually mediate the integration of the provirus sequence to the AAVS1 locus on chromosome 19 (Deyle and Russell, 2009; Kohli et al., 2004). For the design of targeting constructs, the following requirements have to be met: The composition of the target region has to contain less than 50% of interspersed repeats and low complexity DNA sequences, encompass at least one exon for modification and comprise at least 2 kb. This region is then divided into two 1 kb segments, while the central sequence always has to be a region in between processed exons. A synthetic exon promoter trap cassette (SEPT) is placed between these elements to facilitate enrichment of correctly targeted cells by application of antibiotics (Fig. 4.20 a). However, an inserted SEPT cassette prevents the complete transcription of the targeted gene (Kim et al., 2008). In order to eliminate the SEPT cassette, it can be flanked by loxP sites to enable its excision by Cre expression. Here, Cre was transiently expressed by transfection of transgenic lt-NES cells with synthetic modified mRNA to prevent further genomic alteration of the targeted cells. By this method, the expression of an epitope-tagged endogenous protein can be achieved.

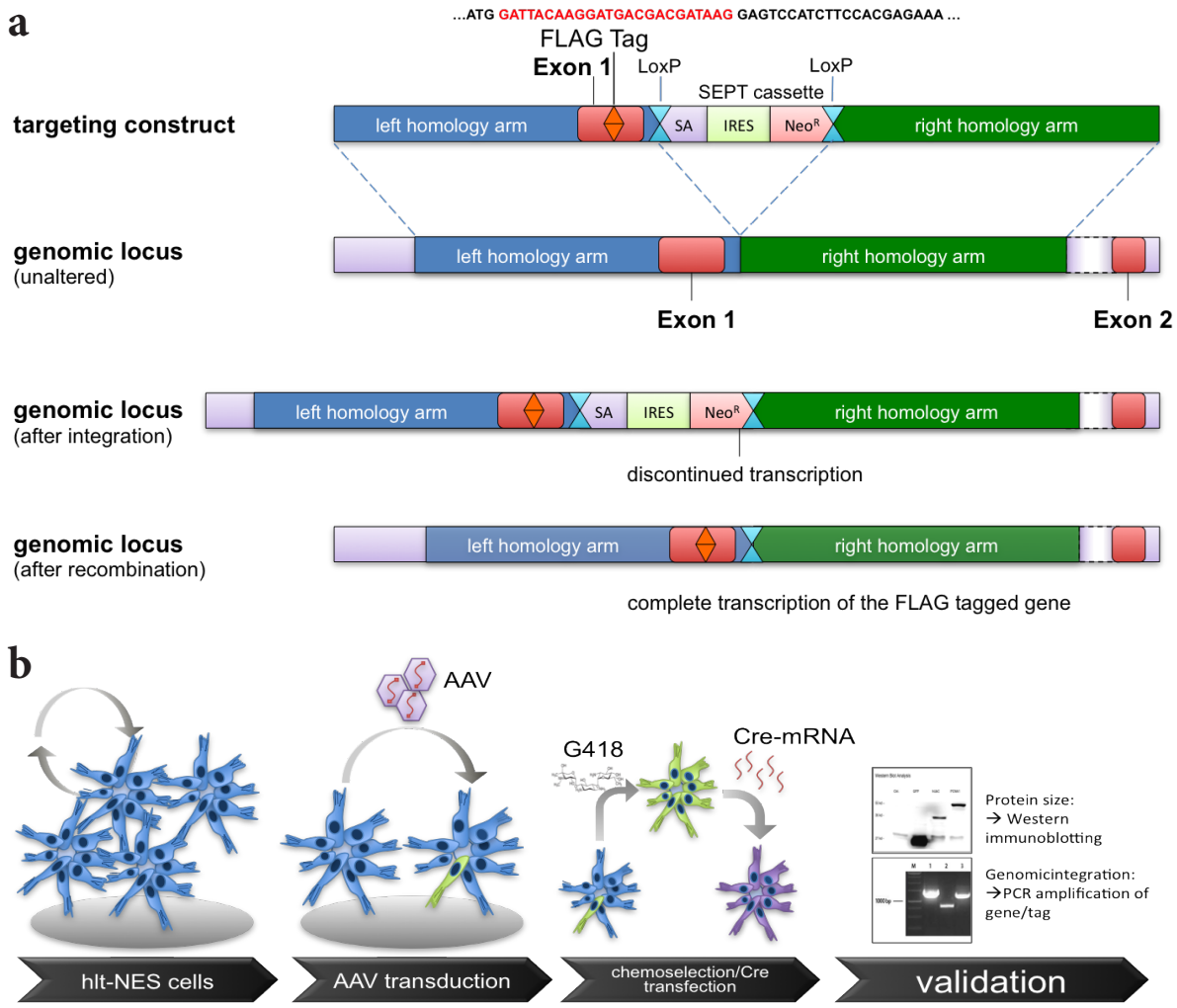


Figure 4.20: Design and experimental procedure of AAV-mediated gene targeting
***a**: The construct for AAV-mediated gene targeting consists of two homology arms (≥ 1000 bp) and the exon containing the FLAG tag. A synthetic exon promoter trap (SEPT) flanked by loxP Cre recombination sites is located between both homology arms. **b**: Five days after transduction of cells, G418 (0,2 - 0,5 µg / ml G418, 168 h) was applied to select for transgenic cells. Cell pools were expanded and transfected with synthetic modified Cre recombinase mRNA in order to….*

4.3.2. Epitope-tagged endogenous proteins can be used to determine their localization pattern and size

Using engineered AAV particles, a precise manipulation of a distinct locus of the genome was conducted to introduce a FLAG tag directly before the initiating methionine, resulting in N-terminally tagged Atxn3 protein variants. Lt-NES cells derived from patients suffering from Machado-Josephs disease were transduced and enriched using antibiotics (0,2 – 0,5 µg / ml G418, 168h). From a pool of $2x10^6$ cells, approximately 5 - 30 resistant colonies emerged, which was reported to be dependent on the quality of the flanking strand (Vasileva and Jessberger, 2005).

The disease-associated isoform of Ataxin-3 contains an increased number of CAG repeats that are causative for disease development (Rüb et al., 2013). In MJD patients, usually both the normal ATXN3 (42 kDa) and an expanded ATXN3 (58 kDa, Koch et al., 2011) can be found and therefore easily be distinguished by their sizes. Successful targeting was analyzed after Cre-mediated excision of the SEPT cassette by Western immunoblotting, resulting in two FLAG-positive proteins of 42 kDa and 58 kDa (Fig. 4.21 a). The observed protein sizes correspond to the calculated protein sizes of FLAG-tagged Atxn3 proteins that are either pathologically elongated (58 kDa) or unaltered (42 kDa). In immunohistochemical analyses using an antibody directed against the FLAG epitope tag, fluorescent signals in targeted MJD lt-NES cells were observed (Fig. 4.21 b).

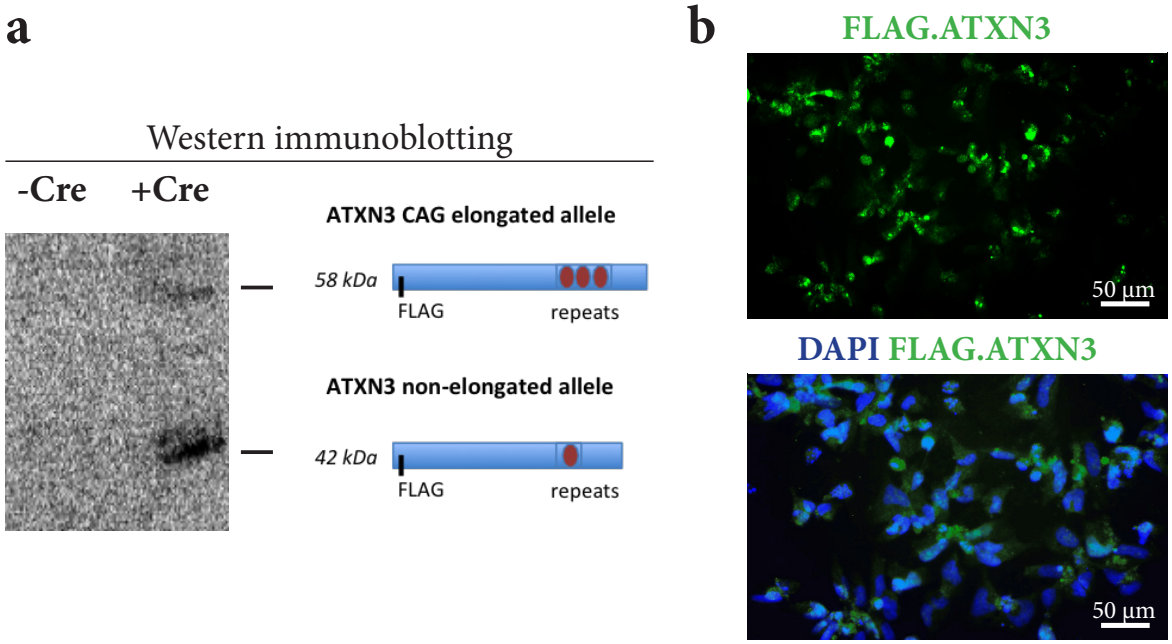


Figure 4.21: AAV-mediated gene targeting in lt-NES cells derived from MJD patients
***a**: A Western immunoblot using FLAG antibody shows two protein variants corresponding to the sizes of non-elongated Atxn-3 and extended poly glutamine Atxn-3, indicating a modification of both alleles present in these cells by introduction of a FLAG tag. **b**: Immunocytochemical analysis of Atxn3-targeted lt-NES cells shows the expression of FLAG tag positive proteins in targeted cells. Scale bars = 50 µm.*

These results demonstrate, that, in principle, the technique of AAV-mediated epitope tagging can be applied to lt-NES cells to generate somatic cell populations with endogenously tagged proteins. The observation that both, the disease and the healthy allele of Atxn3 in lt-NES cells derived from a patient suffering from Machado-Joseph disease were successfully targeted using an AAV-based approach, might enable AAV-based gene correction approaches.

5. Discussion

5.1. Suitability of different cellular systems for the analysis of protein-protein interactions

The high degree of standardization required for reliable protein-protein interaction (PPI) studies deterred scientists for decades from using primary cells obtained from multiple tissue donors in their studies. Instead, naturally occurring immortal cells, derived from various isolated tumors, were employed and provided an identical genetic background particularly suitable for repeatable scientific experiments (HeLa, 293, Scherer et al., 1953; Schneider et al., 1977). The growing demand for cells derived from a specific tissue or certain cellular subtypes resulted in the establishment of many immortalized cell lines using a variety of molecular biological techniques including random mutagenesis, introduction of viral genes, the expression of cell cycle manipulating proteins or the cellular fusion with other tumor-derived cells (Cohen et al., 2007; Graham et al., 1977; Milstein, 1999; Varma and Leavitt, 1988). Most PPI studies rely on the utilization of tagged genes for the analysis of their interaction partners. Immortal cell lines usually represent a genetically readily accessible resource that allows the quick establishment of clonal cell lines. The majority of present findings concerning cell signaling cascades, protein complex composition, the spatial distribution of proteins in subcellular compartments and the detection of targets for pharmaceutical intervention in cellular processes were obtained using immortalized cells. However, their application has some significant drawbacks: Most protocols for the generation of transgenic cell lines are highly adapted to “classic” lines like HeLa or 293 cells and allure scientists to conduct their studies using a set of prefabricated protocols and biological materials, saving time by avoiding an elaborate establishment of adapted protocols. Data generated in this context is rarely tissue specific and in some cases could lead to false-positive results as introduced bait-proteins can interact with proteins their physiological environment normally would prevent them from or to false-negative results when potential interaction partners are absent. For example, if a neuron-specific protein is exogenously expressed in the ovarial carcinoma cell line HeLa, it may interact with proliferation-associated proteins that naturally are nonexistent in postmitotic cells. In many cases, cell lines of tissues corresponding to the physiological setting of the bait protein exist and would enable tissue specific analyses.

The high degree of standardization of cultivation techniques enabled the distribution of these cell lines to different laboratories as well as the independent investigation and validation of specific findings, facilitating a detailed characterization of employed lines. However, tissue-specific cell lines are rarely used for PPI analyses as common lines as e.g. HeLa or 293T are widely spread, backed by a rich collection of adapted protocols and exhibit a broad protein expression pattern suitable for most problems (Havugimana et al., 2012; Trinkle-Mulcahy et al., 2008). In addition, cell lines have acquired significant mutations to become immortal: Immortalization alters the biology of the cell substantially and leads to dysregulation of proteins associated with cell cycle regulation and its associated metabolism. To overcome issues associated with immortal cell lines or tissue specificity, experiments are also conducted in animals – in most cases serving as subsequent validation steps. Here, the function of certain proteins can be assessed directly in the tissue of interest e.g. by the generation of genetically engineered animals or the cultivation of specific cell populations of certain tissues derived from these animals for a limited period of time. While the interpretation of data raised in mice provides impeccable information concerning murine proteins, data generated from investigations of human proteins in such an environment should be treated with caution: Despite conservation of most functional domains and high similarities between murine and human proteins, a difference of one single amino acid can lead to entirely different structures with altered interaction partners or a complete incapacitation of relevant protein binding sites (Anfinsen, 1973). When aiming for most authentic PPI data possible, non-transformed human cells of matching tissue represent the ideal candidates. While primary cells have the advantage to resemble the tissue of interest closely, they lack the capacity for self-renewal or enter senescence after about 15 passages – a circumstance that requires a permanent replenishment of primary cells from varying sources with naturally occurring batch-to-batch variations (Golemis, 2005). Furthermore, the high variability of transfection efficiencies for different primary cells poses major obstacles for the introduction of expression constructs. While numerous methods for transfection, e.g. like different electroporation protocols, liposome-based transfection methods or $CaPO_4$-based complexes, were developed, few of them are comparable to transfection efficiencies of common cell lines. When aiming for the generation of primary cells with stable transgene integration, retroviral-based methods have to be used. As primary cultures consist primarily of postmitotic cells, only lentivirus-based

vectors are suitable for their transduction (Blömer et al., 1997; Graham and van der Eb, 1973).

Pluripotent stem cell-derived cell populations exhibit proteomic environments closely related to somatic cell populations

The cellular system, in which PPI studies are conducted, strongly influences the proteins enriched for the respective bait (Aebersold and Mann, 2003). This phenomenon is attributable to a multitude of factors, acting directly on binding characteristics: The proteome of a cell is subject to change during progression of the cell cycle and can vary in response to extrinsic stimuli or developmental clues, inducing an up- or down regulation of certain protein levels. Additionally, the degree of differentiation of generated cell populations as well as cell fate decisions are able to induce or deplete the existence of large sets of cell type specific proteins, either allowing or preventing interactions with the bait. For in-depth analysis of tissue-specific protein complex compositions or effects of disease-induced changes, such studies ideally should be conducted in the cell type relevant for the effect to be examined. Moreover, investigations aimed at the characterization of disease-linked alterations of protein binding patterns would be desirable in cells affected by the disease of interest. In this context, pluripotent stem cells provide an intriguing platform that can be used to generate any desired cell type, provided a tailored differentiation protocol is available.

