Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

M.Sc. Marie-Therese Mackmull

born in: Mosbach, Germany

Oral examination: 20.04.2017

# The landscape of the nucleocytoplasmic transport system and cell-type specific variations of the nuclear pore complex

Referees: Dr. Darren Gilmour

Prof. Dr. Matthias Mayer

This study was carried out in the Structural and Computational Biology Unit (SCB) at the European Molecular Biology Laboratory (EMBL) Heidelberg from May 2013 to February 2017 supervised by Dr. Martin Beck.

## Summary

The main function of the nuclear pore complex (NPC) is to facilitate and regulate the transport between the cytosol and the nucleus but NPCs are also involved in various other cellular functions, including regulation of gene expression, translational control, signal transduction and cell differentiation. The nucleocytoplasmic transport system is composed of the NPC, that forms a large aqueous channel lined with FG-repeats containing nucleoporins (FG-Nups). FG-Nups constitute a permeability barrier, which prevents the passage of the majority of all proteins. Nuclear transport receptors (NTRs, also called importins, exportins or karyopherins) specifically recognize localization signals of cargo molecules and facilitate their passage through the central channel by transiently interacting with FG-Nups. Classical methods such as affinity purification or measurement of dissociation constants are not well-suited to globally identify NTR-cargo interactions because they are of a very transient nature, the spectrum of recognized cargos is huge and their dynamic concentration range comprises orders of magnitude. The exact cargo spectrum of the majority of NTRs, their specificity and even the extent to which active nucleocytoplasmic transport contributes to protein localization thus remains uncertain. To systematically map cargo-NTR relationships in an unbiased way *in situ*, I used proximity labeling mass spectrometry based on the so-called BioID system. I systematically fused BirA to various NTRs and other factors involved in nucleocytoplasmic transport. I found that at least one third of the human proteome is subject to active nuclear transport. I characterized the specific cargo spectrum of several NTRs and can thus estimate their specificity or overlap with other transport pathways. I identified the responsible transport pathways of various key protein complexes and demonstrated that those and components of pathways tend to be transported by related NTRs. The identification of the exact biotinylation sites provided evidence for the relevant interaction surfaces and sheds light in direct versus piggyback transport mechanisms.

To understand the compositional changes, potentially affecting the nucleocytoplasmic transport, within the NPC and NTR system better, I investigated them during development and within different cell types in *Drosophila* and Zebrafish. For the NPC only Nups, exposed to the environment, showed significant changes. The abundance changes of the NTRs were more dynamic and indicated a more flexible adaptation to changing circumstances

## Zusammenfassung

Die Hauptfunktion der Kernporen (NPC) besteht darin, den Transport zwischen dem Zytoplasms und dem Zellkern zu erleichtern und zu regulieren, aber NPCs sind auch an verschiedenen anderen zellulären Funktionen beteiligt, einschließlich der Regulations von Transkription und Translation, Signaltransduktion und Zelldifferenzierung. Das nukleozytoplasmatische Transportsystem besteht aus dem NPC, der einen großen wässrigen Kanal bildet, der mit Nukleoporinen mit FG-Wiederholungen (FG-Nups) ausgekleidet ist. FG-Nups bilden eine Permeabilitätsbarriere, die die Passage der meisten Proteine verhindert. Kerntransportrezeptoren (NTRs, auch Importine, Exportine oder Karyopherine genannt) erkennen spezifisch Lokalisierungssignale und erleichtern den Durchgang von Cargos durch transiente Interaktionen mit FG-Nups im zentralen Kanal. Klassische Methoden wie die Affinitätsreinigung oder die Messung von Dissoziationskonstanten sind ungeeignet, um NTR-Cargo-Wechselwirkungen global zu identifizieren, da sie sehr transient sind, das Spektrum unterschiedlicher Cargos riesig ist und ihr Konzentrationsbereich mehrere Größenordnungen umfasst. Das genaue Cargospektrum der meisten NTRs, ihre Spezifität und sogar das Ausmaß, in dem der aktive nucleozytoplasmatische Transport zur Proteinlokalisierung beiträgt, bleibt somit ungewiss. Um die NTR-Cargo-Beziehungen systematisch *in situ* abzubilden, habe ich das so genannten BioID-System verwendet in Kombination mit Massenspektrometrie. Ich fusionierte BirA mit verschiedene NTRs und andere Faktoren, die am nukleozytoplasmatischen Transport beteiligt sind. Ich habe heraus gefunden, dass mindestens ein Drittel des menschlichen Proteoms einem aktiven Kerntransport unterliegt. Ich charakterisierte das spezifische Cargospektrum mehrerer NTRs und kann damit deren Spezifität oder Überlappung mit anderen Transportwegen abschätzen. Ich identifizierte die verantwortlichen Transportwege von verschiedenen Schlüsselproteinkomplexen und zeigte, dass diese und Komponenten von Signalwegen durch verwandte NTRs transportiert werden. Die Identifizierung der genauen Biotinylierungsstellen lieferte Hinweise auf relevante Wechselwirkungsoberflächen und direkter Transport oder indirekten Transport via Huckepack.

Um die Veränderungen der Zusammensetzung zu verstehen, die möglicherweise den nukleozyto-plasmatischen Transport innerhalb des NPC und NTR-Systems beeinflussen, untersuchte ich diese während der Entwicklung und in verschiedenen Zelltypen in *Drosophila* und Zebrafisch. Für den NPC zeigten nur Nups, die der Umwelt ausgesetzt waren, signifikante Veränderungen. Die Änderungen der NTRs war dynamischer und zeigte eine flexiblere Anpassung an sich verändernte Umstände.

## Table of Contents

# 1 Introduction

Compartmentalization is a hallmark of eukaryotic cells. One of their compartments is the nucleus, which contains almost all genetic information. This is not an isolated compartment; an active exchange between the nucleus, where DNA transcription occurs, and the cytoplasm, where mRNA translation takes place, is essential. This connection is made by nuclear pore complexes (NPCs), selective gates that are located within the nuclear envelope (NE). The nuclear transport system controls the nucleocytoplasmic localization of various essential cellular components by actively transporting them to either side of the nuclear envelope. It facilitates the import of transcription factors that function downstream of signaling pathways, the import of histones that are required for DNA replication and the export of ribosomes and fully processed mRNAs needed for translation and ultimately proliferation (Dickmanns et al., 2015), to name just a few fundamental functions. Therefore, the nuclear transport system critically contributes to regulate cellular homeostasis.

The NPC, which is embedded in the double membrane surrounding the nucleus, is the only gateway between the nucleus and the cytoplasm during interphase. The main and most apparent function of the NPC is to facilitate and regulate transport of macromolecules between the cytosol and the nucleus (more details see below). Additionally the NPC also acts as a barrier to guard the composition of the nuclear compartment. For example viruses, which need to enter the nucleus for replication, had to develop strategies to overcome this barrier (Kobiler et al., 2012). Besides this, the NPC is involved in several other functions like chromatin organization, replication-coupled DNA repair, regulation of gene expression and cell division during mitosis (Hoelz et al., 2011). During mitosis the nuclear envelope breaks down completely and the NPC is disassembled. Some of the subcomplexes are believed to remain associated and drive the reassembly of the NPC in late anaphase (Güttinger et al., 2009).

## 1.1 The nuclear pore complex (NPC)

### 1.1.1 Structure and composition of the NPC

The human NPC consists of around 30 different proteins, called nucleoporins (Nups) (Figure 1), and has a molecular mass of approximately 112 MDa in vertebrates (Ori et al., 2013; Reichelt et al., 1990). In total this complex is formed of ~1000 proteins (Ori et al., 2013) that are organized in different subcomplexes. Its structure is characterized by an eight-fold rotational symmetry (Hinshaw

and Milligan, 2003). The NPC scaffold consists of an inner ring, formed by the Nup93 subcomplex as well as a cytoplasmic and nucleoplasmic ring, formed primarily by the Nup107-subcomplex (also called Y-complex). The members of these subcomplexes are well structured and consist mainly of alpha helices and beta sheets, which form alpha solenoids and beta propellers, respectively. The overall structure has been extensively studied by combining multiple approaches, including purification of subcomplexes and determining their compositions by Western blot (WB) and mass spectrometry (MS). Confocal microscopy, immunogold labelling and super-resolution microscopy was used to locate the position of specific Nups. Cryo-electron tomography of whole NPCs, structural analysis of single Nups or subcomplexes by X-ray crystallography and single-particle electron microscopy (EM), cross-linking, and targeted mass spectrometry was used to study the overall structure. Recently with cryo electron tomography a resolution of 21.4 Å could be obtained for the scaffold structure of the NPC, and crystal structures of various individual Nups or smaller subcomplexes were fitted into it.

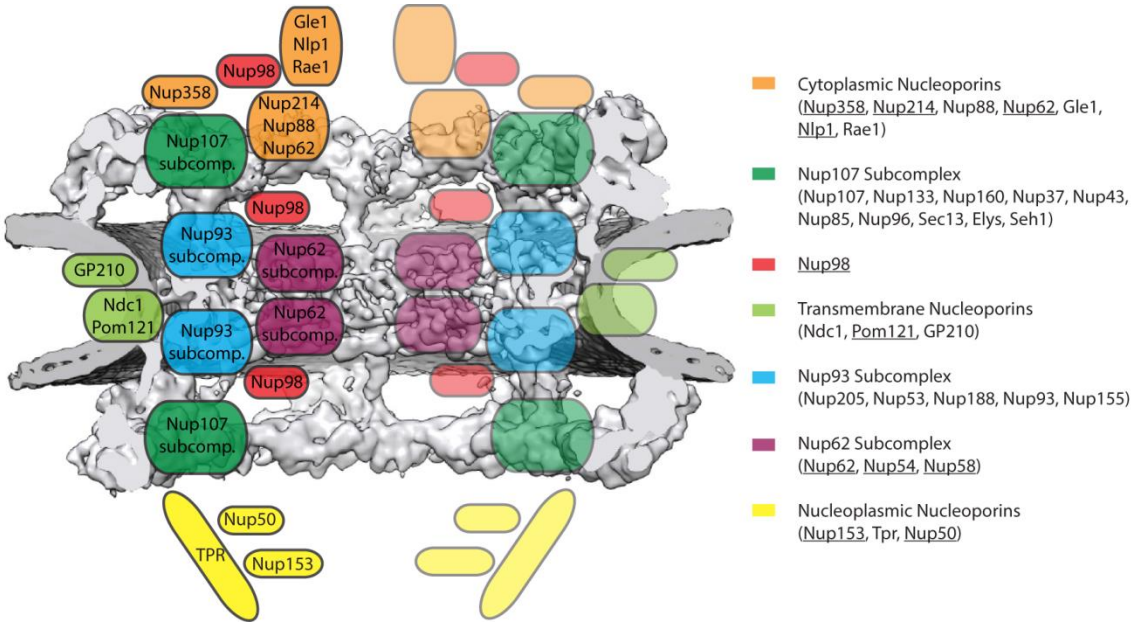


**Figure 1: Nuclear pore complex with subcomplex mapped on a cryo-electron tomographic reconstruction (Protein Data Bank EMD-31-3). Nucleoporins with FG repeats are underlined.**

This scaffold is complemented by three transmembrane Nups, and FG (Phenylalanine-Glycine)-rich Nups that are intrinsically disordered and located in the central channel of the NPC. As a consequence, the inside of the NPC appears to be empty in averaged 3D EM (three dimensional EM) maps. FG-Nups facilitate the binding of transport factors and are localized within the central channel, as well as on the cytoplasmic and nucleoplasmic side of the pore (Wente and Rout, 2010).

FG-Nups consists of 5-50 repeats of FG, FxFG, GLFG, PxFG or SxFG (x stands for any amino acid) connected by linkers of various lengths. These FG repeats are distributed within the FG-Nups in patches and are partly enriched for certain types of repeats (Figure 2). FG-Nups not only consist of FG repeats but also of structured, folded domains which are necessary for proper anchoring in the NPC or to interact with other proteins, like the Ran binding domains of Nup358 or Nup50. FG-Nups form a permeability barrier that prevents the passive diffusion of the vast majority of large macromolecules. In contrast, FG-Nups offer low affinity binding sites for NTRs allowing them to enter the NPC (Cook et al., 2007) and thus providing a gate for nucleocytoplasmic transport. NTRs are the main class of proteins assisting the nuclear import and export of macromolecules and complexes by recognizing specific peptide sequences of cargos that are called nuclear localization signals (NLSs) or nuclear export signals (NESs) depending on the respective destination. This enables a fast and directed transport of cargo to and from places of activity. NTRs make the connection between cargos and FG-Nups and allow the passage through the permeability barrier.