Lt-NES cells and their neuronal progeny as a CNS-model for PPI analyses

Due to the limited availability of standardized and defined primary tissue from healthy or diseased individuals and its restricted accessibility to genetic manipulation, most PPI studies up to now were conducted in transformed or tumor cell lines derived from a subset of tissues. This problem is even more relevant for primary human brain tissue, consisting of a multitude of cell types, delicate to separate, manipulate and even harder to standardize. Therefore, cultures of human neurons generated from hPS cells provide a promising alternative for the study of PPIs in the relevant cell type. One important prerequisite for harnessing human stem cell-derived neuronal cultures for PPIs is the stable and efficient production of highly enriched neuronal cultures. In commonly applied run-through protocols, hPS cells are directly

differentiated into neuronal cultures, yet often contain cell contaminants and significant batch-to-batch variations. Pluripotent stem cell-derived long-term self-renewing neuroepithelial stem cells (lt-NES cells) are a stable and robust cell population, which enable the fast generation of authentic human neurons. Furthermore, this population features continuous expandability, a stable neuronal and glial differentiation competence, a characteristic transcription factor profile, a regional identity resembling an anterior ventral hindbrain specification amenable to regional patterning as well as the capacity to generate functionally mature neurons (Falk et al., 2012). In the course of this study, hES- and iPS cell-derived lt-NES cells were used as a starting population for genetic manipulation, yielding highly enriched neuronal cultures within four weeks of differentiation in both manipulated and wild type condition. Without application of patterning cues, the default neuronal subtypes generated using this approach are GABAergic interneurons (Koch et al., 2009). By application of developmental cues in form of growth factors like RA/FGF2/EGF/SHH or SHH/FGF8/AA, different neuronal subtypes including motoneurons of ventral spinal cord or tyrosine hydroxylase positive neurons with ventral midbrain identity, respectively, can be generated. The preserved regional plasticity of lt-NES cell populations therefore allows for a subtype-specific evaluation of PPI differences (Elkabetz and Studer, 2008). Recapitulating, present protocols for lt-NES- and neuronal cell derivation represent a reliable method for the reproducible generation of highly enriched cultures of human neuronal precursors and mature neurons that should be suitable for standardized mass spectrometric PPI analysis. Furthermore, the option to generate diverse neuronal subtypes enables specific investigations in affected neuronal subpopulations for various diseases.

5.2. Overexpression versus BAC-mediated expression of tagged proteins

Cells have to tightly regulate their proteins in response to cell cycle-dependent, metabolic, environmental or developmental influences to fulfill their physiological role within the tissue. In this work, it was a key prerequisite to introduce tagged proteins and analyze their binding behavior by mass spectrometry in a manner as close to the endogenous situation as possible while retaining the capacity to perform this task in a streamlined and timely efficient manner. The introduction of a tagged protein into a cell can be accomplished by a large variety of techniques with individual benefits and drawbacks. While most highly efficient transgenic strategies rely on the introduction

of a cDNA under control of a strong and small promoter (e.g. CMV), the main mechanisms of endogenous regulation or processing are prevented. A site-specific manipulation by targeted introduction of gene modifications via homologous recombination meets these criteria but used to be highly inefficient until genome-editing techniques like the TALEN or CRISPR/Cas9 platforms were developed (Fu et al., 2013; Miller et al., 2011). An alternative to site-specific genome editing is the application of bacterial artificial chromosome (BAC)-based expression vectors. These constructs are expected to contain most, if not all, regulatory elements and all intronic and exonic elements. Applying these large constructs, stable cell lines can be generated and were successfully used in numerous studies (Hubner et al., 2010a; Placantonakis et al., 2009; Poser et al., 2008; Spruijt et al., 2010). The evaluation of the best strategy for BAC transgenesis in lt-NES cells for more than 30 constructs was enabled by the availability of a vast library of tagged genes present on BACs generated within the framework of the DiGtoP consortium (a NGFN funded network). N-terminal tagged forms contain a GFP fused to the first exon with an artificial intron containing a PGK driven neomycin resistance gene (neo^R) while C-terminal tagged forms exhibit a tag fused to the last exon followed by an IRES-linked neo^R. The generated transgenic lt-NES cell lines were investigated for protein localization by confocal microscopy, the proportion of cells comprising naïve fluorescent GFP-fusion proteins by flow-cytometry, alternative protein spliceforms by western immunoblotting and exemplary functional assays for the evaluation of protein regulation. Correct protein localization is one of the most important aspects when investigating PPIs as the spatial organization of proteins within a cell – and especially in neurons – substantially influences the classes of possible protein interaction partners. Trafficking elements on mRNA level or on protein precursors can influence protein levels in confined compartments of the cell more than the transcriptional control of the gene itself (Snee et al., 2002). An abundant presence of a protein, being the case in an overexpression setting, increases the risk, that it enters incorrect cell compartments and then interacts with proteins that it never normally contacts (Gibson et al., 2013). The BAC-transgenic lt-NES cell lines generated in this study exhibited a protein localization pattern closely resembling previously observed or predicted patterns (Uniprot database and Poser et al., 2008). For HeLa cells, expression levels of the tagged proteins using BAC-transgenesis have been reported previously to correspond to endogenous levels since these constructs contain

unaltered promoter regions (Hubner et al., 2010a; Poser et al., 2008; Stadler et al., 2013). herefore in this study, protein expression was assessed by flow cytometric and Western immunoblotting analysis detecting the fluorescent tag rather than protein-specific Western immunoblotting avoiding the need to establish more than 30 antibody-specific protocols.

The regulation of gene expression is not exclusively accomplished by modulation of the quantity or localization of the encoded protein but also by providing mRNA splice variants resulting in different protein isoforms (Black, 2003). By employing BACs for genetic manipulation of a target cell population, the regulation of this machinery is preserved, relying on the presence of all intronic- and exonic elements, except in the case of alternative first (for N-terminal tagging) or last (for C-terminal tagging) exons.

In this study, the endogenous-like behavior of tagged proteins expressed by BAC transgenesis was exemplarily addressed in context of the cell-cycle dependent protein regulation for Aurora kinase A, a microtubule-associated cell-cycle regulated kinase (AURKA), which was enriched in mitosis and localized to centrosomes and spindle microtubules, as described previously (Ding et al., 2011). Moreover, the differentiation-induced downregulation of a BAC-expressed protein was validated by neuronal differentiation assays using PCNA, a protein exclusively expressed in proliferating cells, were performed. The tagged protein not only was downregulated in the whole culture but also was completely absent in neurons expressing a post-mitotic mature neuronal marker, MAP2ab. The regulation was sufficiently precise to allow for fluorescence-based readouts in multiwell assays applying proliferation-impairing substances. Although this approach relied on the utilization of cell pools rather than clonal lines, neither Western blot analysis nor flow cytometry indicated a truncated protein or an expression in a minor fraction of cells. This behavior could largely be attributed to the IRES-linked expression of the selection marker, rendering it virtually impossible for a truncated form of a protein or the resistance gene alone to be expressed. Taken together, BACs as expression constructs facilitate endogenous-like expression of tagged proteins in a cell-type specific manner and BAC transgenesis can reliably be performed in lt-NES cells.

5.3. Protein-protein interaction analysis

In most cases, the analysis of localization and interaction of an endogenously expressed protein together with phenotypic data from loss-of-function studies is sufficient to analyze its molecular role. Unfortunately, many PPI datasets rely on the overexpression of tagged proteins consisting of a single splice form in transformed cell lines. Transformed and tumor cell lines are known to express a wide range of proteins, which are not assembled within healthy somatic tissue, much less a single somatic cell type (Uhlen et al., 2010b). This circumstance might already have led to well-accepted assumptions of putative protein-protein interactions that physiologically never exist. Likewise, important information regarding weak or indirect protein-protein interactions in non-transformed somatic cells might have gone unnoticed.

In this study, PPI analysis was performed using quantitative interaction proteomics together with BAC transgenesis to introduce a tagged protein expressed at physiological levels in a somatic neural stem cell population or neurons derived thereof. Using this approach, artifacts due to the application of protein-specific antibodies with varying specificity can be avoided since the same antibody for every GFP-tagged protein is used. Furthermore, endogenous expression of the tagged protein facilitates the study of protein complex composition upon differential, environmental- or developmental cues with high specificity (Hubner et al., 2010b). In the following section, available PPI databases and proteins that were used as baits as well as their identified potential interaction partners are discussed in detail.

5.3.1. Publicly available databases for the validation of PPI

Publicly available PPI databases were initially built to serve as a central interface to access data previously generated in multiple experiments from various laboratories. Hereby, different approaches for data collection were chosen to generate an easily accessible comprehensive representation of the current state of knowledge in the field. However, the more data was integrated in these databases, the more data sources had to be employed and this circumstance prevented the preservation of a consistent quality at the expense of credibility. Therefore, most databases can be configured to present a certain analysis level but generally are designed for specific user groups. An example for a very comprehensive PPI database is the ‘Search Tool for the Retrieval of Interacting Genes/Proteins’ (STRING) and can be used for the visualization and generation of functional protein association networks

(http://string-db.org/). It is a database of known and predicted protein interactions and includes direct (physical) and indirect (functional) associations derived from four sources: The genomic context filter allows displaying genes in immediate neighborhood on the genome (within 300 bp on the same strand) while protein conservation can be analyzed by a phylogenetic display. A powerful text-mining algorithm enables the up-to-date analysis of full text PubMed entries and can display whether certain proteins were co-mentioned or can analyze these data in a context-dependent manner. The integration of real experimental data in form of conserved co-expression databases as well as data generated by high-throughput experiments is also possible (Franceschini et al., 2013; Jensen et al., 2009; Snel et al., 2000; Szklarczyk et al., 2011; von Mering et al., 2003; von Mering et al., 2007).

The 'Comprehensive Resource of Mammalian protein complexes' (CORUM) database provides a resource of manually annotated protein complexes from mammalian organisms (http://mips.helmholtz-muenchen.de/genre/proj/corum/). The specialty of this database is that data are collected exclusively from publications in which PPI were validated and exclude data from high throughput experiments. As a consequence, the output of protein complexes is of high quality but seldom provides a uniform result as usually multiple complexes are reported for an individual protein while the process of manual annotation comes at the expense of actuality (Ruepp et al., 2008; Ruepp et al., 2010).

A more specialized database is the 'Biological General Repository for Interaction Datasets' (BioGRID) database and archives and disseminates genetic and protein interaction data from model organisms and humans (http://thebiogrid.org). The content of this database is exclusively extracted from data generated by either high throughput or low throughput associations and can be filtered accordingly. In contrast to the STRING resource, it contains far less entries but each entry is backed by experimental data rather then presenting predictions of likely interactors based on text mining approaches (Breitkreutz et al., 2008; Chatr-Aryamontri et al., 2013; Stark et al., 2011; Stark et al., 2006).

Another valuable tool is the 'Universal Protein Resource' (UniProt), and contains protein sequence and annotation data (http://www.uniprot.org). The database itself is not strictly designed as a resource for reporting individual PPI data but contains many informations of the protein itself, including protein function, subcellular localization, post-translational modification and its subunit structure where

information on established protein interactors can be found. UniProt is a consortium currently consisting of the European Bioinformatics Institute (EMBL-EBI), Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (PIR).
In order to evaluate whether detected interaction partners are new or unknown or why a protein that has been reported to always coprecipitate with a certain protein is at once is missing, all databases have to be consulted to get the best understanding of the particular situation.

5.3.2. PCNA

Proliferating cell nuclear antigen (PCNA) forms a homotrimer and acts as an auxiliary protein of DNA polymerase delta, thereby increasing the processivity of leading strand synthesis during DNA replication (Liu et al., 2003b). PPI analysis for PCNA resulted in the enrichment of proteins of components associated with the function of PCNA: Among those were members of the replication factor c, RFC1, RFC2, RFC3, RFC4 and RFC5, that acts as a clamp loader and catalyzes the loading of PCNA onto DNA (Bowman et al., 2004a), DNA-methyltransferase 1 (DNMT1), which has been reported to associate with DNA replication sites in S phase sustaining the methylation pattern in the newly synthesized strand and thereby maintains epigenetic inheritance (Doyon et al., 2004a; Frauer et al., 2011), RNA Polymerase II-Associated Factor 1 Homolog (PAF1/p15) that has previously been shown to associate with PCNA during DNA repair events (Simpson et al., 2006), members of the cleavage and polyadenylation specific factor complex (CPSF1, 3 and 4, Nagaike et al., 2011), the WD repeat domain 33 (WDR33), which is part of the pre-mRNA 3'processing complex (Chen and Scheller, 2001a) and the Wilms tumor 1-associated protein (WTAP), a regulator of the pre-mRNA splicing machinery (WTAP, Ortega et al., 2003). In the present study, functional assays were conducted in order to validate the characteristic proliferation-associated expression profile of PCNA (Liu et al., 2003a). Here, PCNA.GFP ceased to be expressed upon differentiation and PPI analysis only detected highly specific and enriched proteins in the subfraction of remaining proliferative lt-NES cells with sustained PCNA.GFP expression (section 4.1.6).

5.3.3. CDK2AP1

Cyclin-dependent kinase 2-associated protein 1 (CDK2AP1) is a homodimer and acts as a specific inhibitor of the monomeric unphosphorylated cell-cycle kinase CDK2 (Ertekin et al., 2012). Recently, CDK2AP1 was described as a bona fide subunit of the nucleosome remodeling and histone deacetylase (NuRD) complex (Spruijt et al., 2010). PPI analysis for CDK2AP1 in the present study confirms these results as all members of the NuRD complex, except for histone deacetylase 1 (HDAC1) and GATA zinc finger domain-containing protein 2A (GATAD2A), were detectable: Metastasis associated protein 1 (MTA1), MTA2, MTA3, methyl-CpG-binding domain protein 3 (MBD3), retinoblastoma-binding protein p48 (RBAP48), chromodomain-helicase-DNA-binding protein 4 (MI-2/CHD4), histone deacetylase 2 (HDAC2) and the histone H3.3A variant. The absence of HDAC1 could be linked to its role as transcriptional regulator, deacetylating the N-terminal part of the core histones (H2A, H2B, H3 and H4, Spruijt et al., 2010). Apart from these complex members, only breast cancer anti-estrogen resistance protein 1 (BCAR1), zinc finger protein 219 (ZNF219) and X-Ray Repair Complementing Defective Repair In Chinese Hamster Cells 6 (XRCC6, G22P1) were enriched in our PPI screens. While BCAR1 primarily is reported as an interaction partner of RBAP48 (Yarden and Brody, 1999), ZNF219 has been described to act as a transcriptional repressor and therefore may be involved in cooperation with the histone deacetylation activity of the NuRD complex (Sakai et al., 2003b). The enrichment G22P1 as a DNA-dependent ATP-dependent helicase may support a negative regulation of transcription together with APEX1, which was not detected in the present analysis (Reeves and Sthoeger, 1989b).

Upon differentiation, RBAP48, H3.3A, BCAR1 and G22P1 were no longer enriched in CDK2AP1 PPI analyses while GATAD2A and HDAC1 are appearing. In the context of a predominantly postmitotic cell population, the dissociation of the active histone variant H3.3A together with the recruitment of HDAC1 and GATAD2A, both contributing to an inactive chromatin and associated with processes of differentiation, is plausible (Brackertz et al., 2002). Since RBAP48 is also a core histone-binding subunit and associated with multiple complexes involved in transcriptional repression including the chromatin assembly factor 1 (CAF1) complex, the core histone deacetylase (HDAC) complex and the nucleosome remodeling factor (NURF) complex, a decay of the NuRD complex in differentiating cells might be attributable to

the necessity of developing neurons to express genes that were previously silenced (Zhang et al., 2000).

5.3.4. SETD1B

The histone-lysine N-methyltransferase SET domain containing 1B (SETD1B) as part of the complex proteins associated with set1 (COMPASS) complex specifically methylates the 'Lys-4' residue of histone H3, provided that the neighboring 'Lys-9' residue is not already methylated, thereby activating transcription (Lee et al., 2007; Mohan et al., 2010). In PPI analysis of lt-NES cells, all proteins of the COMPASS complex were detected except for host cell factor 1 (HCF1), WD repeat-containing protein 5 (WDR5) and WD repeat-containing protein 82 (WDR82): The histone-lysine N-methyltransferase SET domain containing 1A (SETD1A), the Set1/Ash2 histone methyltransferase complex subunit Ash2 (Absent, Small, Or Homeotic)-Like (Drosophila) (ASH2L), dpy-30 homolog (DPY30), retinoblastoma-binding protein 5 (RBBP5), CXXC Finger Protein 1 (CXXC1/CFP1) and histone H3.3 (H3.3A). The presence of an active form of histone H3.3A supports the concept of its existence within actively transcribed loci (Tagami et al., 2004b). Although HCF1 has been regarded as a member of the COMPASS complex, this protein was not identified within the complex, probably due to its role as a tethering compound between the transcription activating COMPASS and the repressing SIN3 transcription regulator family member A (Sin3a) complex (Denslow and Wade, 2007b). Both WDR5 and WDR82 are absent in lt-NES PPI analyses, which might be du to their initiating function within the COMPASS complex: WDR82 mediates the recruitment of the complex to specific sites (Lee and Skalnik, 2008a), while WDR5 is postulated to position the N-terminus of histone H3 for efficient trimethylation (Patel et al., 2009).
Upon differentiation, the active histone H3.3A disappeared from the complex as well as DPY30, a core component of several methyl transferase complexes (Wang et al., 2009a). Furthermore, SETD1A was not found in the complex anymore, underlining its non-redundant contribution to the epigenetic control of chromatin structure and gene expression (Lee and Skalnik, 2008b).

5.3.5. RUVBL2

RuvB-Like AAA ATPase 2 (RUVBL2) has a single-stranded DNA-dependent ATPase and 5’ to 3’ ATP-enabled DNA helicase activity (Doyon et al., 2004b). RUVBL2 is a member of multiple protein complexes including the Nu4A (histone acetylation), the

Snf2-Related CREBBP Activator Protein (SRCAP, exchange of H2AZ/H2B dimers with H2A/H2B, Auger et al., 2008) the unconventional prefoldin RPB5 interactor 1 (URI, TOR kinase dependent transcription program, Gstaiger et al., 2003) and the INO complex (chromatin remodeling, Aoyama et al., 2008). In PPI analysis of lt-NES cells, representatives from each of these complexes were detected: E1A Binding Protein P400 (EP400), transformation/transcription domain-associated protein (TRRAP), DNA methyltransferase 1-associated protein 1 (DMAP1), the helicase SRCAP, actin-like protein 6A (ACTL6A, BAF53), RuvB-like AAA ATPase 1 (RUVBL1), actin-related protein 5 (ACTR5), INO80 complex subunit C (INO80C), the DNA helicase INO80 (INO80), Actin-related protein 8 (INO80N), URI and prefoldin subunit 2 (PFDN2).

Upon differentiation, proteins typically associated with proliferation, i.e. members of the INO80 complex, INO80A and INO80N, both involved in DNA replication (Conaway and Conaway, 2009b), a member of the NuA4 complex associated with normal cell cycle progression, TRRAP (Doyon et al., 2004a; Sardiu et al., 2008b), the proliferation associated ACTR5 as a member of the SRCAP complex and PFDN2 as a growth-dependent compound of the URI complex (Jeronimo et al., 2007b), dissociated from the RUVBL2-interacting complexes. Furthermore, BAF53 also was not enriched anymore for RUVBL2 in neuronal cultures. BAF53 is a multifunctional protein present in a multitude of different complexes and is required for the proliferation of neural progenitors. During neural differentiation, which was modeled by withdrawing growth factors from the lt-NES cell cultures, a switch from a neural stem cell population to post-mitotic neurons requires post-mitotic chromatin remodeling. This process is conducted by the neuronal-specific BAF (nBAF) complex with its BAF53 homologue ACTL6B/BAF53B, a protein absent in our proliferative lt-NES cells that emerged only in neuronal cultures (Li and Jin, 2010; Yoo et al., 2011a). In addition, ribophorin 1/2 (RPN1/2) as compounds of the oligosaccharyl transferase (OST) complex and enhancer of polycomb homolog 2 (Drosophila) (EPC2) could not be detected as RUVBL2 interacting candidates in differentiated neuronal cultures.

Taken together, BAC-based PPI analysis in lt-NES cells and their neuronal progeny is sufficiently sensitive and specific to observe differentiation-specific changes in protein complexes/interactions associated with epigenetic imprinting, gene regulation

through histone acetylation/methylation and nucleosomal rearrangement. The observations presented here encompass dynamic changes, which have not yet been reported and may thus contribute to a deeper insight into human-specific PPIs occurring during the differentiation of neural stem cells into postmitotic neurons. The presented method is sufficiently streamlined to perform medium-throughput analyses of PPIs while retaining the sensitivity of detecting second-order or indirect interactors and the specificity of an approach using endogenous protein levels. At the time of composing this thesis, no study aimed at the analysis of PPIs generated postmitotic somatic cells without the application of chemotherapeutics or compromising conditions. The present study combines both highly sophisticated methods to obtain physiological expression of the bait protein and state-of-the-art analysis of interacting proteins by AP-MS/MS with a unique somatic cell population. Lt-NES cells are well suited for this approach as they are amenable to confident genetic manipulation, stable proliferation and exhibit a highly neurogenic potential. Thus, this work presents an unprecedented way to investigate PPIs of a broad range of proteins in a physiological context.

5.4. Limitations of BAC-mediated PPI analysis

Despite many advantages BAC-mediated protein expression in the context of PPI analysis exhibits, this system still suffers from some disadvantages including a large tag size affecting protein folding and binding (I), unpredictable integration sites (II), occasionally multiple integrations of the construct (III), the possibility of non-tagged protein isoforms (IV) and the restriction to proteins already expressed in lt-NES cells, when C-terminal tags are used, since in this case the expression of the chemo-selection resistance gene is linked with the expression of the gene of interest (V).

I) The size of the protein tag

Although GFP is a versatile marker for live cell monitoring of protein localization - as shown exemplarily for AURKA.GFP expression in this study - its size may limit its usefulness as a protein tag. As protein complexes typically consist of a large and - depending on their actual physiological role – fluctuant number of proteins, the inclusion of some members within this complex or the direct interaction with the tagged protein itself may be sterically hindered by the presence of the tag. Additionally, the tag itself may recruit other proteins normally avoiding contact with the complex. During data evaluation, compensation for unspecific binding proteins was performed but displacement due to the tag may permanently prevent detection of bona fide interactors. Furthermore, the GFP tag may affect the stability of the protein and prolong or delay its half-life period.

II) Locus of integration

By using an electroporation-based approach to generate BAC-transgenic cell lines, the position of integration on the genome cannot be predicted as the integration process is predominantly mediated by non-homologous end joining (NHEJ) and therefore may increase the risk for disruption of a tumor-suppressing gen or altering the fate of a cell permanently, both potentially leading to a re-modulation of protein expression patterns in a cell. Derived PPI data therefore may lead to wrong conclusions. Associated with a random integration is an unknown chromatin status of the BAC environment on the genome, a phenomenon previously described as a chromatin positional effect. However, due to the size of BAC-based constructs, they normally encompass most, if not all cis-regulatory elements and therefore are less likely to be influenced by chromatin positional effects. In rare cases, regulation can still take place *in trans* or by epigenetic regulation that might cover chromosomal

regions larger than a BAC construct (Blaas et al., 2012). Therefore an authentic endogenous regulation cannot be assumed.

III) Integration frequency

Although the integration frequency was monitored for several lines by fluorescence in situ hybridization (FISH) in this study, a valid definition guaranteeing BAC integration at only one locus cannot be made. In analyses where only one additional locus was detected, investigations using the fluorescence signal intensity per spot to evaluate the DNA amount per locus in relation to endogenous intensities showed an average integration rate of one BAC copy per locus (lines AURKA.GFP, Jun.GFP) with the exception of PCNA.GFP, featuring two copies per integration. These observations correspond to previous studies, reporting that integration frequencies generally are low but multiple integrations per locus may occur (Sparwasser and Eberl, 2007). Multiple integrations of a BAC within a cell expressing a tagged protein most likely result in a low overexpression if not compensated by regulatory mechanisms. Additionally, the protein itself will be present as tagged and non-tagged variant, originating from three different loci of the genome.

IV) Dependency on splice variants

One major limitation is based on the GFP tag localization, restricting the composition of tagged protein variants to those containing the first CDS-coding exon (for N-terminal tags) or the last CDS-coding exon (C-terminal tags). Hereby, shorter variants or alternatively spliced proteins lacking the first or last exon, respectively, are not subject to investigations.

V) Limitations by the choice of the selection cassette

In principle, two different systems for expression of the chemoselection gene were available from the BACE database (http://hymanlab.mpi-cbg.de:81/bac/) and each showed individual advantages and limitations. BAC constructs harboring an N-terminal GFP tag consisted of an artificial intron comprising an independent, constitutive active small promoter, phosphoglycerate kinase (PGK), facilitating a strong expression of the Neomycin resistance gene (Neo^R). These constructs allow the chemoselection of cells that do not exhibit active expression of the gene introduced by BAC constructs but simultaneously can lead to resistant cells with incompletely integrated constructs, lacking the expression of tagged proteins or featuring an incorrect regulation. Therefore, the integrity of the construct has to be tightly monitored as truncated BAC expression constructs have been described

before (Chandler et al., 2007). In BAC constructs containing a C-terminal protein tag, the expression of the Neo^R gene is directly linked to the regulatory elements of the introduced gene itself via an internal ribosomal entry site (IRES) sequence. Hereby, the enrichment of transgenic cells is limited to genes that are actively expressed by the desired cell population but usually give rise to cell lines containing the complete gene with most of its regulatory elements. Therefore, it was impossible to generate BAC transgenic cell lines of C-terminally tagged genes expressed at very low endogenous levels or genes exclusively present in fully differentiated and postmitotic cells.

Taken together, aspects concerning tag design, integration site and -frequency may be of individually pronounced significance in the outcomes of PPI analysis. However, this approach facilitates a highly sensitive as well as specific PPI assessment most closely resembling an endogenous setting that is still fit for a streamlined approach to screen a broad spectrum of disease-associated proteins within a somatic cell population.

5.5. Alternative approaches for PPI studies using epitope-tagged proteins

Most limitations of a BAC-based expression system concerning tag size, integration site and frequency can be circumvented by the application of systems facilitating the modification of endogenous genes in order to express epitope-tagged proteins. One approach is the utilization of AAV-mediated site-directed manipulation to equip endogenous genes with a small epitope-tag. This approach not only avoids the introduction of an additional gene copy but also results in a smaller tag size, less likely to influence binding affinities of prospective interactors. Complementary studies using authentic endogenously tagged proteins utilizing the AAV technique should be able to tackle these problems and potentially provide benefits over BAC-mediated protein tagging. However, establishment, validation and employment of large numbers of individually engineered AAV vectors not only is impractical but clearly unsuitable for streamlined applications and therefore can only serve as a tool for the investigation of individual proteins addressing particular issues. Alternatively, transcription activator-like effector nucleases (TALEN) represent a versatile tool that can quickly be engineered to bind practically any DNA sequence and induce precisely located double strand breaks. Exploiting the endogenous non-homologous

end joining (NHEJ) machinery or homology directed repair, foreign DNA can be introduced to these sites when present as exogenous double-stranded DNA fragments (Hockemeyer et al., 2011). Another approach relying on the targeted introduction of double strand breaks is the newly established technique of genome editing that uses the RNA-guided Cas family of nucleases based on the clustered regularly interspaced short palindromic repeats (CRISPR) type II system. Here, the precise targeting of genomic sites is guided by target-sequence containing RNAs without the need of generating sequence specific binding proteins. This technique facilitates either the disruption of genomic sites or the integration of foreign DNA sequences using the same cellular repair mechanisms mentioned in context of the TALEN technique (Chang et al., 2013; Cho et al., 2013). Due to simultaneous publicizing of a genome-spanning database together with the CRISPR-based genome editing approach, the generation of suitable targeting constructs is now available to the whole scientific community (Mali et al., 2013). Refined techniques for the validation of correctly targeted sites like the ‘Advanced Analytical’s Fragment Analyzer’ and DNAWorks software specifically composed for the design of necessary oligonucleotides should enable the setup of a streamlined system fit to reliably target arbitrary proteins in a proteome-wide approach. However, a library of constructs containing a donor DNA consisting of the tag flanked by sequences homologue to the targeted site would be necessary as well. The ongoing development of innovative techniques and critical scrutinizing of present data will aid to advance the field of protein interaction analysis to a grade that may be one day at eye level with our current comprehension of the human genome.

5.6. Outlook

Within the framework of the present work, comprehensive systems and methods were established that on the one hand necessitate further validation studies of the generated data and on the other hand give rise to novel technical opportunities. In the following sections, several aspects will be discussed in detail.

Validation of protein-protein interactions

The streamlined approach for the assessment of PPI presented in this study is well suitable for the fast generation of multiple interactomes for several proteins. However, for a detailed analysis of specific interactions that include parameters not measurable by QUBIC or to validate interaction partners, the following approaches are conceivable.

I) In order to validate the interaction of a certain protein, reciprocal PPI studies can be conducted, as demonstrated previously. Here, the availability of QUBIC and an almost genome-spanning library of tagged proteins enables the examination whether a former prey can be used as bait to enrich the former bait. This method illustrates a straightforward and powerful approach that has already been used by a number of studies (Hubner et al., 2010a; Spruijt et al., 2010).

II) In order to address the impact of individual proteins for the complex composition and its stability, loss-of-function studies like RNAi or novel genome editing technologies like the CRISPR-Cas9-system can be used as a tool to deplete specific members of a complex (Carlin et al., 2011; Chang et al., 2013; Krastev et al., 2011).

III) The combination of RNAi and BACs containing RNAi-site-specific silent mutations or BAC-mediated reconstitution of a cellular protein deficiency can further enhance the sensitivity of measurements by depleting endogenous proteins that are competing for binding partners with the tagged variants. Hereby, the proportion of tagged proteins within a complex can be increased and thus contribute to a better understanding of the protein function.

IV) The combination of RNAi and BACs containing a silent and an affecting mutation can be employed to analyze the impact of certain mutations on particular binding sites for interaction partners, leading to a deeper understanding of the involved domains.