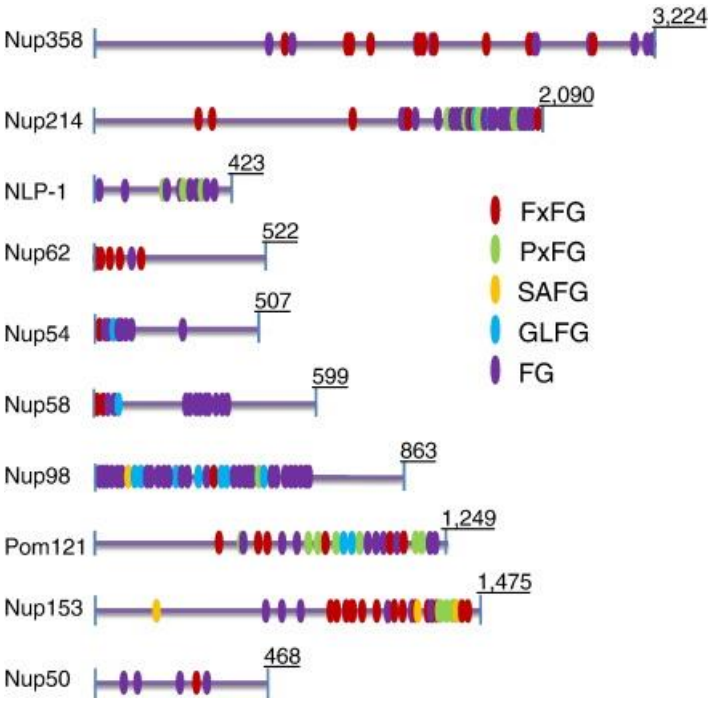


**Figure 2: Distributions of FG repeats in different FG-Nups (image reproduced from (Lemke, 2016)).**

Inside the central channel, which has a diameter of around 40 nm, the concentration of FG-Nups is around 1 mM (Lemke, 2016). Different models are proposed for the organization of FG-repeats and their interaction with NTRs (Wälde and Kehlenbach, 2010). The four most prominent models are briefly described next. The hydrogel model is defined by interaction of FG-Nups and the formation of a sieve-like structure. The polymer brush model does not imply constant interaction of FG-Nups but transient interactions based on thermodynamics. The FG-Nups form an entropic barrier, especially for larger macromolecules, and complexes of NTRs and cargo overcome this barrier by binding to the FG repeats. The model of trees and bushes is based upon collapsed and non-collapsed FG-Nups and proposes the formation of different zones within the central channel.

The forth model implicates the NTRs at the rims of the central channel, thereby reducing the dimensionality by a constant collapse of the unstructured FG domains that interact with NTRs. Although these models might explain some of the experimentally observed properties of the nuclear transport system, they also have some shortcomings. For example, none of these models offers an explanation why some FG-Nups localize asymmetrically to either the cytoplasmic or nucleoplasmic side of the NPC. Therefore, the exact conformation of FG-Nups and the transport mechanisms remain an area of active research and are further investigated, e.g. using various types of fluorescence microscopy (Grünwald et al., 2011).

### 1.1.2 Compositional changes of the NPC across cell types and differentiation

Various studies have demonstrated that NPCs dynamically change depending on developmental states, cell types, health status and many more. NPC composition is thereby modulated in time and space. Nucleoporins that are important for the scaffold structure of the pore remain more stably associated with the NPC over several hours in tissue culture cells and have extremely slow turn-over in differentiated cells (D'Angelo et al., 2009; Savas et al., 2012). For example, the scaffold nucleoporin Nup205 turned over only partly in rat brain in a time period of twelve months (Savas et al., 2012). In contrast, peripheral Nups as e.g. Nup153 and Nup50 that are involved in the translocation of cargos, have mean residence times at the NPC in the seconds to minutes range in tissue culture cells (D'Angelo et al., 2009; Rabut et al., 2004). Targeted proteomics was used to investigate the extent of stoichiometric remodeling of the NPC across different cancer cell lines, tissues and diseases compared with mRNA expression data (Ori et al., 2013). The picture that emerges is that a very static nuclear pore scaffold is decorated with more dynamically associated transport channel Nups that adapt NPC composition to context-specific needs (Figure 3). In case of disease, increased or decreased transport competences might cause for example incorrect nucleocytoplasmic transcription factor distributions.

Various mutations of nucleoporins in different organisms have been linked to tissue-specific disorders or developmental impairments (Capelson and Hetzer, 2009). For example, it has been shown that Nup133 has an important function in cell differentiation in mouse embryos. In different developmental stages and cell types, varying expression levels of the Nup133 mRNA were detected by *in situ* hybridization. The absence of this Nup does not impair the assembly of the NPC and the expression of other nucleoporins (Lupu et al., 2008). However, the authors showed that the expression of Nup133 is necessary for the differentiation of embryonic stem cells *in vitro* and the

formation of fully functional neurons during embryonic development, causing embryonic lethality (Lupu et al., 2008). For Nup50 deficient mouse embryos similar characteristics were observed (Lupu et al., 2008), but in contrast these embryos survived up to birth, which is an indication for a distinct function in development. Another example of the importance of changes in the nuclear pore complex on differentiation is Nup210 (D'Angelo et al., 2012). In normally proliferating myoblasts and embryonic stem cells Nup210 is not present at the NPC but it is expressed during differentiation.

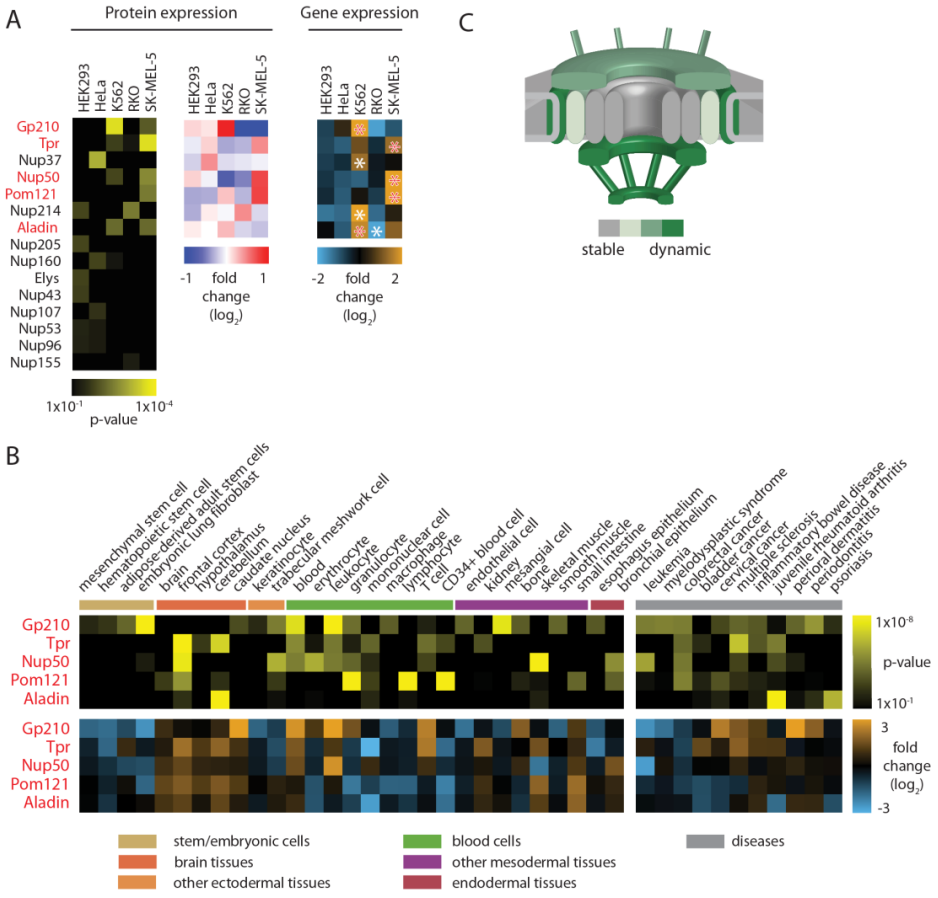


**Figure 3: (A) Nucleoporin abundance measured across 5 different cell lines using targeted mass spectrometry. The Nups highlighted in red showed similar results in mRNA expression data. (B) mRNA expression of selected Nups across many cell types and diseases show different expression pattern. (C) Scaffold Nups are more stable than Nups located either on the cytoplasmic or nucleoplasmic side as well as transmembrane Nups (modified from (Ori et al., 2013)).**

The differentiation of myoblasts into multinucleated myotubes is prevented by Nup210 gene silencing and rescued by transfection with a plasmid encoding a Nup210-GFP fusion that is resistant to RNAi. Additional experiments showed that the depletion of Nup210 did not affect nucleocytoplasmic transport indicating that myogenesis is not regulated by major changes in the transport function of the NPC. Conversely Nup210 knock down had a direct impact on gene

expression. Using DNA microarrays, 255 genes were found to be differently expressed upon Nup210 depletion, with 75% of the effected genes being down regulated, which is an indication for a function of Nup210 in gene expression activation.

### 1.1.3 Functions of Nucleoporins away from the NPC

Besides being an integral part of the NPC, Nups are also found in the nucleoplasm or cytoplasm. In embryonic *Drosophila* cells the nucleoporin Nup93 is located also within the nucleoplasm. In the nucleoplasm Nup93 is bound to active genes and stimulates their expression, while at the NPC Nup93 interacts with intergenic regions or only genes of average expression levels (Kalverda and Fornerod, 2010). Different *Drosophila* nucleoporins, including Nup98, Nup50 and Nup62, were shown to interact with overlapping sets of genes in the nucleoplasm, and to be able to influence their expression (Kalverda et al., 2010). Interactions of Nups with chromatin however also occur in an NPC-associated manner. It has been proposed that NPCs form heterochromatin exclusive zones at the NE, meaning underneath the NPC euchromatin is enriched. Close to the lamina heterochromatin is found, but near NPCs the chromatin is more open, which is an indication for a role of the NPC in this process (Capelson and Hetzer, 2009). The nucleoporin TPR has an important function in establishing decondensed chromatin next to the NPC (Krull et al., 2010). TPR gene silencing caused condensed chromatin to occur close to NPCs. In case of Nup93, interactions with chromatin spread over the whole genome were detected using chromatin immunoprecipitation (ChIP). Histone deacetylase inhibitors (HDACi) treatment significantly changed these bindings sites due to changes in the overall genomic organization (Brown et al., 2008). The binding positions of Nup93 were enriched for sites with higher transcriptional activity after the treatment, an indication for a large-scale reorganization of chromatin in the nucleus with effects on gene expression and NPC recognition. The Nup107 subcomplex stays assembled during mitosis and it had been shown to target kinetochores (Zuccolo et al., 2007). Absence of the Nup107 subcomplex led to chromosome congression and segregation defects.

Nups also occur as fully assembled NPCs in the cytoplasm of oocytes, embryonic and highly transformed cells. They are present in subcellular structures called annulate lamellae (AL), which consist of stacks of ER (endoplasmic reticulum) sheets containing NPCs (Kessel, 1992). Their exact function in the cells remained unclear so far. For more details see section 3.2.1.

### 1.1.4 Nucleoporin ALADIN and the Triple A Syndrome

The Nup ALADIN, only present in higher eukaryotes (Cronshaw et al., 2002), has a size of around 60 kDa (546 amino acids) and is characterized by WD repeats which likely form a seven bladed beta propeller, although a high resolution structure is not yet available. The exact localization of ALADIN within the NPC remains still unknown. So far it has only been shown that NDC1, a transmembrane Nup, directly interacts with ALADIN and is also required for its localization at the NPC (Yamazumi et al., 2009). RNAi experiments targeting ALADIN also showed a mislocalization of NDC1 (Kind et al., 2009). Colocalization experiments showed that ALADIN is rather localized at the cytoplasmic side (colocalization with Nup358 at the cytoplasmic face) than at the nucleoplasmic side (no colocalization with TPR) (Cronshaw and Matunis, 2003).

ALADIN was the first Nucleoporin linked to a genetic disease, namely the Triple A syndrome (achalasia-addisonianism-alacrima syndrome) (Tullio-Pelet et al., 2000). This disease is characterized by difficulties in swallowing due to smooth muscle fibers which cannot relax (achalasia), reduced production of steroid hormones (addisonianism) and lack of tears (alacrima) (Allgrove et al., 1978). Patients suffer from also neurological symptoms (Clark and Weber, 1998). Triple A syndrome can be caused by several mutations (missense, nonsense, frameshift mutations or mutations affecting the splicing sites) in the ALADIN gene (Handschug et al., 2001). Some of these mutations result in a mislocalization of the ALADIN protein from the nuclear pore complex to the cytoplasm (Cronshaw and Matunis, 2003), but seem to have no effect on the morphology of nuclei or nuclear envelopes. Different studies also showed the involvement of ALADIN on nuclear import through the importin-α/β pathway (Hirano et al., 2006). A specific mutation causes the localization of ALADIN into the cytoplasm and proteins like aprataxin and DNA ligase l cannot be imported in the nucleus anymore. These missing proteins lead to an accumulation of damaged DNA in the nucleus which might cause cell death. The generation of a mouse model lacking the gene for ALADIN did not cause any detectable typical symptoms of the Triple A syndrome, female mice seemed to be infertile (Huebner et al., 2006). Recent studies also showed that the down-regulation of ALADIN causes hypersensitivity to oxidative stress (Prasad et al., 2013) and mislocalization of Aurora A, a mitotic kinase, which causes errors in mitosis (Carvalhal et al., 2015).

## 1.2 Nuclear transport receptors (NTRs)

While molecules up to a size of ~40 kDa or ~ 5 nm can freely diffuse through the channel of the NPC, bigger molecules up to a diameter of 39 nm (Panté and Kann, 2002) are actively transported by a process involving transport factors which recognize nuclear localization signals (NLS) as well as nuclear export signals (NES). While nucleoporins are referred to as the stationary phase of the nucleocytoplasmic transport system the soluble (mobile) phase consists of the GTPase Ran, a number of auxiliary and RNA transport factors as well as nuclear transport receptors (NTRs) that transiently interact with FG-repeats but also explore the open space of cytoplasm and nucleoplasm to recruit cargos. The transport capacity of NPCs is estimated to reach 1000 translocation events and up to 100 MDa per second (Ribbeck and Görlich, 2001). The human genome encodes 20 NTRs of the importin β and 7 NTRs of the importin α family (Figure 4, see also Table 7 for alternative names). The NTRs NTF2, the RAN recycling factor, and NXT1, NXT2 and NXF1, responsible for RNA export, are additional NTRs with a broad cargo spectrum. They have different structural features than the class of importin α and β and operate as a homodimer or heterodimer in the nucleocytoplasmic transport process.

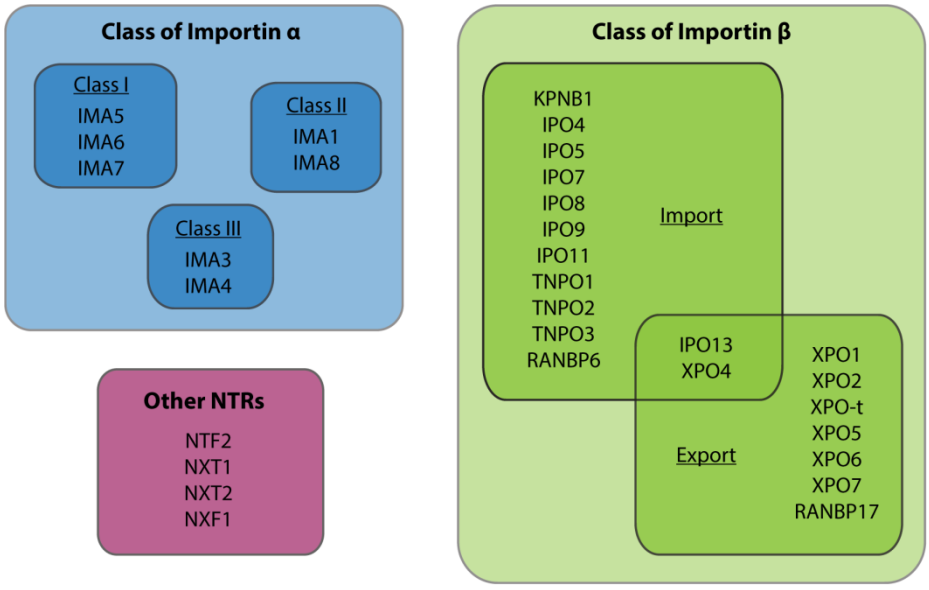


**Figure 4: Nuclear transport receptors are structurally classified into importins α (ARM repeats), with 3 subclasses based on sequence similarities, and importin β (HEAT repeats) which includes NTRs capable of import, export or both. NTF2, NXT1, NXT2 and NXF1 have different structural features and are not classified into the previous two classes.**

The small GTPase Ran fuels the active nuclear transport cycle (see Figure 5, part of recycling). Ran exists in a GTP or GDP bound form and 2 cofactors facilitate their inter-conversion. RCC1, also

known as RanGEF (Ran guanine nucleotide exchange factor), is localized within the nucleus by binding to chromatin and favors the exchange of RanGDP to RanGTP. The antagonist of RCC1 is RanGAP (RAN GTPase-activating protein) in the cytoplasm and triggers the hydrolysis of RanGTP to RanGDP, which is transported back into the nucleus by the RanGDP recycling factor NTF2. This gradient of RanGTP/RanGDP enables the directionality of the nucleocytoplasmic transport. The binding to RanGTP triggers the disassembly of import complexes at the nuclear face of the NPC (Izaurralde et al., 1997). RanGTP and importin-β form a recycling complex to be exported back into the cytoplasm. Ran is also a member of ternary complexes formed together with α-type importins, such as the importin-α-XPO2-RanGTP recycling complex and cargo-exportin-RanGTP complexes (Figure 5). The disassembly of each of the three aforementioned complexes is triggered by GTP hydrolysis once they traveled back to the cytoplasmic face of the NPC. The driving force of Ran allows cargo distribution also against their concentration gradients. However, also Ran-independent transport pathways have been reported (Kotera et al., 2005; Miyamoto et al., 2002). The nuclear export of mRNA is independent of Ran (Clouse et al., 2001) and performed by NXT1/p15 (Katahira et al., 1999) in a heterodimeric complex with NXF1/TAP (Grüter et al., 1998).

The vast majority of previous studies that have reported transport pathways for individual cargos have been designed in a cargo-centric manner (Chook and Süel, 2011; Kimura and Imamoto, 2014; Twyffels et al., 2014). This means that, so far, most interacting pairs were found by 'low-throughput' studies using the cargo as a starting point. Almost all cargo-NTR relationships were detected in different experiments with strongly varying experimental conditions e.g. the cell types and detections methods. This makes it very difficult to define cargo classes and perhaps agreements and relationships within them. Also the localization sequence is often not known. Only for a very limited set of NTRs, systematic approaches to measure their cargo spectrum in an NTR-centric manner have been designed. In permeabilized cells, the rate of import of SILAC-labeled cellular extracts has been measured mass spectrometrically in the presence and absence of transportin to chart its cargo spectrum (Kimura et al., 2013a). In case of exportin 1, its susceptibility to inhibitors (Nishi et al., 1994) and dependency on RanGTP for export complex formation have been exploited to measure their cargo spectrum. However, these experiments were conducted under non-physiological conditions or *in vitro* and, more importantly, are not generalizable to all NTRs. Notably, NPCs also facilitate RNA export from the nucleus, which might be studied e.g. by fluorescence in situ hybridization (FISH) in fixed or living cells (Weil et al., 2010) but is not subject to this thesis.

### 1.2.1 Importin alpha family

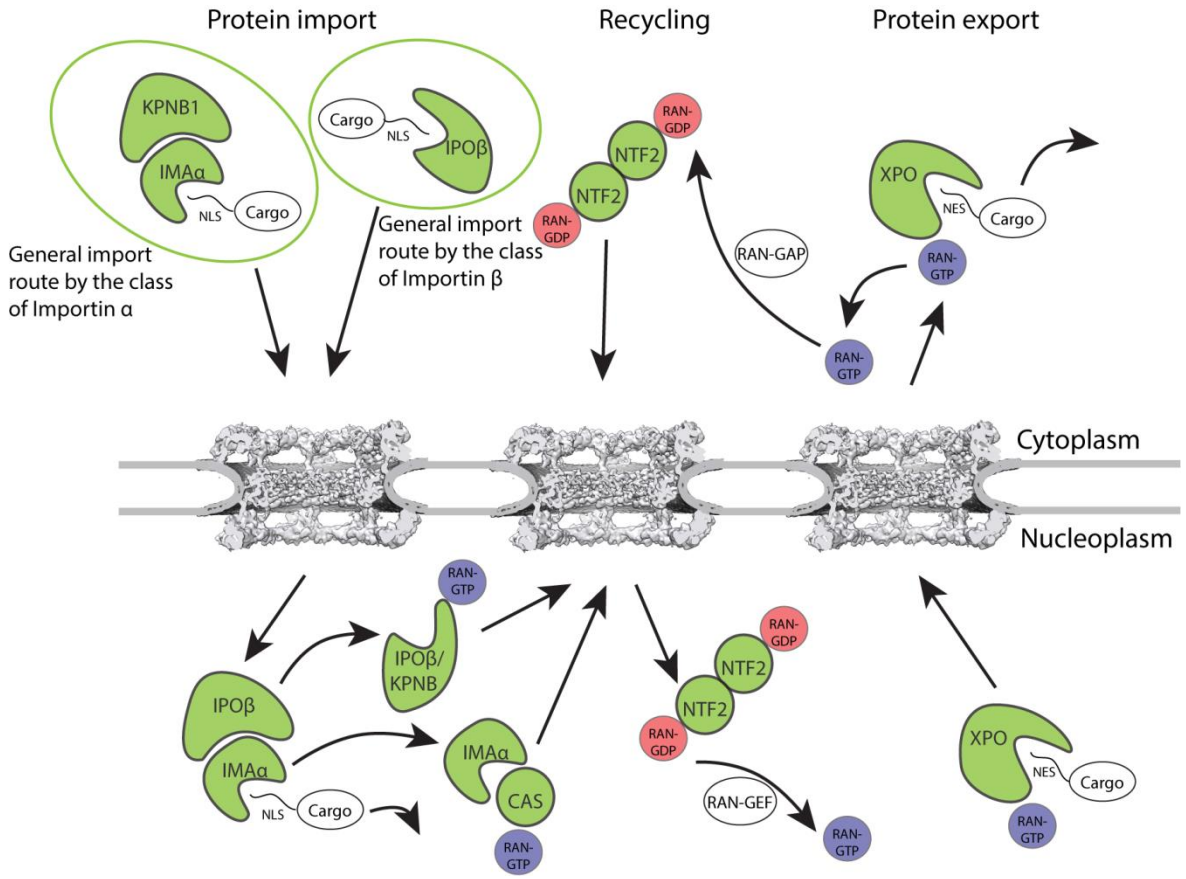


**Figure 5: Overview of the nucleocytoplasmic transport routes used by nuclear transport receptors. Binding of cargo can happen directly via importin ß and exportin or through importin α as an adaptor. Recycling of Ran-GDP is performed through a dimer of NTF2. Exportin 2 (CAS) recycles importin α back to the cytoplasm.**

NTRs of the karyopherin α family (also called importin α) are part of the so called classical import pathway. Three clades of importins of the α-type occur in higher eukaryotes (Figure 4) and humans possess at least seven different importin α proteins (Kelley et al., 2010). Different subfamilies share around 50% sequence identity and within subfamilies it is around 80% (Köhler et al., 1999). Although they are differentially expressed across tissues and during development (Yasuhara et al., 2009), they all bind classical NLSs (cNLSs) and are thought to be functionally redundant, at least to quite some extent (Goldfarb et al., 2004). They often but not always (Kotera et al., 2005) function as adaptors and form a ternary import complex together with their cargo and importin β (Görlich et al., 1995). Each importin α consists of ten ARM (armadillo) repeats, first described in *Drosophila melanogaster* (Riggleman et al., 1989), which are around 40 amino acids long and composed of three α-helices.