Comparative analysis of complex composition in neuronal subtypes

The present study primarily uses human neural stem cells and their GABAergic neuronal derivatives for the analysis of PPI and therefore allows for the analysis of complex composition changes during differentiation for a limited number of proteins. However, it would be conceivable to extend PPI studies to neuronal subtypes such as dopaminergic neurons or motor neurons in order to investigate the PPI network of long-discussed disease-associated proteins or key regulators of developmental processes directly in the *in vitro* correlate of the *in vivo* cell type. It can be assumed that postmitotic human neurons have a highly specific proteome and therefore it is of particular interest, whether PPI exist that are predominantly found in neuronal subtypes preferentially affected by specific neurodegenerative diseases like Parkinson's disease, amyotrophic lateral sclerosis or Alzheimer's disease. Direct comparison of the interactome of lt-NESCs and the various neuronal subpopulations might then reveal neuronal subtype-associated differences in the proteome.

Furthermore, the advent of iPS technology enables the generation and manipulation of homogenous somatic cell types, which will enable the investigation of PPIs in a diseased environment. This approach is of particular interest for neurodegenerative diseases with unknown genetic cause as the assessment of PPI for proteins that might be relevant for disease progression might enable a better characterization and understanding of such cases.

Further applications

Most techniques that were initially developed to overcome a certain problem can be used to address a broader range of topics. For example, BAC transgenesis facilitates the real-time observation of proteins under reasonably physiological conditions, rendering the following exemplary topics attractive applications in context of cellular models of the human CNS.

Reporter constructs for lineage selection

The fact that BAC-based constructs exhibit almost complete regulatory elements, facilitates their application as reporter constructs as these constructs can be regulated almost identically to their endogenous counterparts. In these settings, cells with a specific fate can be identified either by expression of a visually detectable reporter gene like GFP or by expression of a resistance gene. Following identification of suitable reporter genes e.g. HB9 (motoneurons) or Pitx3 (midbrain dopaminergic

neurons), the generation of reporters suitable for the enrichment of specific neuronal subtypes is a straightforward task. Previous studies used a BAC-based technology for the genetic identification of defined hESC-derived neural cell types and thereby established a method for the otherwise difficult enrichment of human motor neurons with proper marker expression and electrophysiological activity (Placantonakis et al., 2009). Another study established a protocol for transposon-mediated integration of BAC-based reporters for studies during development and differentiation in hESCs (Rostovskaya et al., 2012). Except for issues related with poor fluorescence intensity by low expression of such constructs or an altered protein stability caused by the tag, systems of this sort should be suitable for most applications and might facilitate the development of improved differentiation protocols.

Reporter constructs for cellular stress

With the development of ever more sophisticated analytic algorithms, complex assays for the identification of different stressors can be adapted for high-content data analyses by application of BAC transgenomics (Wink et al., 2014). Such constructs can be used as reporters on the genetic level (luciferase reporters) or on the protein level (fluorophore coupled proteins), detecting a broad range of cellular stress signals including oxidative stress (Keap1, Nrf2, Srxn1), endoplasmic reticulum stress (ATF4/XBP1, CHOP/BiP), inflammation (IκBα, RelA, ICAM1), DNA damage (53BP1, p53, p21), heat (HSF-1/2, HSP70/90), heavy metal (MTF-1, MTs) and hypoxia (VHL, HIF-1α, HUMMR, Wink et al., 2014). The high content evaluation of images generated using such assays enables the characterization of cellular stress responses that typically follow cell perturbations at the subcellular organelle level (Chan et al., 2012). However, these analyses are absolutely dependent on good fluorescence signals and therefore may be unsuitable for proteins expressed at very low levels as hereby artifacts of long-exposed images may lead to false-positive results. A particularly interesting feature is the possibility to mutate genes on BAC-constructs for disease modeling. This is particular suitable for applications where genetically-linked protein aggregation is investigated: Instead of using common coaggregating proteins like firefly luciferase mutants as sensors of proteome stress (Gupta et al., 2011), the protein of interest can directly be monitored. However, each system has to be precisely monitored as fluorescent labeling of proteins was repeatedly reported to impair protein functionality (Brieger et al., 2012; Meyer et al., 2007).

6. Abbreviations

Abbreviation	Full name
AA	Ascorbic acid
AAV	Adeno-associated virus
ACTL6A	Actin-like 6A/BAF53
ACTL6B	Actin-like 6B
ACTR5	ARP5 actin-related protein 5 homolog
ALS	Amyotrophic lateral sclerosis
APH1	Anterior pharynx defective 1 homolog
ATP	Adenosine triphosphate
Atxn3	Ataxin-3
AP-MS	Affinity purification coupled mass spectrometry
AURKA	Aurora kinase A
BAC	Bacterial artificial chromosome
BACE1	β-Site APP cleaving enzyme 1
BAF53	BRG1-associated factor 53A
BB	Brainbow
BMP	Bone morphogenic protein
BrdU	Bromdesoxyuridin
Brn2	POU class 3 homeobox 2
BSA	Bovine serum albumin
CA	Cornu ammonis
CAS	CRK-associated substrate
cAMP	Cyclic adenosine monophosphate
CD	Cluster of differentiation
CDK2AP1	Cyclin-dependent kinase 2 associated protein 1
CFP	Cyano fluorescent protein
CHD4	Chromodomain helicase DNA binding protein 4
CKAP4	Cytoskeleton-associated protein 4
CMV	Cytomegalie virus
CNS	Central nervous system
CORO1C	Coronin, actin binding protein, 1C
Cre	Cre recombinase derived from the P1 bacteriophage
CSE1L	CSE1 chromosome segregation 1-like
CSTF3	Cleavage stimulation factor, 3' pre-RNA, subunit 3, 77kDa
CXXC1	PHD finger and CXXC domain-containing protein 1
Dach1	Dachshund homolog 1
DAPI	4',6-Diamidino-2-phenylindole
DAPT	N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester

DMSO	Dimethyl suphoxide
DNA	Deoxyribonucleic acid
DNMT1	DNA-methyltransferase 1
CPP	Cell permeant proteins
cPPT	Central polypurine tract
CPSF	Cleavage and polyadenylation specific factor
dNTPs	Nucleoside triphosphate
Dox	Doxycycline
dsRNA	Double stranded RNA
EB	Embryoid body
EDTA	Ethylenediaminetetraacetic acid
EF1α	Elongation factor α
EGF	Epidermal growth factor
ELISA	Enzyme-linked Immunosorbent Assay
EP400	E1A binding protein p400
EPC2	Enhancer of polycomb homolog 2
ER	Endoplasmatic reticulum
Ex	Exon
FACS	Fluorescence activated cell sorting
FCS	Fetal calf serum
FGD1	FYVE, RhoGEF and PH domain containing 1
FGF2	Fibroblast growth factor 2
FGF8b	Fibroblast growth factor 8b
FISH	Fluorescence in situ hybridization
FLAG	Minimal protein tag
FoxA2	Forkhead box A2
FoxG1	Forkhead box G1
FSC	Forward scatter
GABA	Gamma-aminobutyric acid
GATAD2A	GATA zinc finger domain containing 2A
GFAP	Glial fibrillary acidic protein
GFP	Green fluorescent protein
GMP	Good manufacturing practice
GSK-3α/β	Glycogen synthase kinase 3 α/β
H2A	Histone H2A
H2AZ	H2A histone family, member Z
H2B	Histone H2B
H3K4	Histone H3K4
HDAC	Histone deacetylase
HEK cell	Human embryonic kidney

hES cell	Human embryonic stem cell
hNF	Human neurofilament
hPS cell	Human pluripotent stem cell
HSPC231	PFDN2, Prefoldin 2
IFNα	Interferon α
IFNβ	Interferon β
IKKα	Nuclear factor NF-kappa-B inhibitor kinase alpha
IKKβ	Nuclear factor NF-kappa-B inhibitor kinase beta
IκBα	NF-kappa-B inhibitor alpha
In	Intron
INO80	Putative DNA helicase INO80 complex homolog 1 / INO80A
INO80N	ARP8 actin-related protein 8 homolog
IP10	Interferon-inducible protein 10
iPS cell	Induced pluripotent stem cell
IRES	Internal ribosomal entry site
IRF3	Interferon regulatory factor 3
ISGF3	Interferon regulatory factor 9
JARID1C	Jumonji, AT rich interactive domain 1C
JUN	Jun proto-oncogene
LC	Liquid-Chromatography
LC-MS/MS	Liquid-Chromatography-Mass Spectrometry/Mass Spectrometry
LIF	Leukemia inhibitory factor
Ln	Laminin
loxP	Recognition site of Cre recombinase
lt-NES cell	long-term self-renewing neuroepithelial stem cell
MAP	Microtubuli-associated protein
MAP1LC3B	Microtubule-associated protein 1 light chain 3 beta
MAPK	MAP kinase
MAPT	Microtubuli-associated protein tau
MARK2	MAP/microtubule affinity-regulating kinase 2
Mash1	Achaete-scute homologue ash1
MAVS	Mitochondrial antiviral signaling protein
MBD3	Methyl-CpG binding domain protein 3
mC1-mRNA	Modified C1-capped mRNA
MDA5	Melanoma differentiation-associated protein 5
MECP2	Methyl CpG binding protein 2
MEF	Mouse embryonic fibroblast
MG	Matrigel
MGA	MAX gene associated
MJD	Machado-Joseph disease

MOI	Multiplicity of infection
mRFP	Monomeric red fluorescent protein
miR	micro RNA
mmRNA	Modified mRNA
mRNA	messenger RNA
MTA3	Metastasis associated 1 family, member 3
N-PAC	Nuclear protein of 60 kDa
nBAF	Neuronal BRG1-associated factor
NCL	Nucleolin
NCSTN	Nicastrin
neoR	Neomycin resistance
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
Ngn2	Neurogenin 2
NLS	Nuclear localization signal
NPM1	Nucleophosmin
NuA4	NuA4 histone acetyltransferase complex
NuRD	Nucleosome remodeling and histone deacetylase
Olig2	Oligodendrocyte transcription factor 2
Otx2	Orthodenticle homeobox 2
OST	Oligosaccharyltransferase
P	PreScission cleavage site
p50	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1
p65	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 3
pA	Poly adenylation
PAGE	Polyacrylamid gel electrophoresis
Pax6	Paired box 6
PAF1	RNA polymerase II-associated factor 1 homolog
PBS	Phosphate buffered saline
pCAG	Poly glutamine
PCR	Polymerase chain reaction
PDRG1	p53 and DNA-damage regulated 1
PFA	Paraformaldehyde
PGK	Phosphoglycerate kinase
PLZF	Promyelocytic leukemia-associated zinc finger
PPI	Protein-protein interaction
PRPF40a	PRP40 pre-mRNA processing factor 40 homolog A
PSEN1	Presenilin 1
PSEN2	Presenilin 2
puroR	Puromycin restance
QAP	Quantitative affinity purification

QUBIC	Quantitative BAC InteraCtomics
Rag2	Recombination activating gene 2
RBBP	Retinoblastoma-binding protein
RBPJ	Recombination signal binding protein for immunoglobulin kappa J region
RIG-I	Retinoic acid-inducible gene 1 protein
RNA	Ribonucleic acid
RNAi	RNA interference
RFC	Replication factor C
RPN (1, 2)	Ribophorin (1, 2)
RPS27a	Ribosomal protein S27a
RT-PCR	Reverse-transcriptase PCR
RRE	Rev responsive element
rtTA	reverse tetracycline-controlled transactivator
RuvBL (1, 2)	RuvB (E coli homolog)-like (1, 2)
RV	Rabies virus
S	S-peptide
SA	Splice acceptor
SD	Splice donor
SDS	Sodium dodecyl sulphate
SEPT	Synthetic exon promoter trap
SetD1A	Set1/Ash2 histone methyltransferase complex subunit SET1
SetD1B	SET domain-containing protein 1B
SHH	Sonic Hedgehog
SILAC	Stable isotope labeling by amino acids in cell culture
SIN3A	SIN3 transcription regulator homolog A...
siRNA	Short-interference RNA
SMAD	Mothers against decapentaplegic homolog
SMCHD1	Structural maintenance of chromosomes flexible hinge domain containing 1
SNP	Single-nucleotide polymorphism
Sox2	SRY-box 2
SPATS2	Spermatogenesis associated, serine-rich 2
SRCAP	Snf2-related CREBBP activator protein
SSC	Side scatter
SVZ	Subventricular zone
T	TEV cleavage site
TAB1	TGF-beta activated kinase 1/MAP3K7 binding protein 1
TAB2	TGF-beta activated kinase 1/MAP3K7 binding protein 2
TAB3	TGF-beta activated kinase 1/MAP3K7 binding protein 3
TAK1	TGF-beta-activated kinase 1
TALEN	Transcription activator-like effector nucleases

TANK	TRAF family member-associated NFKB activator
TBK1	TANK-binding kinase 1
TC	Tissue culture
TE	Trypsin/EDTA
TEMED	Tetramethylethylenediamine
tetOn	Tetracycline regulatable gene induction system
TF	Transcription factor
TGFβ	Transforming growth factor β
TH	Tyrosine hydroxylase
TI	Trypsin inhibitor
TLR3	Toll-like receptor 3
TMD	Transmembrane domain
TMEM132A	Transmembrane protein 132A
TOR	Target of rapamycin
TRAF3	TNF receptor-associated factor 3
TRAF6	TNF receptor-associated factor 6
TREtight	Tetracyclin-responsive element
TRRAP	Transformation/transcription domain-associated protein
TTC	Tetanus toxin fragment C
Txl	Taxol
UbC	Ubiquitin C
UHRF1	Ubiquitin-like with PHD and ring finger domains 1
URI	Unconventional prefoldin RPB5 Interactor
VCP	Valosin containing protein
WGA	Wheat germ agglutinin
WPRE	Regulatory element of woodchuck hepatitis virus
WTAP	Wilms tumor 1-associated protein
Y2H	Yeast two hybrid
YFP	Yellow fluorescent protein
Zic1	Zinc finger protein 1
ZNF	Zinc finger
ZO1	Zona occludens 1

7. Summary

While most approaches in cell-based disease modeling are focused on the effects of defined mutations on the molecular or cellular phenotype, the assessment of underlying alterations in the interactomes of disease-relevant proteins has faced several technical challenges. First, experiments were typically conducted using overexpression paradigms resulting in unphysiologically high protein levels and thus promoting unspecific interactions. Second, such studies have been relying mostly on transformed cell lines, which enable mass production of transgenic cells but do not exhibit a tissue-specific proteomic environment. For that reason, the present study aimed at addressing these issues by bacterial artificial chromosome (BAC)-based expression of tagged proteins in pluripotent stem cell-derived long-term neuroepithelial like stem cells (lt-NES cells), a stable and robust cell population, which generates authentic human neurons with high fidelity. Tagged proteins were found to be expressed at endogenous levels, and fluorescence in situ hybridisation (FISH) analyses revealed an average integration rate of one copy per genome for the majority of cell lines analyzed. Correct compartmentalization and size of the tagged proteins could be confirmed by high-resolution confocal and live cell imaging as well as Western immunoblotting analysis, respectively. Employing this approach, multiple cell lines were generated harboring tagged proteins associated with human developmental disorders, cancer and neurodegeneration. Representatives of these groups include Proliferating Cell Nuclear Antigen (PCNA), Aurora Kinase A (AURKA), Cyclin-Dependent Kinase 2-Associated Protein 1 (CDK2AP1), Set Domain-Containing Protein 1B (SETD1B), RuvB-Like 2 (RUVBL2), the Methyl CpG Binding Protein 2 (MECP2) and the Alzheimer's disease-associated proteins Nicastrin (NCSTN) and Valosin-Containing Protein (VCP). Using a label-free, quantitative affinity purification-mass spectrometry approach, numerous novel interaction partner candidates of these proteins were identified. Direct comparison of protein-specific interactomes of proliferating lt-NES cells and their neuronal progeny further revealed changes in the composition of several chromatin-remodeling complexes, suggesting that this system is sufficiently sensitive and specific to identify the dynamic differential recruitment of individual proteins as a response to developmental switches. In a proof-of-concept study, the approach of BAC-mediated expression of tagged proteins with a subsequent analysis of interacting proteins was successfully transferred to induced pluripotent stem cell (iPS)-derived lt-NES cells in order to enable PPI analyses in the context of complex diseases in following studies.