These ARM repeats are the site of interaction with the NLS of cargos. The first NLS was discovered in SV40 T antigen (Kalderon et al., 1984) and consists of one cluster of basic amino acids. This type of NLS is known as monopartite NLS. A more complex NLS consisting of 2 stretches of basic amino acids connected by a 10 to 12 amino acid long linker is termed bipartite NLS and was first discovered in nucleoplasmin in *Xenopus laevis* (Dingwall et al., 1988). Still there are many more variants of NLS which also show different interaction preferences with the ARM domain of importin α (Kosugi et al., 2009). Some known cargos do not seem to possess a NLS in their sequence and only structural analysis of the protein revealed potential recognition sites, so called conformational NLS (Sessler and Noy, 2005). This makes the prediction of potential cargos extremely difficult. A second important feature of importin α is the importin β binding (IBB) domain located at the N-terminus.

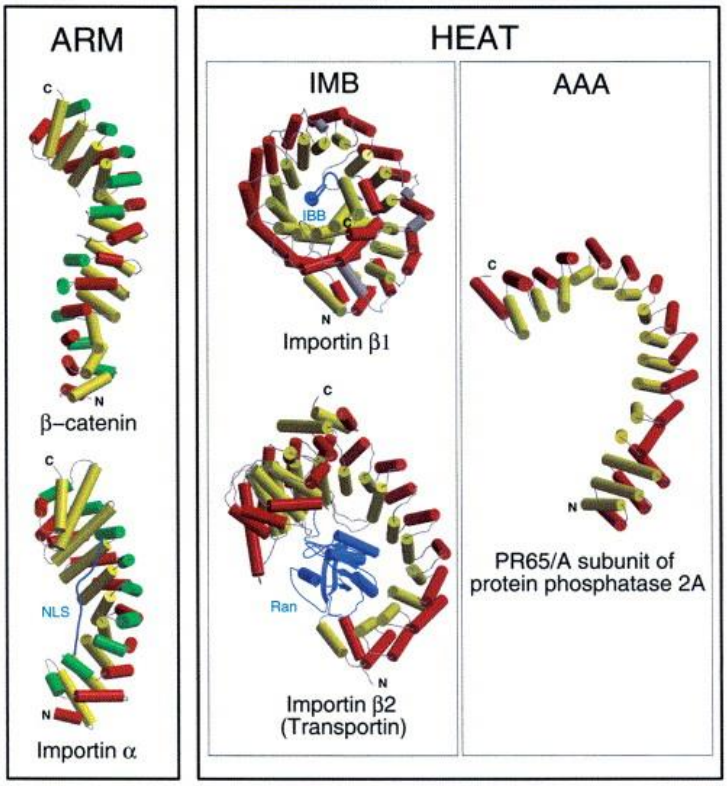


**Figure 6: Proteins containing ARM or HEAT repeats typical for the class of Importin α and Importin β, respectively (image reproduced from (Andrade et al., 2001)). Importin α is bound to an NLS, importin β1 to an IBB domain and Transportin to Ran.**

The IBB domain has a similar sequence to cNLS and also interacts with the NLS binding pocket on the ARM repeats in the absence of cargo. Therefore, the binding of importin α to importin β has to occur before the NLS of a cargo can be recognized. The release of the cargo in the nucleus happens through binding of RanGTP (Lee et al., 2005). Importin β is separated from importin α and both are

transported separately in complex with RanGTP back into the nucleus. Importin α additionally needs Exportin 2 (XPO2/CAS) for a successful relocalization into the cytoplasm (Kutay et al., 1997). Some studies showed Nup50 to be involved in the release of cargo as well, by accelerating the disassembly of the import complex by displacing the cargo from importin α (Matsuura et al., 2003; Matsuura and Stewart, 2005).

### 1.2.2 Importin beta family

The class of importin β share 19 to 20 characteristic HEAT (**H**untingtin, **e**longation factor 3, protein phosphatase 2**A**, **T**OR1) repeats (Andrade and Bork, 1995). Each HEAT repeat consists of 40-45 amino acids which folds into two anti-parallel α-helices and form a superhelical fold. The sequence similarity between the different importin β members is rather low but they all share a similar organization and molecular weight (Quan et al., 2008). HEAT repeats are more flexible than ARM repeats. All importin β are capable of binding cargo directly, but importin β1 (KPNB1) functions mainly as an adaptor protein for cargo loaded importin α. Depending on the directionality of cargo transport they are referred to as importins and exportins, and facilitate cargo import into or export from the nucleus, respectively. But this strict subdivision does not hold true for all of them. For example, Importin 13 (IPO13) functions mainly as an importin but it also has been shown to have export capacity for the translation initiation factor eIF1A (Mingot et al., 2001). Also Exportin 4 (XPO4) has a bidirectional transport capability (Gontan et al., 2009). It has been shown that besides its export capability, it conducts the import of transcription factors relevant in development, SRY and Sox2. A specialized member of the β-karyopherin family is the cellular apoptosis susceptibility (CAS/XPO2) protein that is a recycling factor for α-karyopherins (Kutay et al., 1997). Many members of the class of importin β are little studied and only few cargos are known (Chook and Süel, 2011). Transportin-1 (TNPO1, also known as Kapβ2 or TRN-1) and Transportin-2 (TNPO2, also known as Kapβ2B or TRN-2A) are an exception. Transportins consist of 20 HEAT motifs and have a characteristic, 62 amino acid long, acidic loop within the eighth HEAT repeat. This acidic loop has an important function in cargo release inside the nucleus. After biding of RanGTP the conformation of Transportin changes and the loop pushes the cargo out of its binding pocket (Chook et al., 2002). Several cargos have been experimentally validated (Twyffels et al., 2014) and also a specific type of NLS, the so-called PY-NLS, has been defined. The general motif consists of a hydrophobic or basic stretch of amino acids followed by R/H/K$X_{(2-5)}$PY (Lee et al., 2006). Many proteins with PY-NLS are related to RNA, either by binding, transcription or processing. Most of the proteins with a PY-NLS are

exclusively transported by Transportin. Other proteins with classical NLS or unknown signal sequences transported by Transportin tend to employ other importin β pathways as well, like various ribosomal proteins (Jäkel and Görlich, 1998) or histones (Baake et al., 2001).

The exportins, a subclass of karyopherin β, bind their cargos in a Ran dependent manner and release it through hydrolysis of RanGTP to RanGDP. One of the best studied members is Exportin 1 (XPO1, also called CRM1). It has a wide cargo spectrum and recognizes a leucine-rich nuclear export signal (NES) (Fornerod et al., 1997) which was first identified in the HIV-1 Rev protein (Fischer et al., 1995). Leptomycin B, an antifungal antibiotic isolated from *Streptomyces* (Hamamoto et al., 1983), is a specific inhibitor of XPO1 (Kudo et al., 1998). This inhibitor was utilized to identify more potential cargos of XPO1 (Thakar et al., 2013) but it also affects the cell cycle severely (Lu et al., 2012), which prevents longer cell treatments or applications in medicine.

All other exportins, except XPO2/CAS, the recycling factor for Importin α, are studied to a much lesser extend and only few cargos are known so far.

### 1.2.3 NTRs and cell differentiation

It has been hypothesized that the composition of the FG network might vary between different cell types, probably on account of the presence of different transport factors (Capelson and Hetzer, 2009) that recognize the nuclear import or export signals of cargos (Wente and Rout, 2010). Transport factors might thus modulate the nuclear transport activity in a cell-type specific manner and are involved in key events during development and cell differentiation (Okada et al., 2008; Yasuhara et al., 2009).

When mouse embryonic stem cells differentiate into neuronal cells, importin α expression changes its subtype from α2 to α1 (Yasuhara et al., 2013). These subtypes of importin α regulate the import of various transcription factors important for cell fate and can therefore indirectly regulate the differentiation process. To keep embryonic stem cells in an undifferentiated state the transport of the transcription factor Oct3/4 into the nucleus is necessary and executed by Importin α2. Oct6, an important transcription factor to drive neuronal differentiation, is capable of binding Importin α2 but at a position which inhibits nuclear transport. Only after the expression of Importin α1 Oct6 and other transcription factors are imported and neuronal differentiation proceeds. Similar results point to an involvement of the importin β family (Sangel et al., 2014), whose mRNA level varies across

different cell types. The effects of knocking down several NTRs in mouse embryonic stem cell (mESC) differentiation were analyzed. For example, the knockdown of Importin 7 inhibited the differentiation into neural ectoderm, but the formation of other lineages remained unaffected. Many cases of differentially expressed importins have been experimentally detected so far and have been linked to various developmental stages (overview in (Kimura and Imamoto, 2014)).

Since the tight regulation of developmental signaling is crucial, possible dysregulation of NTR expression might lead to malignant transformation (Winkler et al., 2016). Indeed various clinically manifested mutations map to components of the nuclear transport system, and malfunctioning transport pathways have been suggested to cause human diseases (Kimura and Imamoto, 2014).

### 1.2.4 Regulation of cargo binding

The adjusted expression of NTRs is one way to regulate the nucleocytoplasmic exchange. Another level of regulation is achieved by altering the binding of NTR and cargos (Christie et al., 2016) through posttranslational modifications (PTMs) like phosphorylation (Nardozzi et al., 2010), acetylation (Spilianakis et al., 2000), methylation (Oudhoff et al., 2013), ubiquitination (Chen and Mallampalli, 2009) and sumoylation (Truong et al., 2012). PTMs on or close to a NLS can increase or decrease the binding affinity to a NTR. Pho4, a transcription factor of *Saccharomyces cerevisiae,* has multiple sites of phosphorylation (Komeili and O'Shea, 1999). One phosphorylation site, located within the NLS, prevents its nuclear import. Two other sites, just upstream of the NLS, promote the nuclear export of Pho4. Sometimes PTMs are also necessary to enable access to a buried NLS. The human Hepatits B virus (HBV) depends on the RNA polymerase II of the host and therefore has to enter the nucleus. Phosphorylation of a core protein of HBV induces a conformation change and exposes a NLS which is then recognized by importin α1 (Kann et al., 1999; Rabe et al., 2003). Another example of an intra-molecular masked NLS is the STAT1 (Signal Transducer and activators of Transcription) dimer (Meyer et al., 2002). Phosphorylation leads to a different conformation of the STAT1 dimer, exposing a dsNLS (dimer specific NLS). Not only conformational limitation within a protein mask a NLS, but also interactions with other proteins or DNA might limit the accessibility of a NLS. The transcription factor dimer NF-κB is in a complex with IκB which retains the complex in the cytoplasm. Upon phosphorylation, IκB is degraded and NF-κB is transported amongst other by importin α3 and α4 into the nucleus (Fagerlund et al., 2005; Oeckinghaus et al., 2011).

PTMs not only occur on cargos but also influence and regulate nucleocytoplasmic transport by modifying NTRs (Figure 7). For example acetylation on a specific Lysine of Importin α1 enables the binding to importin β (Bannister et al., 2000) or the nuclear import of the RNA binding protein HuR depends on acetylation as well as phosphorylation on importin α1 (Wang et al., 2004).

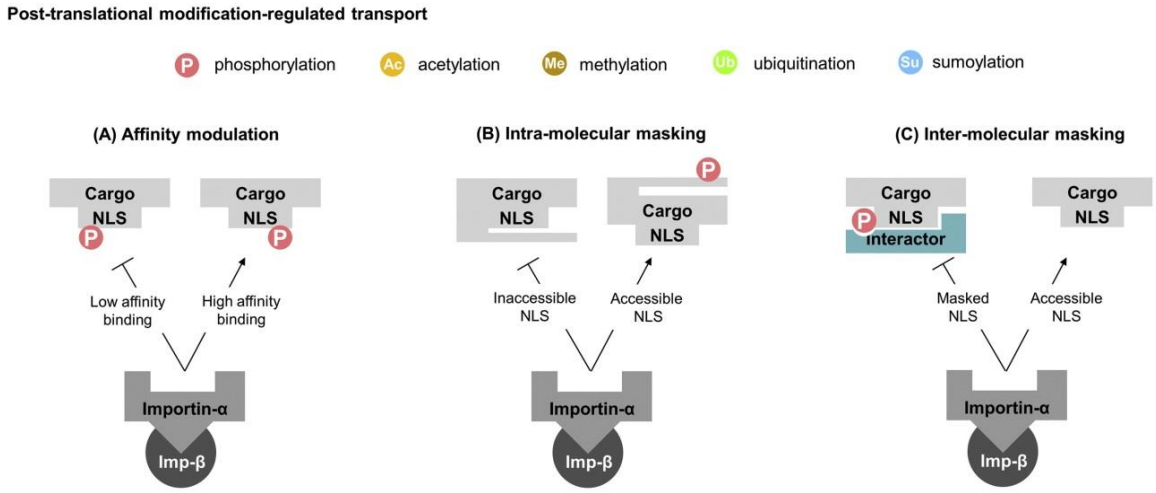


**Figure 7: Different ways to regulate the nucleocytoplasmic transport through post translational modifications (modified from (Christie et al., 2016)). See text for more details.**

Alternative mechanisms for a regulated nucleocytoplasmic transport are directed transport of cargos to the nuclear envelope, transport within a bigger complex or competition for NTRs by an increase of NLSs. Viruses, like adenoviruses, use microtubules to reach the nucleus faster after entry into the cells (Leopold and Pfister, 2006). Similar strategies have also been shown for human proteins like p53 and PTHrP (parathyroid hormone-related protein) (Roth et al., 2011). A dUTPase has in different *Drosophila* strains two isoforms, one with and the other one without a NLS, containing up to three NLSs in a trimetric complex. The number of NLSs varies between one and three within the complex formed by these isoforms and significantly affects the nucleocytoplasmic distribution of these complexes (Róna et al., 2014).

## 1.3 Capturing transient interactions by combining proximity labeling and identification using Mass Spectrometry

### 1.3.1 Quantitative Proteomics

The ultimate goal of quantitative proteomics is to identify and determine the quantity of all proteins in one sample as well as their modifications and variations under different biological conditions, such as in different cell types and environments (Domon and Aebersold, 2010; Gstaiger and Aebersold, 2009). On one side the composition of one sample can be analyzed; on the other side changes in different samples can be compared.

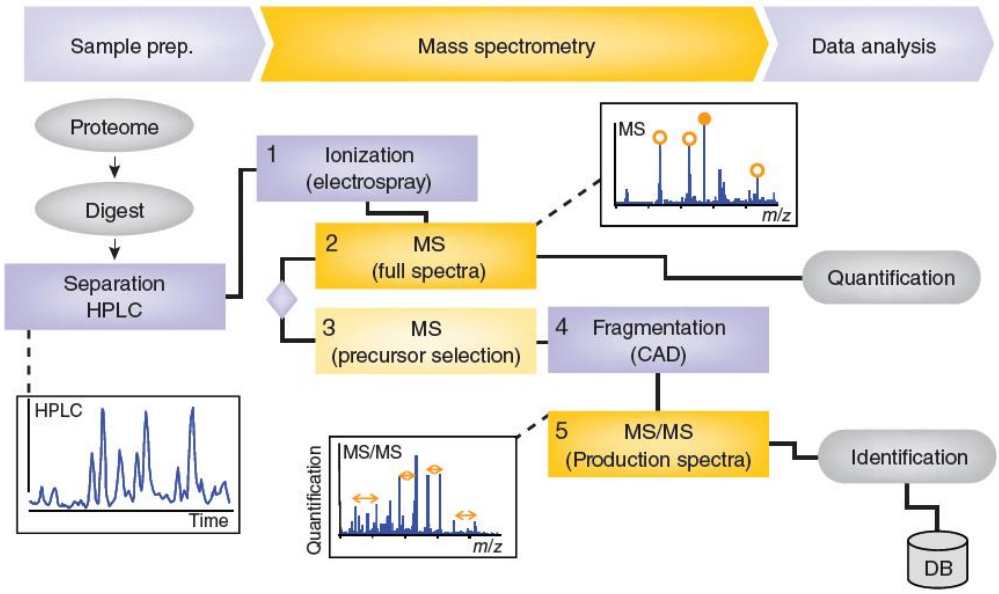


**Figure 8: Overview of the basic steps in quantitative proteomics (image reproduced from (Domon and Aebersold, 2010)).**

An overview of the basic steps in a classical bottom-up quantitative proteomics workflow is shown in Figure 8. Sample preparation begins with the solubilization of proteins generally from whole cell lysates, specific organelles like the nucleus, or affinity-purified complexes and proteins. After the enzymatic digestion of the proteins, the obtained peptides need to be separated to reduce the complexity of the sample. The standard separation method is reverse-phase liquid chromatography, which is often on-line coupled to the mass spectrometer (LC-MS). If samples are highly complex or a deeper insight into the proteins composition is necessary, additional sample fractionation at the protein or peptide level can be performed off-line by using e.g. high pH reverse-phase fractionation. After this step the analytes need to be ionized and transferred into gas phase. The most commonly

used method to achieve this is electrospray ionization (ESI), mainly because it can be directly coupled with a high performance liquid chromatography (HPLC). The liquid from the LC flows through a narrow capillary with an applied potential. At the end of the capillary the drops scatter in smaller drops and a charge is transferred onto the eluting peptides.

In tandem mass spectrometry (LC-MS/MS), also known as shot-gun proteomics, two mass analyzers are coupled. Typical instruments used for this kind of analysis are linear ion trap coupled with orbitrap (LIT-OT) or quadrupole coupled with time of flight (Q-TOF) mass spectrometers. First, the mass to charge ratios of eluting peptides are recorded in a full scan spectrum (MS1) (Cañas et al., 2006; Yates et al., 2009). From this spectrum precursor ions with the highest intensity are selected, isolated and fragmented. Fragmentation can be obtained by collision activated dissociation (CAD), also referred to as collision-induced dissociation (CID). During this process ions collide with neutral gas molecules like Helium, Neon or Argon. Fragmentation generally occurs at peptide bonds generating so called b- and y-ions. For each selected precursor ion, the mass to charge ratios of fragment ions are recorded in a second (MS2) spectrum subsequently to isolation and fragmentation. The combined information from both MS1 and MS2 spectra is used to identify peptides and the modifications they might bear, by searching against a database of protein sequences.

Quantification of relative amounts of protein might be obtained based on isotope labeling, or alternatively, by label-free techniques, which allow quantification of any protein without limiting it to labeled compounds and do not produce an increase in sample complexity introduced by isotopes. In this case samples that are to be compared are processed separately and the output data is used for comparison. The quantification itself might be based on spectral count or, better, extracted precursor ion chromatograms on the MS1 level (Zhu et al., 2010). Inaccuracies are caused by differences between sample preparation and injection as well as shifts in the retention times due to variable interaction strength of peptides with the column. Due to variations between different runs, data normalization is required and several replicates need to be analyzed in order to achieve reliable quantification.

### 1.3.2 Proximity labeling using BioID

Classical affinity purifications are well suited to detect specific and direct interactions but have weaknesses for the discovery of transient and weak interactions, like the ones that occur between

NTRs and cargos. Additionally, the cargo spectrum might be broad and thus only highly abundant interactions partners can be detected. Therefore to chart the relatively transient interactions occurring between FG-Nups, NTRs and cargos the use of a system called BioID (Roux et al., 2012), which is depicted in Figure 9, is highly suitable. The system is based on a biotin ligase (BirA) from *Escherichia coli*, which biotinylates a subunit of the acetyl-CoA carboxylase (Chapman-Smith and Cronan, 1999). The original biotin ligase has only one target and the release of the reactive biotinyl-5'-AMP only occurs after successful recognition of the active site. By introducing a R118G mutation in the biotin ligase the interaction between the biotin ligase and biotinyl-5'-AMP is significantly reduced (Kwon and Beckett, 2000) and the biotinylation is not restricted to one target anymore (Choi-Rhee et al., 2004). Although being less specific than classical affinity purification, the system has been proven to capture also transient interactions (reviewed in (Varnaitė and MacNeill, 2016).

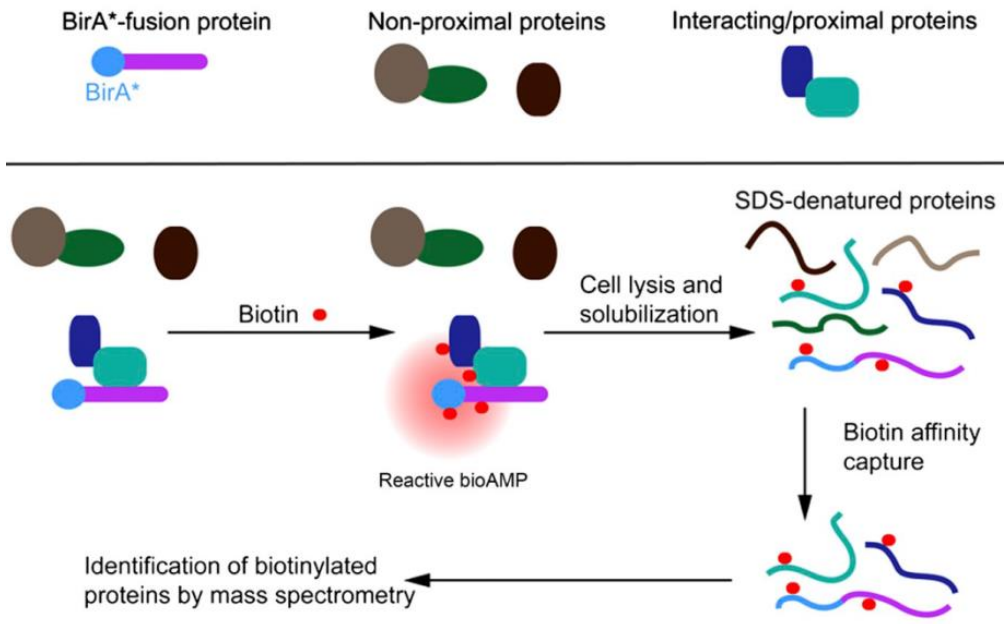


**Figure 9: Overview of the basic principle of the BioID system (modified from (Roux et al., 2012)).**

In stable cell lines the expression of a protein of interest fused to the 35 kDa mutated biotin ligase, can be induced and biotinylated proteins extracted for analysis. The BioID system offers various advantages compared to other methods. First, the labeling of proteins in proximity to the protein of interest is performed *in vivo* (or *ex vivo*) and except for stable over expression of the BirA construct or transiently transfecting the cells no additional interference is necessary. Second, the method is well suited for transient interactions or proteins in close proximity. Not only direct interactions, but also interactions in close proximity are detectable. The labeling radius was first estimated to be 20-

30 nm by fusing BirA to lamin A (Roux et al., 2012). By fusing BirA to various members of the NPC, especially the Y-complex with excellent knowledge of structure, position and dimensions, the labeling radius was estimated to be ~ 10 nm (Kim et al., 2014). Third, the biotin stays as an identification mark on proteins also after the interaction occurred. Forth, this also allows using stringent lysis conditions for the following purifications. Fifth, the interaction of biotin with streptavidin is one of the strongest interactions in biology (Green, 1975). For purification of the biotinylated proteins, beads coupled to streptavidin are used, allowing harsh washing steps and minimizing therefore the amount of background identified in mass spectrometry (MS). Since the interaction is so strong, on-bead digestion is commonly used to identify interacting proteins by MS indirectly, i.e based on non-biotinylated peptides eluting after digestion. Thereby, an important layer of information is lost, namely the exact biotinylation sites.

Since the introduction of the BioID method in 2012 many different application possibilities have been published. The NE and NPC were studied as well as the centrosome (Gupta et al., 2015), cell junctions (Fredriksson et al., 2015) and the Hippo pathway (Couzens et al., 2013), an important cell signaling pathway.

# 2 Motivation and Objectives

The vast majority of previous studies that have reported NTR interactions for individual cargos were designed in a cargo-centric manner. Although powerful for studying the contribution of the nuclear transport system to individual cellular mechanisms, these approaches are not suitable to globally chart its cargo spectrum. The respective experiments were often conducted under non-physiological conditions *ex vivo* and more importantly, are not generalizable to all NTRs. The cargo spectrum of the vast majority of NTRs thus remains incompletely charted. It remains unknown to which extent the functions of the different importins, exportins and transportins are distinct or functionally redundant and even the extent to which the nuclear transport system contributes to nucleocytoplasmic protein localization has been questioned. The goal of this thesis was to address this challenge and to systematically determine cargo-NTR relationships in an unbiased way *in vivo* using proximity labeling mass spectrometry based on the so-called BioID system. The expression of fusion proteins of human NTRs to BirA in tissue culture cells was expected to label even transient interactors that are subsequently affinity purified and identified by mass spectrometry. Also specific NTR- FG-Nup interactions during cargo translocation should be identified and contribute to our understanding of nuclear transport pathways.