Finally, an adeno-associated virus based approach for epitope tagging of endogenous genes in iPS-derived lt-NES cells from a patient suffering from Machado-Joseph disease allowed the generation of cell pools exhibiting both the diseased and healthy isoform of N-terminal FLAG-tagged Ataxin-3. The present work demonstrates a successful establishment of two different methods for protein tagging in somatic cell populations that subsequently can be employed for a multitude of analytical techniques including fluorescent microscopic visualization of protein localization, dynamics of protein recruitment or the detection of PPI.

8. Zusammenfassung

Während zellbasierte Ansätze der Krankheitsmodellierung vornehmlich auf die Auswirkungen definierter Mutationen des molekularen sowie zellulären Phänotyp fokussiert sind, müssen bei der Untersuchung zugrunde liegender Veränderungen im Interaktom von krankheitsrelevanten Proteine eine Reihe technischer Herausforderungen berücksichtigt werden. Einerseits werden Experimente typischerweise unter Überexpressions-Paradigmen durchgeführt, in deren Konsequenz durch unphysiologisch hohe Protein-Niveaus unspezifische Interaktionen gefördert werden. Andererseits wurden solche Studien meist unter Verwendung transformierter Zelllinien durchgeführt, was zwar eine einfache Bereitstellung großer Mengen transgenen Zellmaterials ermöglicht, jedoch in keinem gewebsspezifischen Kontext stattfindet. Vor diesem Hintergrund war es das Ziel der vorliegenden Arbeit, die auf bakteriellen artifiziellen Chromosomen basierende Expression markierter Proteinen in von pluripotenten Stammzellen abgeleiteten neuroepithelialen Stammzellen (lt-NES Zellen) zu etablieren. Diese stabile und robuste Zellpopulation ermöglicht es, authentische humane Neuronen in hoher Effizienz zu generieren. Hierbei konnte eine Expression der markierten Proteine auf endogenem Niveau beobachtet werden während gleichzeitig lediglich eine BAC-Integration von einer Kopie pro Genom für die meisten etablierten Linien durch Fluoreszenz-*in-situ*-Hybridisierungsstudien (FISH) nachgewiesen wurde. Eine korrekte Größe sowie subzelluläre Lokalisation konnten anhand von Western-Blot-Analysen sowie mit hochauflösender konfokaler und Lebendzellmikroskopie gezeigt werden. Durch Anwendung dieser Methodik konnten mehrere Zelllinien erstellt werden, welche markierte Proteine exprimieren, die eine veränderte Expression in humanen entwicklungsassoziierten Erkrankungen, Tumorerkrankungen oder neurodegenerativen Erkrankungen aufweisen. Vertreter dieser Gruppen umfassen das Proliferating Cell Nuclear Antigen (PCNA), Aurora Kinase A (AURKA), Cyclin-Dependent Kinase 2-Associated Protein 1 (CDK2AP1), Set Domain-Containing Protein 1B (SETD1B), RuvB-Like 2 (RUVBL2), das Methyl CpG Binding Protein 2 (MECP2) und Alzheimer-assoziierten Proteine Nicastrin (NCSTN) und Valosin-Containing Protein (VCP). Durch das Verwenden einer Label-freien, quantitativen affinitätschromatographischen Anreicherung mit nachfolgender massenspektrometrischer Analyse konnte eine Vielzahl neuer möglicher Protein-Interaktionen identifiziert werden. Ein direkter Vergleich von proteinspezifischen Interaktomen von proliferierenden lt-NES Zellen und von diesen abgeleiteten Neuronen konnten zudem Veränderungen in der Zusammensetzung von einigen Chromatin-Remodellierungskomplexen zeigen. Dies deutet darauf hin, dass das verwendete System sensitiv und spezifisch genug ist, um die dynamische Rekrutierung von einzelnen Proteinen als Auswirkung von Entwicklungs-assoziierten Prozessen zu detektieren. In einer Machbarkeitsstudie wurde die etablierte Methodik auf von iPS-Zellen abgeleiteten lt-NES

Zellen übertragen, um so eine Basis für die Analyse von Protein-Protein-Interaktionen im Kontext krankheits- und patientenspezifischer Zellen zu etablieren. Um Probleme, die bei der zufälligen Integration eines BACs entstehen können, zu umgehen, wurde schließlich ein Adeno-assoziiertes virales System zur Epitopmarkierung endogener Gene etabliert. Es wurden exemplarisch von an Machado-Joseph Krankheit erkrankten Patienten abgeleitete It-NES Zellen generiert, welche jeweils eine N-terminal markierte krankhafte sowie eine gesunde Isoform des Proteins Ataxin-3 aufweisen.

Die vorliegende Arbeit zeigt eine erfolgreiche Etablierung verschiedener Methodiken um Proteinmarkierungen in somatische Zellpopulationen einzubringen, welche sich durch eine weitgehend endogene Expression auszeichnen. Hierdurch ergibt sich die Möglichkeit, unter anderem Interaktionsstudien mit hoher Spezifität und Sensitivität an humanen gewebe- und krankheitsspezifischen Zellpopulationen durchzuführen als auch analytische Methoden wie fluoreszenzmikroskopische Lokalisierungsstudien oder Dynamiken der Proteinrekrutierungen anzuwenden.

9. References

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

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nature genetics *23*, 185-188.

Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. Science (New York, NY) *181*, 223-230.

Anokye-Danso, F., Trivedi, C., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P., Epstein, J.*, et al.* (2011). Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency. Cell stem cell *8*, 376-388.

Aoyama, N., Oka, A., Kitayama, K., Kurumizaka, H., and Harata, M. (2008). The actin-related protein hArp8 accumulates on the mitotic chromosomes and functions in chromosome alignment. In Exp Cell Res, pp. 859-868.

Auger, A., Galarneau, L., Altaf, M., Nourani, A., Doyon, Y., Utley, R.T., Cronier, D., Allard, S., and Côté, J. (2008). Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants. Molecular and cellular biology *28*, 2257-2270.

Baddour, J.A., Sousounis, K., and Tsonis, P.A. (2012). Organ repair and regeneration: an overview. Birth defects research Part C, Embryo today : reviews *96*, 1-29.

Barberi, T., Willis, L.M., Socci, N., and Studer, L. (2005). Derivation of multipotent mesenchymal precursors from human embryonic stem cells. PLoS medicine *2*, e161.

Benschop, J.J., Brabers, N., van Leenen, D., Bakker, L.V., van Deutekom, H.W., van Berkum, N.L., Apweiler, E., Lijnzaad, P., Holstege, F.C., and Kemmeren, P. (2010). A consensus of core protein complex compositions for Saccharomyces cerevisiae. Molecular cell *38*, 916-928.

Blaas, L., Musteanu, M., Grabner, B., Eferl, R., Bauer, A., and Casanova, E. (2012). The use of bacterial artificial chromosomes for recombinant protein production in mammalian cell lines. Methods in molecular biology (Clifton, NJ) *824*, 581-593.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annual review of biochemistry *72*, 291-336.

Blömer, U., Naldini, L., Kafri, T., Trono, D., Verma, I.M., and Gage, F.H. (1997). Highly efficient and sustained gene transfer in adult neurons with a lentivirus vector. Journal of virology *71*, 6641-6649.

Boehnke, K., Falkowska-Hansen, B., Stark, H.J., and Boukamp, P. (2012). Stem cells of the human epidermis and their niche: composition and function in epidermal regeneration and carcinogenesis. Carcinogenesis *33*, 1247-1258.

Bongso, A., Fong, C.Y., Ng, S.C., and Ratnam, S. (1994). Isolation and culture of inner cell mass cells from human blastocysts. Human reproduction (Oxford, England) *9*, 2110-2117.

Bowman, G.D., O'Donnell, M., and Kuriyan, J. (2004a). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. In Nature, pp. 724-730.

Bowman, G.D., O'Donnell, M., and Kuriyan, J. (2004b). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature *429*, 724-730.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G.*, et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Brackertz, M., Boeke, J., Zhang, R., and Renkawitz, R. (2002). Two highly related p66 proteins comprise a new family of potent transcriptional repressors interacting with MBD2 and MBD3. The Journal of biological chemistry *277*, 40958-40966.

Breitkreutz, B.J., Stark, C., Reguly, T., Boucher, L., Breitkreutz, A., Livstone, M., Oughtred, R., Lackner, D.H., Bähler, J., Wood, V.*, et al.* (2008). The BioGRID Interaction Database: 2008 update. Nucleic Acids Research *36*, D637-640.

Brieger, A., Plotz, G., Hinrichsen, I., Passmann, S., Adam, R., and Zeuzem, S. (2012). C-terminal fluorescent labeling impairs functionality of DNA mismatch repair proteins. PloS one *7*, e31863.

Bronson, S.K., Plaehn, E.G., Kluckman, K.D., Hagaman, J.R., Maeda, N., and Smithies, O. (1996). Single-copy transgenic mice with chosen-site integration. Proceedings of the National Academy of Sciences of the United States of America *93*, 9067-9072.

C. Harris, D. (2010). Quantitative Chemical Analysis. 750.

Cai, Y., Jin, J., Tomomori-Sato, C., Sato, S., Sorokina, I., Parmely, T.J., Conaway, R.C., and Conaway, J.W. (2003). Identification of new subunits of the multiprotein mammalian TRRAP/TIP60-containing histone acetyltransferase complex. The Journal of biological chemistry *278*, 42733-42736.

Cappello, F., Ribbene, A., Campanella, C., Czarnecka, A.M., Anzalone, R., Bucchieri, F., Palma, A., and Zummo, G. (2006). The value of immunohistochemical research on PCNA, p53 and heat shock proteins in prostate cancer management: a review. European journal of histochemistry : EJH *50*, 25-34.

Carlin, L.M., Evans, R., Milewicz, H., Fernandes, L., Matthews, D.R., Perani, M., Levitt, J., Keppler, M.D., Monypenny, J., Coolen, T.*, et al.* (2011). A targeted siRNA screen identifies regulators of Cdc42 activity at the natural killer cell immunological synapse. Science signaling *4*, ra81.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005). LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. Development (Cambridge, England) *132*, 885-896.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Chan, K.S., Xu, J., Wardan, H., McColl, B., Orkin, S., and Vadolas, J. (2012). Generation of a genomic reporter assay system for analysis of γ- and β-globin gene regulation. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *26*, 1736-1744.

Chandler, K.J., Chandler, R.L., Broeckelmann, E.M., Hou, Y., Southard-Smith, E.M., and Mortlock, D.P. (2007). Relevance of BAC transgene copy number in mice: transgene copy number variation across multiple transgenic lines and correlations with transgene integrity and expression. Mammalian genome : official journal of the International Mammalian Genome Society *18*, 693-708.

Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J., and Xi, J.J. (2013). Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. Nature Publishing Group *23*, 465-472.

Chatr-Aryamontri, A., Breitkreutz, B.J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L.*, et al.* (2013). The BioGRID interaction database: 2013 update. Nucleic Acids Research *41*, D816-823.

Chen, Y.A., and Scheller, R.H. (2001a). SNARE-mediated membrane fusion. In Nat Rev Mol Cell Biol, pp. 98-106.

Chen, Y.A., and Scheller, R.H. (2001b). SNARE-mediated membrane fusion. Nature reviews Molecular cell biology *2*, 98-106.

Cheng, L.T., Sun, L.T., and Tada, T. (2012). Genome editing in induced pluripotent stem cells. Genes to cells : devoted to molecular & cellular mechanisms *17*, 431-438.

Cho, S., Kim, S., Kim, J., and Kim, J. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nature Biotechnology *31*, 230-232.

Chow, C.M., Georgiou, A., Szutorisz, H., Maia e Silva, A., Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C., and Dillon, N. (2005). Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. EMBO reports *6*, 354-360.

Ciruela, F., Vilardaga, J.P., and Fernández-Dueñas, V. (2010). Lighting up multiprotein complexes: lessons from GPCR oligomerization. Trends in biotechnology *28*, 407-415.

Clements, W.K., and Traver, D. (2013). Signalling pathways that control vertebrate haematopoietic stem cell specification. Nature reviews Immunology *13*, 336-348.

Cohen, S.B., Graham, M.E., Lovrecz, G.O., Bache, N., Robinson, P.J., and Reddel, R.R. (2007). Protein composition of catalytically active human telomerase from immortal cells. Science (New York, NY) *315*, 1850-1853.

Collins, S.R., Kemmeren, P., Zhao, X.C., Greenblatt, J.F., Spencer, F., Holstege, F.C., Weissman, J.S., and Krogan, N.J. (2007a). Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. Molecular & cellular proteomics : MCP *6*, 439-450.

Collins, S.R., Miller, K.M., Maas, N.L., Roguev, A., Fillingham, J., Chu, C.S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M.*, et al.* (2007b). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature *446*, 806-810.

Colman, A., and Dreesen, O. (2009). Pluripotent stem cells and disease modeling. Cell stem cell *5*, 244-247.

Colucci-D'amato, L., Farina, A., Vissers, J., and Chambery, A. (2011). Quantitative Neuroproteomics: Classical and Novel Tools for Studying Neural Differentiation and Function. Stem cell reviews *7*, 77-93.

Conaway, R.C., and Conaway, J.W. (2009a). The INO80 chromatin remodeling complex in transcription, replication and repair. Trends in biochemical sciences *34*, 71-77.

Conaway, R.C., and Conaway, J.W. (2009b). The INO80 chromatin remodeling complex in transcription, replication and repair. In Trends Biochem Sci, pp. 71-77.

Costa, M.d.o.C., and Paulson, H.L. (2012). Toward understanding Machado-Joseph disease. Progress in neurobiology *97*, 239-257.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature Biotechnology *26*, 1367-1372.

Cusack, S. (1999). RNA-protein complexes. Current opinion in structural biology *9*, 66-73.

Dalton, S. (2012). Signaling networks in human pluripotent stem cells. Current opinion in cell biology.

David Sparkman, O. (2000). Mass spectrometry desk reference. 106.

Denslow, S., and Wade, P. (2007a). The human Mi-2/NuRD complex and gene regulation. Oncogene *26*, 5433-5438.

Denslow, S., and Wade, P. (2007b). The human Mi-2/NuRD complex and gene regulation. In Oncogene, pp. 5433-5438.

Deyle, D.R., and Russell, D.W. (2009). Adeno-associated virus vector integration. Current opinion in molecular therapeutics *11*, 442-447.