Since compositional variations of the NPC and NTR abundance might influence its transport competence, it is important to understand under which conditions they occur. In collaboration with Drs. Bernhard Hampoelz, Alessandro Ori and Amparo Andres-Pons, I have investigated potential compositional variations across cell types and developmental states. I have use quantitative mass spectrometry to analyze NPC composition in developing *Drosophila* and Zebrafish embryos.

# 3 Results

## 3.1 Changes in NPC composition upon neuronal differentiation in zebrafish

### 3.1.1 Background: Advantages of zebrafish as model organisms for studies on differentiation

Zebrafish, as a model organism, offers various benefits. Individual matings generate hundreds of eggs and can be synchronized by a light cycle. This allows the collection of many embryos at the same stage of development and guarantees enough starting material for further experiments. Additionally, their development is fast (Figure 10) and occurs outside of the mother within a clear chorion and a transparent body. Early cell divisions only take around 15 minutes, development can be easily monitored light microscopy and embryos can be sorted based on morphological criteria (Figure 10). The genome of zebrafish is completely sequenced and 70% of human protein-coding genes have orthologous in zebrafish (Howe et al., 2013).

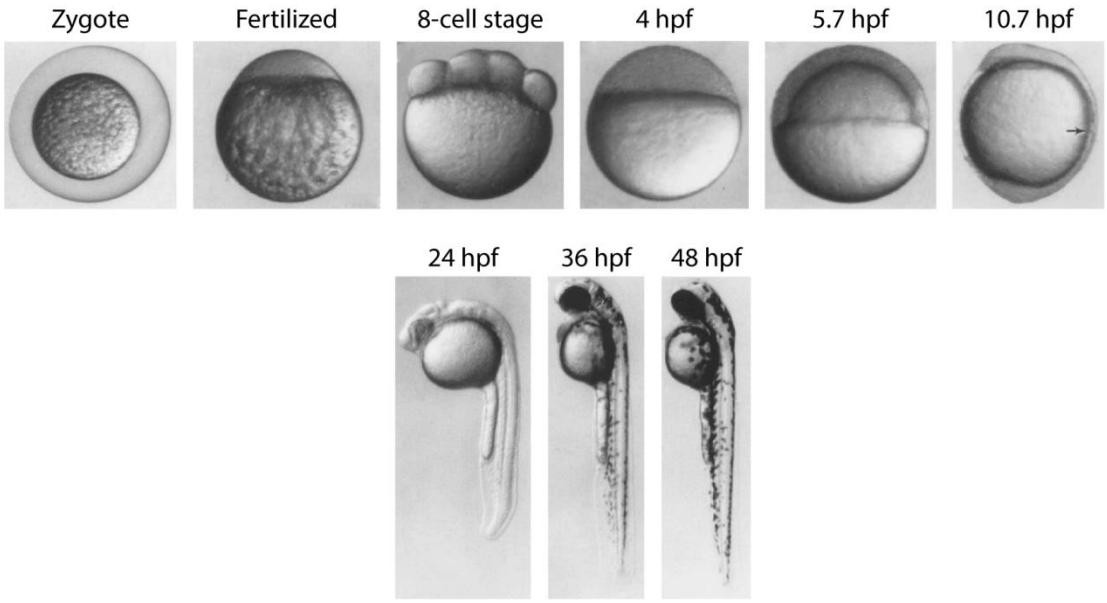


**Figure 10: Stages of embryonic development up to 48 hpf (hours post fertilization) (modified from (Kimmel et al., 1995)).**

### 3.1.2 Compositional characterization of neuronal cells

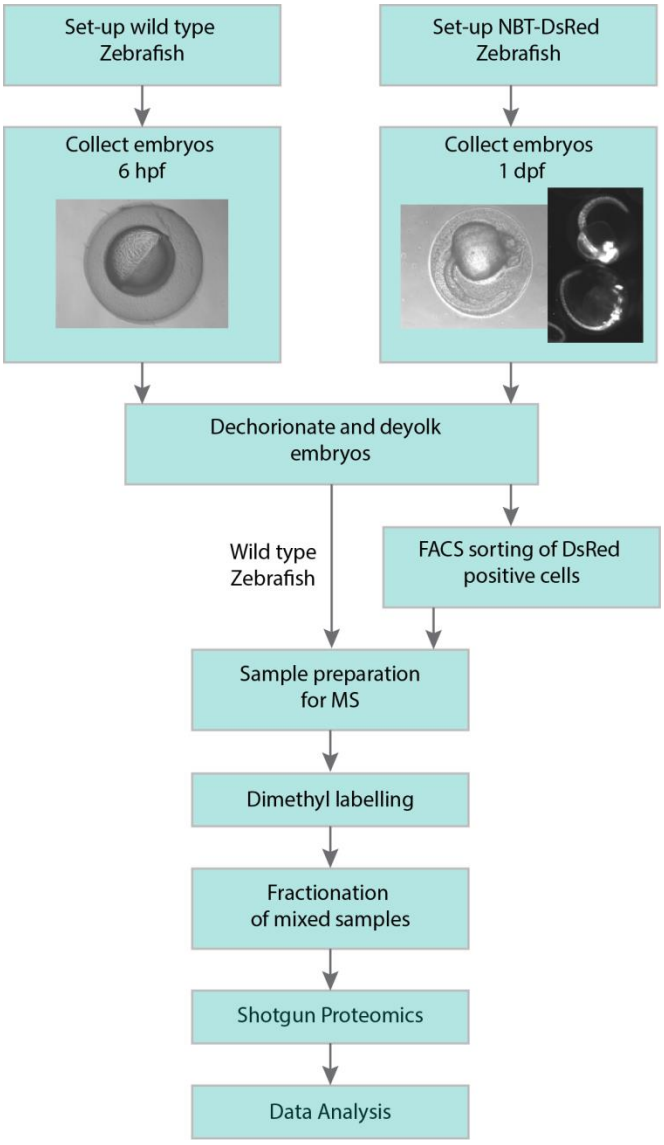


**Figure 11: Workflow for extraction and processing of undifferentiated and neuronal cells of zebrafish.**

To analyze tissue-specific remodeling of the NPC composition and the nuclear transport system in zebrafish, I co-developed an experimental workflow (together with Drs. Amparo Andres-Pons and Alessandro Ori) that allows the isolation of specifically labeled cell types (Figure 11). Undifferentiated cells were extracted of wild-type fish 6 hpf (hours post fertilization). For the collection of neuronal cells, NBT-DsRed zebrafish 1 dpf (day post fertilization) were used. In this fish strain the red-fluorescent-protein is under the control of a neural-specific tubulin promoter of *Xenopus* (Peri and Nüsslein-Volhard, 2008), which allowed the selective sorting of neuronal cells by FACS (fluorescence-activated cell sorting), getting a cell suspension highly enriched. Already 24 hpf

the neuronal development has progressed considerably and embryos that carry the neuronal marker, since the parents were heterozygous, can be sorted easily using a light microscope (see Figure 11 in which the eyes and spinal cord, due to high enrichment for neuronal cells, are clearly visible). Stable isotope dimethyl labeling (Boersema et al., 2009) was used to prepare samples for mass spectrometry. Formaldehyde and the D-isotopomer of formaldehyde react with primary amines of peptides (N-terminus and side chains of lysines) and generate a mass shift of 4 Da. The samples were mixed, fractionated and measured in parallel. Since the labeling is carried out on the peptide level it is cheap and only small amounts of sample are necessary. The samples were fractionated into 32 different fractions using high pH reversed-phase liquid chromatography (Yang et al., 2012) and pooled into 10 fractions which were then analyzed by shotgun mass spectrometry. High pH fractionation increases the dynamic range and improves the number of identified proteins and peptides by about 1.6 to 1.8 fold.

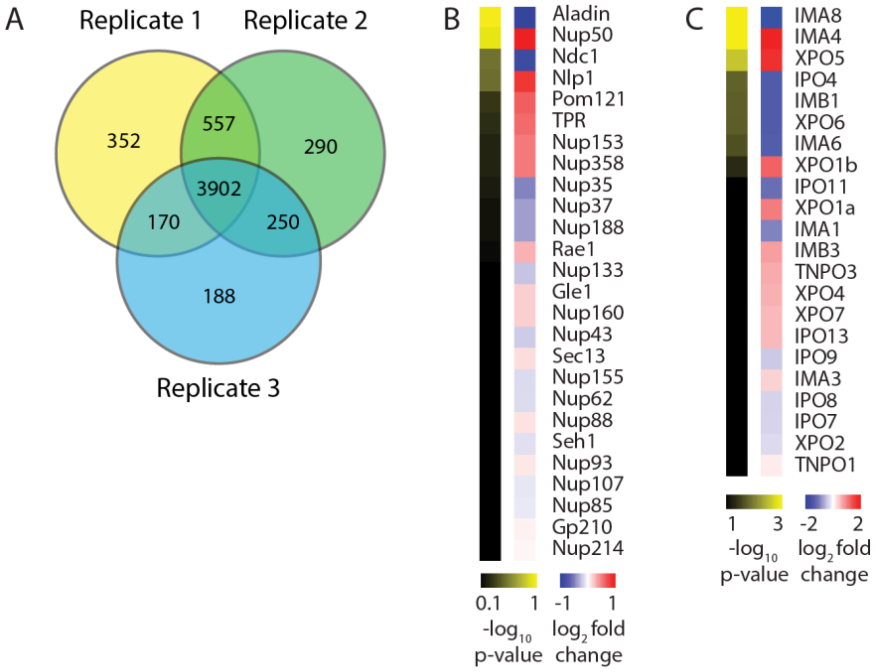


**Figure 12: (A) Venn Diagram of identified protein groups per replicate and overlap between them. (B) and (C) Heatmap of changes of nucleoporins and nuclear transport receptors during differentiation towards neuronal cells in Zebrafish. The proteins are sorted by a decreasing -log$_{10}$ p-values.**

Three biological replicates of 10 pooled fractions were analyzed, and 27000 to 34000 peptides per replicate, which corresponds to 4400 to 4900 protein groups, were identified. In total, I identified 5700 unique proteins of which almost 5000 were seen in at least 2 replicates (Figure 12).

For the analysis of the changes of the nuclear pore complex composition and the nuclear transport receptors in the data set obtained from undifferentiated and neuronal cells of zebrafish, I used a data analysis pipeline that was previously developed within the laboratory. With this pipeline stoichiometric changes, after normalization to the absolute abundance of the respective protein complex, can be identified (Ori et al., 2013). Figure 12B and C show the results of this analysis. Negative fold changes indicate enrichment in undifferentiated cells and positive fold changes enrichment in neuronal cells. ALADIN is significantly down regulated and Nup50 is significantly up regulated in neuronal cells. Also Ndc1, Nlp1 and Pom121, all belonging to the more dynamic parts of the NPC, show a similar trend. I used the same approach for the relative abundance of nuclear transport receptors (Figure 12C). In comparison to the nucleoporins, larger fold changes are detected, implying a more dynamic adaption over differentiation. 7 different NTRs have significant fold changes. IMA8, IPO4, IMB1, XPO6 and IMA6 are less abundant in neuronal cells. Instead, IMA4 and XPO5 display an increase in abundance as compared to undifferentiated cells. The top two NTRs are both importin alphas. In other species, like mouse (Yasuhara et al., 2013), a switch in the expression of NTRs has been shown to be essential for differentiation.

Differentiation requires complex changes in transcription and activation of specific regulators. During the development from undifferentiated cells to neuronal cells only dynamic Nups show different protein abundances. The exact mechanism and contribution of them to the differentiation process has to be more extensively studied in the future.

### 3.1.3 CRISPR/Cas9 knockdown of ALADIN in Zebrafish

ALADIN is one of the nucleoporins identified by our proteomic screens and has previously been linked to a genetic disease. Mutations within the coding sequence or splice sites generating proteins both properly localized at the nuclear envelope and mislocalized cause the disease Triple A syndrome. Heterozygous mutations caused no obvious defects in human (Huebner et al., 2002), but how the effect of an homozygous knockout is not known.

The nucleoporin ALADIN had the most significant and strongest fold change between undifferentiated und neuronal cells. This is consistent with the finding that Triple A syndrome is mainly caused by dysfunctions of the autonomic nervous system. I therefore wanted to study if ALADIN is required for neuronal differentiation in the zebrafish system.

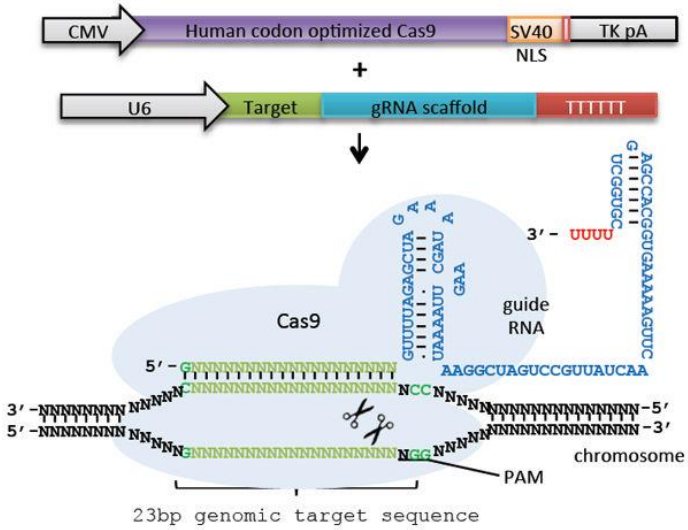


**Figure 13: Type II CRISPR System: RNA-guided gene targeting in cells coexpressing Cas9 with a nuclear localization signal (image reproduced from (Mali et al., 2013)).**

I used the CRISPR-Cas9 system (clustered regularly interspaced short palindromic repeats) from *Streptococcus pyogenes*, modified to be used for genome editing (Figure 13). This project was planned and executed together with Dr. Sarfarazhussain Farooqui and Andreas Kunze from the Gilmour laboratory at EMBL. In this system, the Cas9 nuclease is directed to a specific location in the genome by a small single-guide RNA sequence (sgRNA), consisting of a 20-nt guide sequence and a gRNA scaffold. The only requirement for the target sequence is that it has to be located directly at 5' of a so called PAM (protospacer adjacent motif), which is NGG in the case of Cas9 from *Streptococcus pyogenes.* The nuclease Cas9 makes a blunt cut 3 bp upstream of the PAM motif. This double strand break is then repaired by the endogenous NHEJ (non-homologous end joining) pathway, which has a high error rate, creating random indel mutations at the site of repair. Those mutations are likely to result in frameshifts and stop codons, allowing the creation of a gene knockout when the target sequence is in the coding sequence of a gene.

First, I designed 2 different sgRNAs located within the first exon of the gene (Figure 14A). The first exon was chosen to either cause a stop codon or a frame shift early on in the gene. The first exon consists of only 98 bp and even if the first amino acids are properly translated, the resulting protein is extremely unlikely to be functional. The second target site also contained a restriction site for PstI, which should be lost after a successful CRISPR/Cas9 induced mutation and therefore makes it easy to predict success rates of editing. First, I used the mRNA of Cas9 and the sgRNA for injection into the fertilized egg at the one cell stage. These experiments did not lead to a successful mutation in

the ALADIN gene. Limitations of this approach are the inefficient association of the sgRNA and Cas9 and possible degradation of the sgRNA within the cell. A more efficient alternative is the direct injection of purified Cas9 proteins already in complex with the sgRNA (Gagnon et al., 2014).

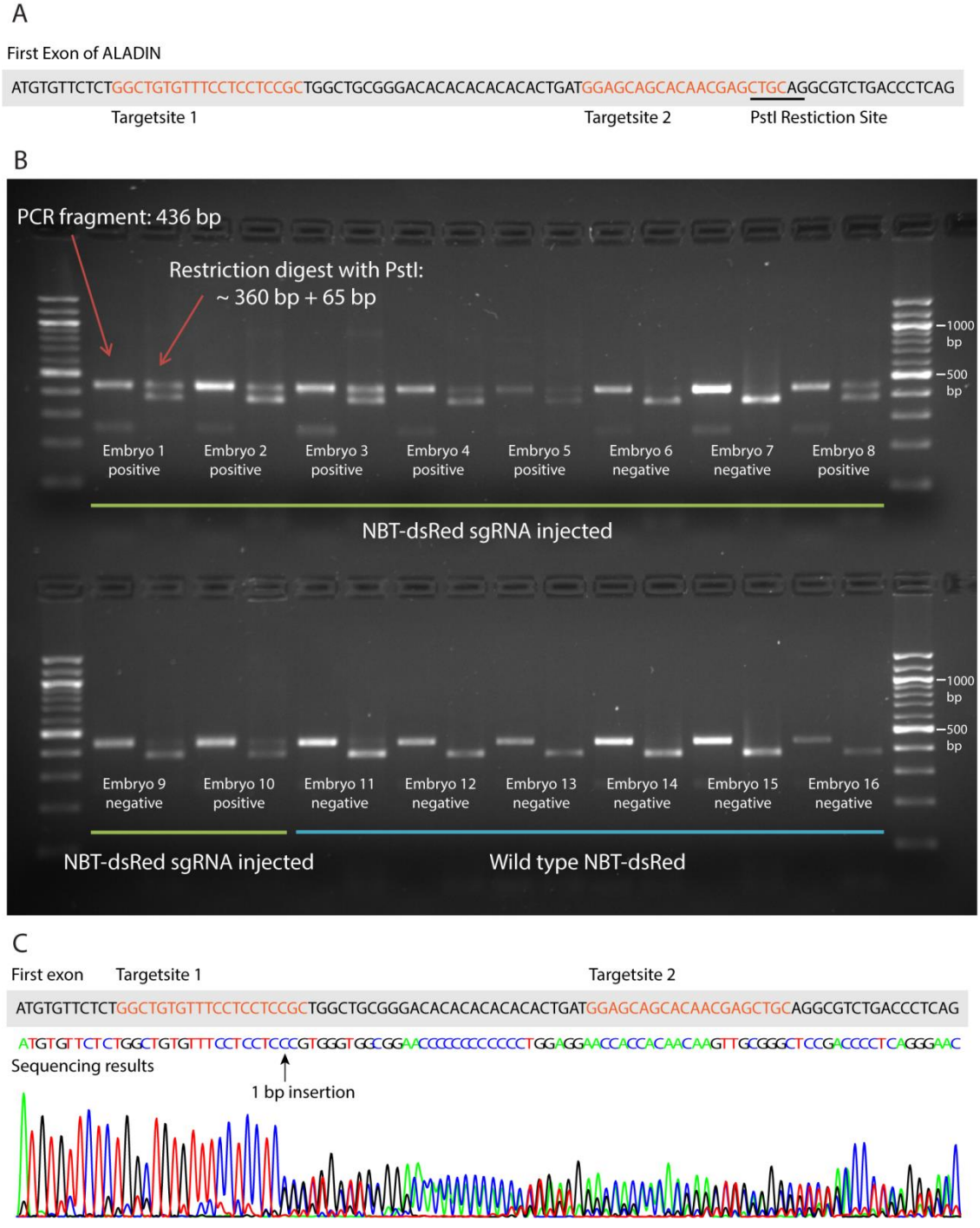


**Figure 14: (A) First exon of ALADIN with the first and second target site as well as the PstI restriction site used for quick verification of a successful mutation. (B) Restriction digest using PstI of embryos 24 hpf to check for successful mutations within the second sgRNA target site. (C) Sequencing result of a typical heterozygous sequence. After the first sgRNA target site the intensity of the sequencing trace is cut in half and traces are overlaying.**

Both sgRNAs were used in parallel to increase the mutation rate. Another advantage of the ready for action complex is that it induces double strand breaks faster and therefore has a higher chance to act already during the one cell stage. This method enabled the generation of mutations and also increased the survival rate of the embryos dramatically. Mutations were confirmed 24 hours after injection by restriction digest (Figure 14B) using the restriction enzyme PstI site in the 5' region of the second target site. Embryos that were not injected were raised and used as reference set, to exclude mutations within the sgRNA target site of the parents. Six wild type NBT-DsRed embryos were used for restriction digest and all showed the same pattern. Most of the embryos which were injected with the Cas9/sgRNA mix were heterozygous. The embryos of the $F_0$ generation were raised for 12 weeks and tested for successful mutations by DNA sequencing. Typically the intensity of the peaks drops by 50% at the target site in heterozygous embryos and cause overlaying reads of differentially ALADIN genes (Figure 14C). This is caused by insertions or deletions of base pairs by errors in the double strand break. Several fish from the $F_0$ generation were selected for further experiments, based on their mutation within the ALADIN locus. Preferred mutations should lead to a frame shift or introduction of a stop codon to ensure a complete disruption of the gene. Several fish had to be excluded because of 3 bp insertions or deletions or effects on the first splice site, as the number of base pairs within the first intron is divisible by three and therefore its loss leads only to the insertion of additional amino acids but no frame shift. This extra loop does not guarantee a disruption of the proper protein function. The main goal of generating a $F_1$ generation by outcrossing fish of the $F_0$ generations with a preferred mutation to wild type NBT-Ds Red zebrafish was to have a successful mutation being transmitted in the germ line. 24 hpf embryos were analyzed by restriction digest and sequencing, and raised till adulthood. The fish of the $F_1$ generation did not show any obvious defects and also the development to adulthood did not seem to be impaired. This was expected since only homozygous or compound heterozygous mutations cause the Triple A Syndrome in humans.

Adult fish were mated by combining the same or highly similar mutations in order to generate homozygous knockouts in the $F_2$ generations. By matching two heterozygous $F_1$ zebrafish 25% of the embryos are expected to carry a homozygous mutation within the ALADIN gene. Only few eggs were laid and succefully fertilized. Some embryos had obvious defects as shown in Figure 15. In comparison to wild type fish with no externally visible defect (Figure 15 A, embryo I.) various different defects could be detected. For some fish the red-fluorescent-protein under the control of a neural-specific tubulin protein seemed to be less abundant in the brain and fluid was retained in the

pericardial cavity (Figure 15 A, embryo II., see arrows). These embryos also had a slower heartbeat in comparison to unaffected ones. In Figure 15 B and C embryos with malformed eyes and curved tail fins with a kink at the tip are shown. The survival rate for the defect embryos was limited to 6 to 10 days after fertilization.

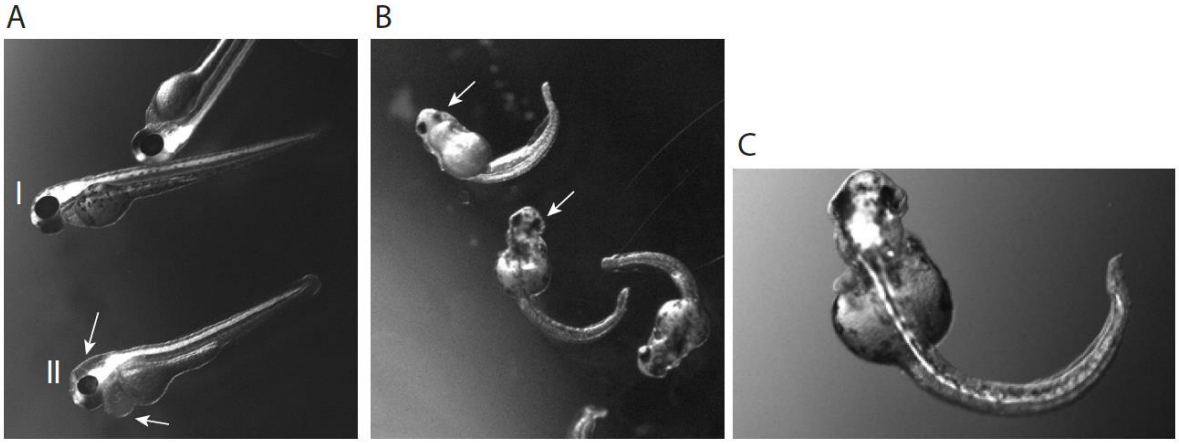


**Figure 15: $F_2$ generation embryos with obvious defects and limited survival rate. (A) I. Embryo with no externally visible defects. II Embryos with fluid retention and the red-fluorescent-protein under the control of a neural-specific tubulin protein is less abundant in the brain. (B) and (C) Embryos with malformed eyes and a curved and kinked tail.**

This data might indicate a more important, possibly essential role of ALADIN in development than one would have anticipated based on published mouse knockout phenotypes that were claimed to be relatively mild (Huebner et al., 2006). However, further work is required to confirm the reduction of ALADIN protein levels and altered NPC composition in these embryos. Also the exact functional role of ALADIN needs to be further studied.