Dietrich, D.R. (1993). Toxicological and pathological applications of proliferating cell nuclear antigen (PCNA), a novel endogenous marker for cell proliferation. Critical reviews in toxicology *23*, 77-109.

Ding, J., Swain, J.E., and Smith, G.D. (2011). Aurora kinase-A regulates microtubule organizing center (MTOC) localization, chromosome dynamics, and histone-H3 phosphorylation in mouse oocytes. Molecular reproduction and development *78*, 80-90.

Doyon, Y., Selleck, W., Lane, W.S., Tan, S., and Côté, J. (2004a). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. In Mol Cell Biol, pp. 1884-1896.

Doyon, Y., Selleck, W., Lane, W.S., Tan, S., and Côté, J. (2004b). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Molecular and cellular biology *24*, 1884-1896.

Dubinsky, L., Krom, B.P., and Meijler, M.M. (2012). Diazirine based photoaffinity labeling. Bioorganic & medicinal chemistry *20*, 554-570.

Dwane, S., and Kiely, P.A. (2011). Tools used to study how protein complexes are assembled in signaling cascades. Bioengineered bugs *2*, 247-259.

Elder, G.A., Gama Sosa, M.A., and De Gasperi, R. (2010). Transgenic mouse models of Alzheimer's disease. The Mount Sinai journal of medicine, New York *77*, 69-81.

Elion, E.A. (2006). Detection of protein-protein interactions by coprecipitation. Current protocols in neuroscience / editorial board, Jacqueline N Crawley [et al] *Chapter 5*, Unit 5.25.

Elkabetz, Y., and Studer, L. (2008). Human ESC-derived neural rosettes and neural stem cell progression. Cold Spring Harbor symposia on quantitative biology *73*, 377-387.

Ertekin, A., Aramini, J.M., Rossi, P., Leonard, P.G., Janjua, H., Xiao, R., Maglaqui, M., Lee, H.W., Prestegard, J.H., and Montelione, G.T. (2012). Human cyclin-dependent kinase 2-associated protein 1 (CDK2AP1) is dimeric in its disulfide-reduced state, with natively disordered N-terminal region. The Journal of biological chemistry *287*, 16541-16549.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Falk, A., Koch, P., Kesavan, J., Takashima, Y., Ladewig, J., Alexander, M., Wiskow, O., Tailor, J., Trotter, M., Pollard, S.*, et al.* (2012). Capture of Neuroepithelial-Like Stem Cells from Pluripotent Stem Cells Provides a Versatile System for In Vitro Production of Human Neurons. PloS one *7*, e29597.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Folger, K.R., Wong, E.A., Wahl, G., and Capecchi, M.R. (1982). Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. Molecular and cellular biology *2*, 1372-1387.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C.*, et al.* (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Research *41*, D808-815.

Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M., Antes, I., Leonhardt, H., and Xu, S. (2011). Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain. In PLoS ONE, pp. e21306.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotechnology *31*, 822-826.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proceedings of the Japan Academy Series B, Physical and biological sciences *85*, 348-362.

Garapaty, S., Xu, C.F., Trojer, P., Mahajan, M.A., Neubert, T.A., and Samuels, H.H. (2009). Identification and characterization of a novel nuclear protein complex involved in nuclear hormone receptor-mediated gene regulation. The Journal of biological chemistry *284*, 7542-7552.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B.*, et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631-636.

Gibson, T.J., Seiler, M., and Veitia, R.A. (2013). The transience of transient overexpression. Nature Publishing Group *10*, 715-721.

Golemis, E. (2005). Protein-Protein Interactions: A Molecular Cloning Manual. Cold Spring Harbor Laboratory Pr; 2nd edition, 682.

Gore, A., Li, Z., Fung, H.L., Young, J.E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M.A., Kiskinis, E.*, et al.* (2011). Somatic coding mutations in human induced pluripotent stem cells. Nature *471*, 63-67.

Graham, F.L., Smiley, J., Russell, W.C., and Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. The Journal of general virology *36*, 59-74.

Graham, F.L., and van der Eb, A.J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology *52*, 456-467.

Grant, S., Bradfield, J., Zhang, H., Wang, K., Kim, C., Annaiah, K., Santa, E., Glessner, J., Thomas, K., Garris, M.*, et al.* (2009). Investigation of the Locus Near MC4R With Childhood Obesity in Americans of European and African Ancestry. Obesity, 1-5.

Gstaiger, M., Luke, B., Hess, D., Oakeley, E.J., Wirbelauer, C., Blondel, M., Vigneron, M., Peter, M., and Krek, W. (2003). Control of nutrient-sensitive transcription programs by the unconventional prefoldin URI. Science (New York, NY) *302*, 1208-1212.

Guirouilh-Barbat, J., Huck, S., Bertrand, P., Pirzio, L., Desmaze, C., Sabatier, L., and Lopez, B.S. (2004). Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. Molecular cell *14*, 611-623.

Gupta, R., Kasturi, P., Bracher, A., Loew, C., Zheng, M., Villella, A., Garza, D., Hartl, F.U., and Raychaudhuri, S. (2011). Firefly luciferase mutants as sensors of proteome stress. Nature Methods *8*, 879-884.

Haber, J.E., Ira, G., Malkova, A., and Sugawara, N. (2004). Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. Philosophical transactions of the Royal Society of London Series B, Biological sciences *359*, 79-86.

Han, Y., Miller, A., Mangada, J., Liu, Y., Swistowski, A., Zhan, M., Rao, M.S., and Zeng, X. (2009). Identification by automated screening of a small molecule that selectively eliminates neural stem cells derived from hESCs but not dopamine neurons. PloS one *4*, e7155.

Hartl, M., Bader, A.G., and Bister, K. (2003). Molecular targets of the oncogenic transcription factor jun. Current cancer drug targets *3*, 41-55.

Hassed, S.J., Wiley, G.B., Wang, S., Lee, J.Y., Li, S., Xu, W., Zhao, Z.J., Mulvihill, J.J., Robertson, J., Warner, J.*, et al.* (2012). RBPJ mutations identified in two families affected by Adams-Oliver syndrome. American journal of human genetics *91*, 391-395.

Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S.*, et al.* (2012). A census of human soluble protein complexes. Cell *150*, 1068-1081.

Hay, D.C., Fletcher, J., Payne, C., Terrace, J.D., Gallagher, R.C., Snoeys, J., Black, J.R., Wojtacha, D., Samuel, K., Hannoun, Z.*, et al.* (2008). Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. Proceedings of the National Academy of Sciences of the United States of America *105*, 12301-12306.

Helms, V. (2008). Principles of Computational Cell Biology. 289.

Hervouet, E., Lalier, L., Debien, E., Cheray, M., Geairon, A., Rogniaux, H., Loussouarn, D., Martin, S.A., Vallette, F.M., and Cartron, P.F. (2010). Disruption of Dnmt1/PCNA/UHRF1 interactions promotes tumorigenesis from human and mice glial cells. PloS one *5*, e11333.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C.*, et al.* (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nature Biotechnology *29*, 731-734.

Hsu, Y.C., Lee, D.C., and Chiu, I.M. (2007). Neural stem cells, neural progenitors, and neurotrophic factors. Cell transplantation *16*, 133-150.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010a). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. The Journal of cell biology *189*, 739-754.

Hubner, N.C., Bird, A.W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., and Mann, M. (2010b). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. In J Cell Biol, pp. 739-754.

Hussein, S.M., Batada, N.N., Vuoristo, S., Ching, R.W., Autio, R., Närvä, E., Ng, S., Sourour, M., Hämäläinen, R., Olsson, C.*, et al.* (2011). Copy number variation and selection during reprogramming to pluripotency. Nature *471*, 58-62.

Initiative, I.S.C., Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P.W., Beighton, G., Bello, P.A., Benvenisty, N., Berry, L.S.*, et al.* (2007). Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nature Biotechnology *25*, 803-816.

Iourov, I.Y., Soloviev, I.V., Vorsanova, S.G., Monakhov, V.V., and Yurov, Y.B. (2005). An approach for quantitative assessment of fluorescence in situ hybridization (FISH) signals for applied human molecular cytogenetics. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society *53*, 401-408.

Itskovitz-Eldor, J., Schuldiner, M., Karsenti, D., Eden, A., Yanuka, O., Amit, M., Soreq, H., and Benvenisty, N. (2000). Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. Molecular medicine (Cambridge, Mass) *6*, 88-95.

Ivanković, M., Cukusić, A., Gotić, I., Skrobot, N., Matijasić, M., Polancec, D., and Rubelj, I. (2007). Telomerase activity in HeLa cervical carcinoma cell line proliferation. Biogerontology *8*, 163-172.

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M.*, et al.* (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Research *37*, D412-416.

Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Thérien, C., Bergeron, D., Bourassa, S., Greenblatt, J.*, et al.* (2007a). Systematic analysis of the protein interaction network for

the human transcription machinery reveals the identity of the 7SK capping enzyme. Molecular cell *27*, 262-274.

Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Thérien, C., Bergeron, D., Bourassa, S., Greenblatt, J.*, et al.* (2007b). Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. In Mol Cell, pp. 262-274.

Kajiwara, M., Aoi, T., Okita, K., Takahashi, R., Inoue, H., Takayama, N., Endo, H., Eto, K., Toguchida, J., Uemoto, S.*, et al.* (2012). Donor-dependent variations in hepatic differentiation from human-induced pluripotent stem cells. Proceedings of the National Academy of Sciences of the United States of America *109*, 12538-12543.

Kehat, I., and Gepstein, L. (2003). Human embryonic stem cells for myocardial regeneration. Heart failure reviews *8*, 229-236.

Kim, H., Lee, G., Ganat, Y., Papapetrou, E., Lipchina, I., Socci, N., Sadelain, M., and Studer, L. (2011). miR-371-3 Expression Predicts Neural Differentiation Propensity in Human Pluripotent Stem Cells. Stem Cell *8*, 695-706.

Kim, H.S., Lee, G.S., Kim, J.H., Kang, S.K., Lee, B.C., and Hwang, W.S. (2006). Expression of leptin ligand and receptor and effect of exogenous leptin supplement on in vitro development of porcine embryos. Theriogenology *65*, 831-844.

Kim, J., Bonifant, C., Bunz, F., Lane, W., and Waldman, T. (2008). Epitope tagging of endogenous genes in diverse human cell lines. Nucleic acids research *36*, e127-e127.

Kirschner, L.S., and Stratakis, C.A. (2000). Structure of the human ubiquitin fusion gene Uba80 (RPS27a) and one of its pseudogenes. Biochemical and biophysical research communications *270*, 1106-1110.

Koch, P., Breuer, P., Peitz, M., Jungverdorben, J., Kesavan, J., Poppe, D., Doerr, J., Ladewig, J., Mertens, J., ting, T.T.*, et al.* (2011). Excitation-inducedataxin-3aggregationinneurons from patients with Machado–Joseph disease. Nature, 1-7.

Koch, P., Opitz, T., Steinbeck, J.A., Ladewig, J., and Brüstle, O. (2009). A rosette-type, self-renewing human ES cell-derived neural stem cell with potential for in vitro instruction and synaptic integration. Proceedings of the National Academy of Sciences of the United States of America *106*, 3225-3230.

Kohli, M., Rago, C., Lengauer, C., Kinzler, K.W., and Vogelstein, B. (2004). Facile methods for generating human somatic cell gene knockouts using recombinant adeno-associated viruses. Nucleic acids research *32*, e3.

Kotin, R.M., Siniscalco, M., Samulski, R.J., Zhu, X.D., Hunter, L., Laughlin, C.A., McLaughlin, S., Muzyczka, N., Rocchi, M., and Berns, K.I. (1990). Site-specific integration by adeno-associated virus. Proceedings of the National Academy of Sciences of the United States of America *87*, 2211-2215.

Krastev, D.B., Slabicki, M., Paszkowski-Rogacz, M., Hubner, N.C., Junqueira, M., Shevchenko, A., Mann, M., Neugebauer, K.M., and Buchholz, F. (2011). A systematic RNAi synthetic interaction screen reveals a link between p53 and snoRNP assembly. Nature cell biology *13*, 809-818.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P.*, et al.* (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637-643.

Kunkanjanawan, T., Noisa, P., and Parnpai, R. (2011). Modeling neurological disorders by human induced pluripotent stem cells. Journal of biomedicine & biotechnology *2011*, 350131.

La Spada, A.R., and Taylor, J.P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. Nature reviews Genetics *11*, 247-258.

Ladewig, J., Koch, P., Endl, E., Meiners, B., Opitz, T., Couillard-Despres, S., Aigner, L., and Brüstle, O. (2008). Lineage selection of functional and cryopreservable human embryonic stem cell-derived neurons. Stem cells (Dayton, Ohio) *26*, 1705-1712.

Laflamme, M.A., Chen, K.Y., Naumova, A.V., Muskheli, V., Fugate, J.A., Dupras, S.K., Reinecke, H., Xu, C., Hassanipour, M., Police, S*., et al.* (2007). Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. Nature Biotechnology *25*, 1015-1024.

Lebel, R.R., May, M., Pouls, S., Lubs, H.A., Stevenson, R.E., and Schwartz, C.E. (2002). Non-syndromic X-linked mental retardation associated with a missense mutation (P312L) in the FGD1 gene. Clinical genetics *61*, 139-145.

Lee, J., Sayed, N., Hunter, A., Au, K., Wong, W., Mocarski, E., Pera, R., Yakubov, E., and Cooke, J. (2012). Activation of Innate Immunity Is Required for Efficient Nuclear Reprogramming. Cell *151*, 547-558.

Lee, J.H., and Skalnik, D.G. (2008a). Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. Molecular and cellular biology *28*, 609-618.

Lee, J.H., and Skalnik, D.G. (2008b). Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. In Mol Cell Biol, pp. 609-618.

Lee, J.H., and Skalnik, D.G. (2012). Rbm15-Mkl1 interacts with the Setd1b histone H3-Lys4 methyltransferase via a SPOC domain that is required for cytokine-independent proliferation. PloS one *7*, e42965.

Lee, J.H., Tate, C.M., You, J.S., and Skalnik, D.G. (2007). Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. The Journal of biological chemistry *282*, 13419-13428.

Li, X., and Jin, P. (2010). Roles of small regulatory RNAs in determining neuronal identity. In Nat Rev Neurosci, pp. 329-338.

Lichner, Z., Páll, E., Kerekes, A., Pállinger, E., Maraghechi, P., Bosze, Z., and Gócza, E. (2011). The miR-290-295 cluster promotes pluripotency maintenance by regulating cell cycle phase distribution in mouse embryonic stem cells. Differentiation; research in biological diversity *81*, 11-24.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'malley, R., Castanon, R., Klugman, S*., et al.* (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature, 1-8.

Liu, L., Rodriguez-Belmonte, E.M., Mazloum, N., Xie, B., and Lee, M.Y. (2003a). Identification of a novel protein, PDIP38, that interacts with the p50 subunit of DNA polymerase delta and proliferating cell nuclear antigen. The Journal of biological chemistry *278*, 10041-10047.

Liu, L., Rodriguez-Belmonte, E.M., Mazloum, N., Xie, B., and Lee, M.Y. (2003b). Identification of a novel protein, PDIP38, that interacts with the p50 subunit of DNA polymerase delta and proliferating cell nuclear antigen. In J Biol Chem, pp. 10041-10047.

Mali, P., Yang, L., Esvelt, K., Aach, J., Guell, M., Dicarlo, J., Norville, J., and Church, G. (2013). RNA-Guided Human Genome Engineering via Cas9. Science (New York, NY) *339*, 823-826.

Marcon, E., and Moens, P.B. (2005). The evolution of meiosis: recruitment and modification of somatic DNA-repair proteins. BioEssays : news and reviews in molecular, cellular and developmental biology *27*, 795-808.

Mayshar, Y., Ben-David, U., Lavon, N., Biancotti, J.C., Yakir, B., Clark, A.T., Plath, K., Lowry, W.E., and Benvenisty, N. (2010). Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell stem cell *7*, 521-531.

Meyer, T., Begitt, A., and Vinkemeier, U. (2007). Green fluorescent protein-tagging reduces the nucleocytoplasmic shuttling specifically of unphosphorylated STAT1. The FEBS journal *274*, 815-826.

Mikhaylova, O., Stratton, Y., Hall, D., Kellner, E., Ehmer, B., Drew, A.F., Gallo, C.A., Plas, D.R., Biesiada, J., Meller, J.*, et al.* (2012). VHL-regulated MiR-204 suppresses tumor growth through inhibition of LC3B-mediated autophagy in renal clear cell carcinoma. Cancer cell *21*, 532-546.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J.*, et al.* (2011). A TALE nuclease architecture for efficient genome editing. Nature Biotechnology *29*, 143-148.

Milstein, C. (1999). The hybridoma revolution: an offshoot of basic research. BioEssays : news and reviews in molecular, cellular and developmental biology *21*, 966-973.

Miyoshi, N., Ishii, H., Nagano, H., Haraguchi, N., Dewi, D.L., Kano, Y., Nishikawa, S., Tanemura, M., Mimori, K., Tanaka, F.*, et al.* (2011). Reprogramming of Mouse and Human Cells to Pluripotency Using Mature MicroRNAs. Cell stem cell *8*, 633-638.

Modzelewska, K., Newman, L., Desai, R., and Keely, P. (2006). Ack1 Mediates Cdc42-dependent Cell Migration and Signaling to p130Cas. Journal of Biological Chemistry *281*, 37527-37535.

Mohan, M., Lin, C., Guest, E., and Shilatifard, A. (2010). Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. Nature reviews Cancer *10*, 721-728.

Muotri, A.R., Nakashima, K., Toni, N., Sandler, V.M., and Gage, F. (2005). Development of functional human embryonic stem cell-derived neurons in mouse brain. Proceedings of the National Academy of Sciences of the United States of America *102*, 18644-18648.

Nagaike, T., Logan, C., Hotta, I., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J.L. (2011). Transcriptional activators enhance polyadenylation of mRNA precursors. Molecular cell *41*, 409-418.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *90*, 8424-8428.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nature Biotechnology *26*, 101-106.

Ni, Z., Olsen, J.B., Emili, A., and Greenblatt, J.F. (2011). Identification of mammalian protein complexes by lentiviral-based affinity purification and mass spectrometry. Methods in molecular biology (Clifton, NJ) *781*, 31-45.

Nistor, G.I., Totoiu, M.O., Haque, N., Carpenter, M.K., and Keirstead, H.S. (2005). Human embryonic stem cells differentiate into oligodendrocytes in high purity and myelinate after spinal cord transplantation. Glia *49*, 385-396.

O'Connor, M., Peifer, M., and Bender, W. (1989). Construction of large DNA segments in Escherichia coli. Science (New York, NY) *244*, 1307-1312.

Ohmine, S., Dietz, A.B., Deeds, M.C., Hartjes, K.A., Miller, D.R., Thatava, T., Sakuma, T., Kudva, Y.C., and Ikeda, Y. (2011). Induced pluripotent stem cells from GMP-grade hematopoietic progenitor cells and mononuclear myeloid cells. Stem cell research & therapy *2*, 46.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317.

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K.*, et al.* (2011). A more efficient method to generate integration-free human iPS cells. Nature Methods *8*, 409-412.

Ortega, A., Niksic, M., Bachi, A., Wilm, M., Sánchez, L., Hastie, N., and Valcárcel, J. (2003). Biochemical function of female-lethal (2)D/Wilms' tumor suppressor-1-associated proteins in alternative pre-mRNA splicing. In J Biol Chem, pp. 3040-3047.

Park, S.S., Caballero, S., Bauer, G., Shibata, B., Roth, A., Fitzgerald, P.G., Forward, K.I., Zhou, P., McGee, J., Telander, D.G.*, et al.* (2012). Long-term effects of intravitreal injection of GMP-grade bone-marrow-derived CD34+ cells in NOD-SCID mice with acute ischemia-reperfusion injury. Investigative ophthalmology & visual science *53*, 986-994.

Patel, A., Dharmarajan, V., Vought, V.E., and Cosgrove, M.S. (2009). On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. The Journal of biological chemistry *284*, 24242-24256.

Paul, F.E., Hosp, F., and Selbach, M. (2011). Analyzing proteinâ€“protein interactions by quantitative mass spectrometry. Methods *54*, 387-395.

Peng, H., Shintani, S., Kim, Y., and Wong, D.T. (2006). Loss of p12CDK2-AP1 expression in human oral squamous cell carcinoma with disrupted transforming growth factor-beta-Smad signaling pathway. Neoplasia (New York, NY) *8*, 1028-1036.

Perrier, A.L., Tabar, V., Barberi, T., Rubio, M.E., Bruses, J., Topf, N., Harrison, N.L., and Studer, L. (2004). Derivation of midbrain dopamine neurons from human embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *101*, 12543-12548.

Peters, A., Burridge, P.W., Pryzhkova, M.V., Levine, M.A., Park, T.S., Roxbury, C., Yuan, X., Péault, B., and Zambidis, E.T. (2010). Challenges and strategies for generating therapeutic patient-specific hemangioblasts and hematopoietic stem cells from human pluripotent stem cells. The International journal of developmental biology *54*, 965-990.

Peura, T., Bosman, A., Chami, O., Jansen, R.P., Texlova, K., and Stojanov, T. (2008). Karyotypically normal and abnormal human embryonic stem cell lines derived from PGD-analyzed embryos. Cloning and stem cells *10*, 203-216.

Placantonakis, D., Tomishima, M., Lafaille, F., Desbordes, S., Jia, F., Socci, N., Viale, A., Lee, H., Harrison, N., Tabar, V.*, et al.* (2009). BAC transgenesis in human embryonic stem cells as a novel tool to define the human neural lineage. Stem cells (Dayton, Ohio) *27*, 521-532.

Porteus, M. (2011). Homologous recombination-based gene therapy for the primary immunodeficiencies. Annals of the New York Academy of Sciences *1246*, 131-140.

Poser, I., Sarov, M., Hutchins, J.R., Hériché, J.K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W.*, et al.* (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nature methods *5*, 409-415.

Pournasr, B., Khaloughi, K., Salekdeh, G.H., Totonchi, M., Shahbazi, E., and Baharvand, H. (2011). Concise review: alchemy of biology: generating desired cell types from abundant and accessible cells. Stem cells (Dayton, Ohio) *29*, 1933-1941.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods (San Diego, Calif) *24*, 218-229.

Radtke, F., and Clevers, H. (2005). Self-renewal and cancer of the gut: two sides of a coin. Science (New York, NY) *307*, 1904-1909.

Ramalingam, S., London, V., Kandavelou, K., Cebotaru, L., Guggino, W.G., Civin, C., and Chandrasegaran, S. (2012). Generation and Genetic Engineering of human induced pluripotent stem cells Using Designed Zinc Finger Nucleases. Stem cells and development.

Reeves, W.H., and Sthoeger, Z.M. (1989a). Molecular cloning of cDNA encoding the p70 (Ku) lupus autoantigen. The Journal of biological chemistry *264*, 5047-5052.

Reeves, W.H., and Sthoeger, Z.M. (1989b). Molecular cloning of cDNA encoding the p70 (Ku) lupus autoantigen. In J Biol Chem, pp. 5047-5052.

Reubinoff, B.E., Itsykson, P., Turetsky, T., Pera, M.F., Reinhartz, E., Itzik, A., and Ben-Hur, T. (2001). Neural progenitors from human embryonic stem cells. Nature Biotechnology *19*, 1134-1140.

Riesen, F., Rothen-Rutishauser, B., and Wunderli-Allenspach, H. (2002). A ZO1-GFP fusion protein to study the dynamics of tight junctions in living cells. Histochemistry and Cell Biology *117*, 307-315.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nature biotechnology *17*, 1030-1032.

Ronen, D., and Benvenisty, N. (2012). Genomic stability in reprogramming. Current opinion in genetics & development.

Rostovskaya, M., Fu, J., Obst, M., Baer, I., Weidlich, S., Wang, H., Smith, A., Anastassiadis, K., and Stewart, A. (2012). Transposon-mediated BAC transgenesis in human ES cells. Nucleic acids research, 1-13.

Rousseau, B., Ménard, L., Haurie, V., Taras, D., Blanc, J.F., Moreau-Gaudry, F., Metzler, P., Hugues, M., Boyault, S., Lemière, S*., et al.* (2007). Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma. Hepatology (Baltimore, Md) *46*, 1108-1118.

Rüb, U., Schöls, L., Paulson, H., Auburger, G., Kermer, P., Jen, J.C., Seidel, K., Korf, H., and Deller, T. (2013). Clinical features, neurogenetics and neuropathology of the polyglutamine spinocerebellar ataxias type 1, 2, 3, 6 and 7. Progress in neurobiology *104*, 38-66.

Ruepp, A., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Stransky, M., Waegele, B., Schmidt, T., Doudieu, O.N., Stümpflen, V*., et al.* (2008). CORUM: the comprehensive resource of mammalian protein complexes. Nucleic Acids Research *36*, D646-650.

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes--2009. Nucleic Acids Research *38*, D497-501.

Sabatino, L., Fucci, A., Pancione, M., Carafa, V., Nebbioso, A., Pistore, C., Babbio, F., Votino, C., Laudanna, C., Ceccarelli, M*., et al.* (2012). UHRF1 coordinates peroxisome proliferator activated receptor gamma (PPARG) epigenetic silencing and mediates colorectal cancer progression. Oncogene *31*, 5061-5072.

Sager, J.J., Bai, Q., and Burton, E.A. (2010). Transgenic zebrafish models of neurodegenerative diseases. Brain structure & function *214*, 285-302.

Sakai, T., Hino, K., Wada, S., and Maeda, H. (2003a). Identification of the DNA binding specificity of the human ZNF219 protein and its function as a transcriptional repressor. DNA research : an international journal for rapid publication of reports on genes and genomes *10*, 155-165.

Sakai, T., Hino, K., Wada, S., and Maeda, H. (2003b). Identification of the DNA binding specificity of the human ZNF219 protein and its function as a transcriptional repressor. In DNA Res, pp. 155-165.

Sanchez, C., Lachaize, C., Janody, F., Bellon, B., Röder, L., Euzenat, J., Rechenmann, F., and Jacq, B. (1999). Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, an Internet database. Nucleic acids research *27*, 89-94.

Sardiu, M., Cai, Y., Jin, J., Swanson, S.K., Conaway, R.C., Conaway, J.W., Florens, L., and Washburn, M.P. (2008a). Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. Proceedings of the National Academy of Sciences of the United States of America *105*, 1454-1459.

Sardiu, M., Cai, Y., Jin, J., Swanson, S.K., Conaway, R.C., Conaway, J.W., Florens, L., and Washburn, M.P. (2008b). Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. In Proc Natl Acad Sci USA, pp. 1454-1459.

Scherer, W., Syverton, J., and Gey, G. (1953). Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. The Journal of experimental medicine *97*, 695-710.

Schirle, M., Bantscheff, M., and Kuster, B. (2012). Mass Spectrometry-Based Proteomics in Preclinical Drug Discovery. Chemistry & Biology *19*, 72-84.

Schneider, U., Schwenk, H.U., and Bornkamm, G. (1977). Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. International journal of cancer Journal international du cancer *19*, 621-626.

Seki, A., Coppinger, J.A., Jang, C.Y., Yates, J.R., and Fang, G. (2008). Bora and the kinase Aurora a cooperatively activate the kinase Plk1 and control mitotic entry. Science (New York, NY) *320*, 1655-1658.

Sigala, B., Edwards, M., Puri, T., and Tsaneva, I.R. (2005). Relocalization of human chromatin remodeling cofactor TIP48 in mitosis. Experimental cell research *310*, 357-369.

Simpson, F., Lammerts van Bueren, K., Butterfield, N., Bennetts, J.S., Bowles, J., Adolphe, C., Simms, L.A., Young, J., Walsh, M.D., Leggett, B.*, et al.* (2006). The PCNA-associated factor KIAA0101/p15(PAF) binds the potential tumor suppressor product p33ING1b. In Exp Cell Res, pp. 73-85.

Smith, E., Lin, C., and Shilatifard, A. (2011). The super elongation complex (SEC) and MLL in development and disease. Genes & Development *25*, 661-672.

Smithies, O., Koralewski, M.A., Song, K.Y., and Kucherlapati, R.S. (1984). Homologous recombination with DNA introduced into mammalian cells. Cold Spring Harbor symposia on quantitative biology *49*, 161-170.

Smits, A.H., Jansen, P.W., Poser, I., Hyman, A.A., and Vermeulen, M. (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic acids research *41*, e28.

Snee, M., Kidd, G.J., Munro, T.P., and Smith, R. (2002). RNA trafficking and stabilization elements associate with multiple brain proteins. Journal of cell science *115*, 4661-4669.

Snel, B., Lehmann, G., Bork, P., and Huynen, M.A. (2000). STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Research *28*, 3442-3444.

Solovei, I., Cavallo, A., Schermelleh, L., Jaunin, F., Scasselati, C., Cmarko, D., Cremer, C., Fakan, S., and Cremer, T. (2002). Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3D-FISH). Experimental cell research *276*, 10-23.

Sommer, C.A., Sommer, A.G., Longmire, T.A., Christodoulou, C., Thomas, D.D., Gostissa, M., Alt, F.W., Murphy, G.J., Kotton, D.N., and Mostoslavsky, G. (2010). Excision of reprogramming transgenes improves the differentiation potential of iPS cells generated with a single excisable vector. STEM CELLS *28*, 64-74.

Sparwasser, T., and Eberl, G. (2007). BAC to immunology--bacterial artificial chromosome-mediated transgenesis for targeting of immune cells. Immunology *121*, 308-313.

Spruijt, C., Bartels, S., Brinkman, A., Tjeertes, J., Poser, I., Stunnenberg, H., and Vermeulen, M. (2010). CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex. Molecular BioSystems *6*, 1700.

Stadler, C., Rexhepaj, E., Singan, V., Murphy, R., Pepperkok, R., Uhlén, M., Simpson, J., and Lundberg, E. (2013). Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. Nature methods *10*, 315-323.

Stark, C., Breitkreutz, B.J., Chatr-Aryamontri, A., Boucher, L., Oughtred, R., Livstone, M.S., Nixon, J., Van Auken, K., Wang, X., Shi, X.*, et al.* (2011). The BioGRID Interaction Database: 2011 update. Nucleic Acids Research *39*, D698-704.

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Research *34*, D535-539.

Steinbeck, J.A., Koch, P., Derouiche, A., and Brüstle, O. (2011). Human embryonic stem cell-derived neurons establish region-specific, long-range projections in the adult brain. Cellular and Molecular Life Sciences, 1-10.

Strelchenko, N., Verlinsky, O., Kukharenko, V., and Verlinsky, Y. (2004). Morula-derived human embryonic stem cells. Reproductive biomedicine online *9*, 623-629.

Surosky, R.T., Urabe, M., Godwin, S.G., McQuiston, S.A., Kurtzman, G.J., Ozawa, K., and Natsoulis, G. (1997). Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. Journal of virology *71*, 7951-7959.

Suter, B., Kittanakom, S., and Stagljar, I. (2008). Two-hybrid technologies in proteomics research. Current opinion in biotechnology *19*, 316-323.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P.*, et al.* (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Research *39*, D561-568.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004a). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51-61.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004b). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. In Cell, pp. 51-61.

Takahashi, K., Okita, K., Nakagawa, M., and Yamanaka, S. (2007). Induction of pluripotent stem cells from fibroblast cultures. Nature protocols *2*, 3081-3089.

Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell *51*, 503-512.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science (New York, NY) *282*, 1145-1147.

Trinkle-Mulcahy, L., Boulon, S., Lam, Y.W., Urcia, R., Boisvert, F.M., Vandermoere, F., Morrice, N.A., Swift, S., Rothbauer, U., Leonhardt, H.*, et al.* (2008). Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. The Journal of cell biology *183*, 223-239.

Tsuchiya, S., Yamabe, M., Yamaguchi, Y., Kobayashi, Y., Konno, T., and Tada, K. (1980). Establishment and characterization of a human acute monocytic leukemia cell line (THP-1). International journal of cancer Journal international du cancer *26*, 171-176.

Tunster, S.J., Van De Pette, M., and John, R.M. (2011). BACs as tools for the study of genomic imprinting. Journal of biomedicine & biotechnology *2011*, 283013.

Turetsky, T., Aizenman, E., Gil, Y., Weinberg, N., Shufaro, Y., Revel, A., Laufer, N., Simon, A., Abeliovich, D., and Reubinoff, B.E. (2008). Laser-assisted derivation of human embryonic stem cell lines from IVF embryos after preimplantation genetic diagnosis. Human reproduction (Oxford, England) *23*, 46-53.

Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S.*, et al.* (2010a). Towards a knowledge-based Human Protein Atlas. Nature Publishing Group *28*, 1248-1250.

Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S.*, et al.* (2010b). Towards a knowledge-based Human Protein Atlas. In Nature Publishing Group, pp. 1248-1250.

Unger, C., Skottman, H., Blomberg, P., Dilber, M.S., and Hovatta, O. (2008). Good manufacturing practice and clinical-grade human embryonic stem cell lines. Human Molecular Genetics *17*, R48-53.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nature reviews Genetics *11*, 636-646.

Vallier, L., Reynolds, D., and Pedersen, R.A. (2004). Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. Developmental biology *275*, 403-421.

Varma, M., and Leavitt, J. (1988). Macromolecular changes accompanying immortalization and tumorigenic conversion in a human fibroblast model system. Mutation research *199*, 437-447.

Vasileva, A., and Jessberger, R. (2005). Precise hit: adeno-associated virus in gene targeting. Nature reviews Microbiology *3*, 837-847.

Vermeulen, M., Hubner, N.C., and Mann, M. (2008). High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. Current opinion in biotechnology *19*, 331-337.

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M.*, et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.

von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: a database of predicted functional associations between proteins. Nucleic Acids Research *31*, 258-261.

von Mering, C., Jensen, L.J., Kuhn, M., Chaffron, S., Doerks, T., Krüger, B., Snel, B., and Bork, P. (2007). STRING 7--recent developments in the integration and prediction of protein interactions. Nucleic Acids Research *35*, D358-362.

Wang, A., and Sander, M. (2012). Generating cells of the gastrointestinal system: current approaches and applications for the differentiation of human pluripotent stem cells. Journal of molecular medicine (Berlin, Germany) *90*, 763-771.

Wang, X., Lou, Z., Dong, X., Yang, W., Peng, Y., Yin, B., Gong, Y., Yuan, J., Zhou, W., Bartlam, M.*, et al.* (2009a). Crystal structure of the C-terminal domain of human DPY-30-like protein: A component of the histone methyltransferase complex. In J Mol Biol, pp. 530-537.

Wang, X., Lou, Z., Dong, X., Yang, W., Peng, Y., Yin, B., Gong, Y., Yuan, J., Zhou, W., Bartlam, M.*, et al.* (2009b). Crystal structure of the C-terminal domain of human DPY-30-like protein: A component of the histone methyltransferase complex. Journal of molecular biology *390*, 530-537.

Warren, L., Manos, P., Ahfeldt, T., Loh, Y., Li, H., Lau, F., Ebina, W., Mandal, P., Smith, Z., and Meissner, A. (2010). Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. Cell stem cell *7*, 618-630.

Wernig, M., Tucker, K.L., Gornik, V., Schneiders, A., Buschwald, R., Wiestler, O.D., Barde, Y.A., and Brüstle, O. (2002). Tau EGFP embryonic stem cells: an efficient tool for neuronal lineage selection and transplantation. Journal of neuroscience research *69*, 918-924.

Wesselschmidt, R.L. (2011). The teratoma assay: an in vivo assessment of pluripotency. Methods in molecular biology (Clifton, NJ) *767*, 231-241.

Wink, S., Hiemstra, S., Huppelschoten, S., Danen, E., Niemeijer, M., Hendriks, G., Vrieling, H., Herpers, B., and van de Water, B. (2014). Quantitative High Content Imaging of Cellular Adaptive Stress Response Pathways in Toxicity for Chemical Safety Assessment. Chemical research in toxicology.

Wong, M.M., Cox, L.K., and Chrivia, J.C. (2007). The chromatin remodeling protein, SRCAP, is critical for deposition of the histone variant H2A.Z at promoters. The Journal of biological chemistry *282*, 26132-26139.

Xia, Z., Wei, P., Zhang, H., Ding, Z., Yang, L., Huang, Z., and Zhang, N. (2013). AURKA governs self-renewal capacity in glioma-initiating cells via stabilization/activation of β-catenin/Wnt signaling. Molecular cancer research : MCR *11*, 1101-1111.

Yan, Y., Yang, D., Zarnowska, E.D., Du, Z., Werbel, B., Valliere, C., Pearce, R.A., Thomson, J.A., and Zhang, S.C. (2005). Directed differentiation of dopaminergic neuronal subtypes from human embryonic stem cells. Stem cells (Dayton, Ohio) *23*, 781-790.

Yao, N., Coryell, L., Zhang, D., Georgescu, R.E., Finkelstein, J., Coman, M.M., Hingorani, M.M., and O'Donnell, M. (2003). Replication factor C clamp loader subunit arrangement within the circular pentamer and its attachment points to proliferating cell nuclear antigen. The Journal of biological chemistry *278*, 50744-50753.

Yarden, R.I., and Brody, L.C. (1999). BRCA1 interacts with components of the histone deacetylase complex. Proceedings of the National Academy of Sciences of the United States of America *96*, 4983-4988.

Yoo, A.S., Sun, A.X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., Lee-Messer, C., Dolmetsch, R.E., Tsien, R.W., and Crabtree, G.R. (2011a). MicroRNA-mediated conversion of human fibroblasts to neurons. In Nature, pp. 1-5.

Yoo, A.S., Sun, A.X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., Lee-Messer, C., Dolmetsch, R.E., Tsien, R.W., and Crabtree, G.R. (2011b). MicroRNA-mediated conversion of human fibroblasts to neurons. Nature, 1-5.

Zan, Y., Haag, J.D., Chen, K.S., Shepel, L.A., Wigington, D., Wang, Y.R., Hu, R., Lopez-Guajardo, C.C., Brose, H.L., Porter, K.I.*, et al.* (2003). Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. Nature biotechnology *21*, 645-651.

Zhang, Q., Vo, N., and Goodman, R.H. (2000). Histone binding protein RbAp48 interacts with a complex of CREB binding protein and phosphorylated CREB. Molecular and cellular biology *20*, 4970-4978.

Zhang, S.C., Wernig, M., Duncan, I.D., Brüstle, O., and Thomson, J.A. (2001). In vitro differentiation of transplantable neural precursors from human embryonic stem cells. Nature Biotechnology *19*, 1129-1133.

Zhu, S., Li, W., Zhou, H., Wei, W., Ambasudhan, R., Lin, T., Kim, J., Zhang, K., and Ding, S. (2010). Reprogramming of human primary somatic cells by OCT4 and chemical compounds. Cell stem cell *7*, 651-655.

10. Danksagung

Zuerst möchte mich bei Prof. Dr. O. Brüstle für die Möglichkeit der Promotion an seinem Institut unter optimalen Forschungsbedingungen bedanken. Dies gilt auch für sein stetes Interesse an den Ergebnissen sowie dem Fortschritt meiner Projekte.

Großer Dank gilt auch Dr. Philipp Koch, welcher mit exzellentem Rat, Kritik und wertvollen Konzepten ein wunderbares wissenschaftliches Arbeiten ermöglichte und große Teile meines wissenschaftlichen Denkens geprägt hat.

Meinem guten Freund und Kollegen Jérôme Mertens, mit dem stets sämtliche privaten und fachlichen Diskussionen zu einer positiven Erfahrung wurden.

Daniel Poppe, für seine vielen weisen Ratschläge, gute fachliche Diskussionen und eine wunderbare Freundschaft.

Meinen guten Freunden Thomas Zillinger und Eugen Lounkine, welche stets gezeigt haben, dass Forschung mit einem gewissen Maß an Unbeschwertheit eben doch ein Kinderspiel ist.

Den Mitgliedern der AG Koch durch welche ein ausgezeichnetes Arbeitsklima zustande kam.

Besonderer Dank gilt meinen Kollegen Carolin Haubenreich, Kathrin Stüber, Julia Fischer, Anke Leinhaas und David Kühne mit denen das Arbeiten immer zum Vergnügen wurde.

Dr. Ina Poser, Dr. Irina Solovei, Marco Hein, Annika Böckenhoff und Dr. Martin Schwarz für die tolle Kollaboration bei den vielen gemeinsamen Experimenten.

Meiner Freundin Anna-Maria Herzner danke ich für alles. Und viel mehr!

Meinen Eltern Serena und Bernhard Doerr dafür dass es mich gibt.

11. Erklärung

Hiermit versichere ich, dass diese Dissertation von mir persönlich, selbständig und ohne jede unerlaubte Hilfe angefertigt wurde. Die Daten, die im Rahmen einer Kooperation gewonnen wurden sind ausnahmslos gekennzeichnet. Die aus anderen Quellen übernommenen Daten, Abbildungen und Konzepte sind unter Angabe der jeweiligen Quelle gekennzeichnet.

Ergebnisse dieser Arbeit trugen in Teilen zu folgenden Veröffentlichungen bei:

Koch P., Breuer P., Peitz M., Jungverdorben J., Kesavan J., Poppe D., **Doerr J.**, Ladewig J., Mertens J., Tüting T., Hoffmann P., Klockgether T., Evert BO., Wüllner U., Brüstle O.;
Excitation-induced ataxin-3 aggregation in neurons from patients with Machado–Joseph disease
Nature, 2011 Nov 23;480(7378):543-6

Ladewig J., Mertens J., Kesavan J., **Doerr J**., Poppe D., Glaue F., Herms S., Wernet P., Kögler G., Müller FJ., Koch P. and Brüstle O.;
Small molecules enable highly efficient neuronal conversion of human fibroblasts
Nature Methods, 2012 Jun;9(6):575-8

Mertens J., Stüber K., Poppe D., **Doerr J.**, Ladewig L., Brüstle O., Koch P.;
Embryonic Stem Cell-Based Modeling of Tau Pathology in Human Neurons
American Journal of Pathology, 2013 May;182(5):1769-79

Doerr J., Hein MY., Poser I., Solovei I., Mertens J., Leonhardt H., Hyman AA., Mann M., Koch P. and Brüstle O.;
BAC-based assessment of protein-protein interactions in human pluripotent stem cell-derived neural stem cells and their neuronal progeny
in preparation

Doerr J., Böckenhoff A., Mertens J., Ewald B., Ladewig J., Matzner U., Eckhardt M., Gieselmann V. and Brüstle O.*, Koch P.*
Genetically corrected human iPSC-derived neural progenitor cells reduce central nervous sulfatide storage in a mouse model of metachromatic leukodystrophy
in preparation

Doerr J., Solovei I., Haubenreich C., Herms S., Hoffman P., Leonhardt H. and Brüstle O., Koch P.
Jumping translocations during long-term culture of pluripotent stem cell derived neural stem cells are random
in preparation

Doerr J., Schwarz M., Leinhaas A., Koch P. and Brüstle O.
A quantification of network integration of human neuroepithelial-like stem cells into murine host circuitries using a Rabies-based transsynaptic labeling technique in vivo
in preparation

Die vorliegende Arbeit wurde an keiner anderen Hochschule als Dissertation eingereicht. Ich habe früher noch keinen Promotionsversuch unternommen.

_____

Jonas Martin Doerr,
Bonn, den 3.2.2014

Academic Achievements

Publications

Koch P., Breuer P., Peitz M., Jungverdorben J., Kesavan J., Poppe D., **Doerr J.**, Ladewig J., Mertens J., Tüting T., Hoffmann P., Klockgether T., Evert BO., Wüllner U., Brüstle O.;
Excitation-induced ataxin-3 aggregation in neurons from patients with Machado–Joseph disease
Nature, 2011 Nov 23;480(7378):543-6

Ladewig J., Mertens J., Kesavan J., **Doerr J**., Poppe D., Glaue F., Herms S., Wernet P., Kögler G., Müller FJ., Koch P. and Brüstle O.;
Small molecules enable highly efficient neuronal conversion of human fibroblasts
Nature Methods, 2012 Jun;9(6):575-8

Mertens J., Stüber K., Poppe D., **Doerr J.**, Ladewig L., Brüstle O., Koch P.;
Embryonic Stem Cell-Based Modeling of Tau Pathology in Human Neurons
American Journal of Pathology, 2013 May;182(5):1769-79

Doerr J., Hein MY., Poser I., Solovei I., Mertens J., Leonhardt H., Hyman AA., Mann M., Koch P. and Brüstle O.;
BAC-based assessment of protein-protein interactions in human pluripotent stem cell-derived neural stem cells and their neuronal progeny
in preparation

Doerr J., Böckenhoff A., Mertens J., Ewald B., Ladewig J., Matzner U., Eckhardt M., Gieselmann V. and Brüstle O.*, Koch P.*
Genetically corrected human iPSC-derived neural progenitor cells reduce central nervous sulfatide storage in a mouse model of metachromatic leukodystrophy
in preparation

Doerr J., Solovei I., Haubenreich C., Herms S., Hoffman P., Leonhardt H. and Brüstle O., Koch P.
Jumping translocations during long-term culture of pluripotent stem cell derived neural stem cells are random
in preparation

Doerr J., Schwarz M., Leinhaas A., Koch P. and Brüstle O.
A quantification of network integration of human neuroepithelial-like stem cells into murine host circuitries using a Rabies-based transsynaptic labeling technique
in preparation