## 3.2 NPC composition at the Nuclear Envelope (NE) in comparison to Annulate Lamellae (AL) in *Drosophila*

Annulate lamellae (AL) are pools of cytoplasmic nuclear pore complexes, stacked into the membranes of the endoplasmic reticulum (ER). During my PhD studies Dr. Bernhard Hampoelz found that these pools allow a rapid growth of nuclei during early development in *Drosophila* before gastrulation. I contributed to this work by a compositional analysis of annulate lamellae NPCs in comparison to nuclear envelope NPCs (Hampoelz et al., 2016).

### 3.2.1 Background: Annulate Lamellae in embryonic development

Cell division and nuclear growth necessitates the constant formation or reassembly of new NPCs. During the interphase of differentiated cells new NPCs are assembled from scratch (D'Angelo et al., 2006) and inserted into the NE, which requires the fusion of the double membrane system of the NE (Doucet and Hetzer, 2010). In higher eukaryotes, open mitosis leads to a complete disassembly of NPCs into stable subcomplexes. Here, NPC re-assembly at the end of mitosis does not require membrane fusion, because the NE is reforming simultaneously (Schooley et al., 2012). In *Drosophila,* mitosis proceeds in a semi-closed fashion, meaning that a fenestrated NE is maintained although NPCs disassemble. The early development of *Drosophila* requires a massive de novo insertion of NPCs into the nuclear envelope in order to keep up with the enormous increase of nuclear surface area during the fast division of nuclei within the syncytial blastoderm. The first 13 synchronized divisions last on average 10 minutes. During each mitosis the nuclei divide and grow without a significant drop of NPC density within the nuclear envelope (Hampoelz et al., 2016). It has been surmised but not proven that this is accomplished with the help of stock-piles of NPCs, stored in so-called annulate lamellae (AL), which are part of the ER crossed by structures similar to NPCs. 2D transmission electron micrographs of AL-NPC are shown in Figure 16A. AL are absent from many cell types but present in various cultured cell lines (Cordes et al., 1996) and highly abundant especially in germ cells and early embryos of various species.

### 3.2.2 Characterization of annulate lamellae using quantitative mass spectrometry

How and if these AL contribute to the amount of NPC located within the NE remained unclear. Dr. Bernhard Hampoelz analyzed AL in the early development of *Drosophila* and determined how AL become part of the NE by an en bloc insertion of membrane sheets (Figure 16B). In the early *Drosophila* embryo and later, the ER and NE thus have to be conceptualized as one membrane

continuum. Because NPCs have some components that asymmetrically distribute across the NE, this work raised the question if AL-NPCs have the same composition than NPCs within the NE. If so, mechanisms have to exist that prevent their wrong, so to say 'upside down' insertion into the NE. Alternatively, NPCs in AL could be reduced in asymmetric components. I addressed this question using mass spectrometry as described in the following. The results are published in (Hampoelz et al., 2016).

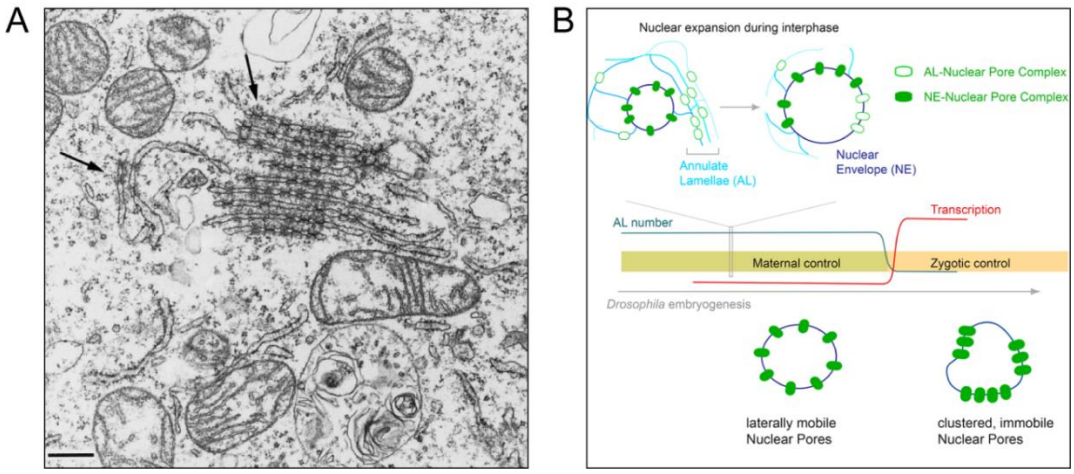


**Figure 16: (A) Stacks of annulate lamellae in the cytoplasm of bovine mammary gland epithelial cells. Arrows indicate structures highly similar to NPCs located within ER sheets (modified from (Cordes et al., 1996)). Scale bar: 20 µm (B) Incorporation of AL-NPCs within the syncytial blastoderm during *Drosophila* embryogenesis (image reproduced from (Hampoelz et al., 2016)).**

To address this question, blastoderm embryos were fractionated into nuclei, cytosol and microsomal membranes (Figure 17A) (done by Dr. Bernhard Hampoelz). Each fraction was analyzed by shotgun mass spectrometry and also Western blot (the latter done by Dr. Amparo Andres-Pons) to control for the quality of the subcellular fractionation of the embryo (Figure 17 B and C). Both Western blot and shotgun proteomics demonstrated that typical subcellular marker proteins are highly enriched or exclusively present in the respective fractions. The distribution of Nups across these fractions is shown in Figure 17D. In both the cytoplasmic membranes and the nuclei fraction many Nups are being identified. Within the soluble cytosol the measured intensities of Nups are significantly lower or below the detection limit with the exception of Sec13, which is also a member of a cytosolic coat protein complex ll (COPll). As expected, the nuclear fraction contained all well-characterized Nups. Therefore, NE-NPCs appear to be fully functional even during early development. Also AL-NPCs contain the Nup107 and Nup93 subcomplexes that constitute the scaffold of each NPC. Further, the transmembrane Nups and the linker Nup98, that is required for scaffold assembly, are present. However, the Nup62 subcomplex and most of the asymmetrically distributed Nups, that is Nup214,

Mbo (Nup88 in human), Mtor (Tpr in human), Nup153 and Nup50 are not detectable. The asymmetric components of the CR and NR, Nup358 and Elys, respectively, are an exception from this trend and occur in both, AL-NPCs and NE-NPCs. I conclude that AL-NPCs comprise fully assembled NPC scaffolds that are devoid of most FG-Nups and asymmetrically distributed Nups, except for Elys and Nup358.

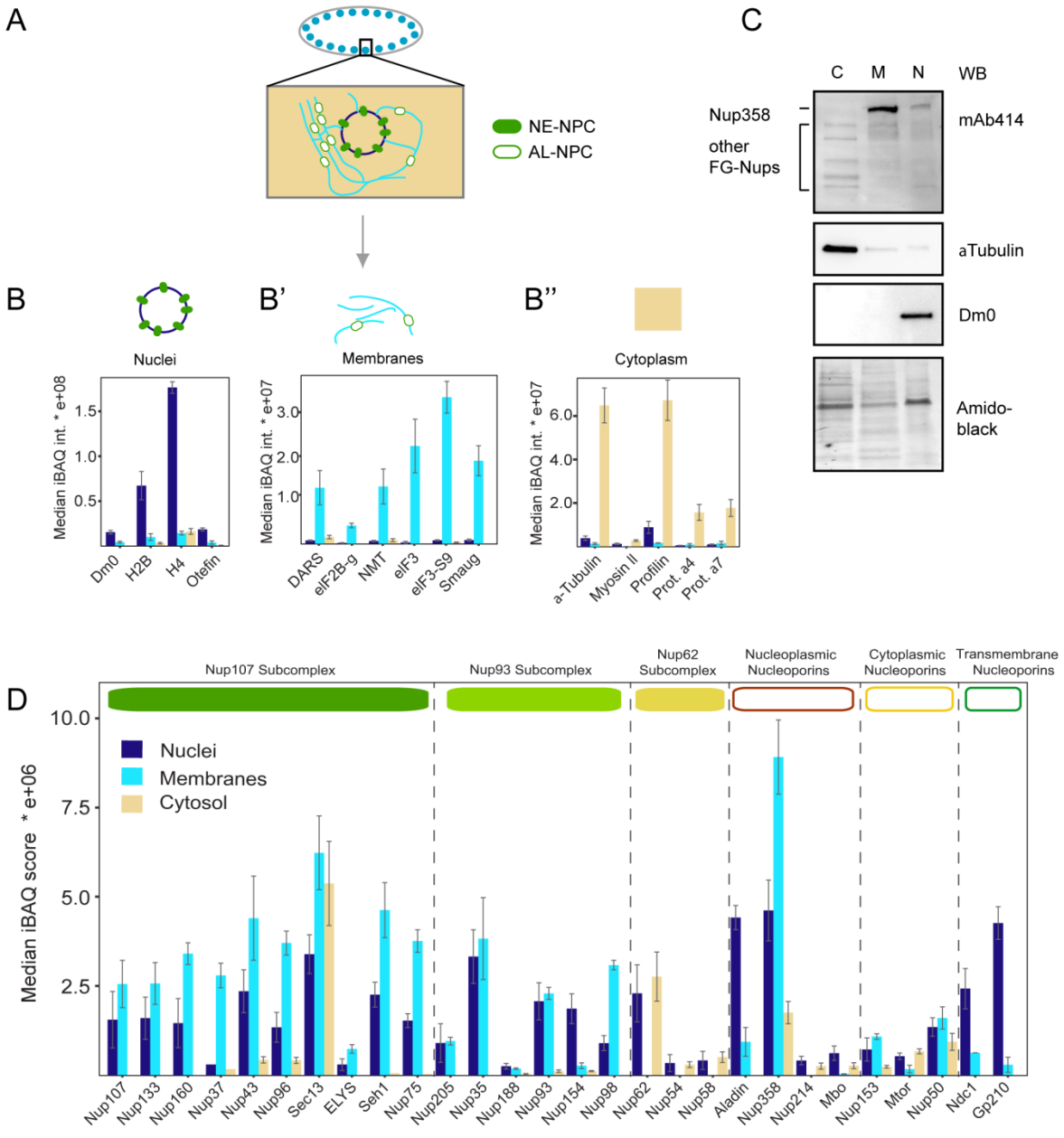


**Figure 17: (A) Subcellular fractionation of *Drosophila* syncytial blastoderm embryos. (B) and (C) Shotgun proteomics and Western blot of Nuclei, Membranes and Cytoplasm fraction to verify a clean fractionation procedure. (D) Shotgun proteomics of NPC members as barplot of median iBAQ scores. All MS-experiments were carried out in biological triplicate and technical duplicate (modified from (Hampoelz et al., 2016)).**

## 3.3 Nuclear transport receptor cargo specificity

### 3.3.1 Background: Previous methods of characterizing NTR-cargo relationships proteome wide

The nucleocytoplasmic transport system is essential for cellular homeostasis. The key players are the NTRs, cargos as well as the NPC itself, and a tight regulation of the relevant interactions is crucial. To better understand these complex processes, it is important to know which cargos are transported by which NTRs and also if there are preferred routes through the NPC. Previous studies were mainly focusing on individual cargos and their specific transport pathways. Some approaches charted specific nucleocytoplasmic transport pathways and I want to highlight a few in an exemplifying manner in the following paragraphs.

SILAC combined with quantitative MS to identify XPO1 cargos

Thakar and colleagues used stable isotope labeling by amino acids in cell culture (SILAC) combined with quantitative mass spectrometry to identify potential cargos of XPO1 with and without its inhibitor leptomycin B (Thakar et al., 2013). The general workflow is depicted in Figure 18. First, HeLa cells were either grown in cell culture medium with light labeled amino acids or with heavy labeled amino acids and additional leptomycin B. After mixing, subcellular fractionation and in-gel digestion the cytosolic and nuclear fraction were measured by LC-MS/MS. With this method more than 100 potential XPO1 cargos could be identified either by depletion from the cytoplasm or enrichment within the nucleus upon leptomycin B treatment. Both known cargos and proteins with previously unknown NES were detected. Selected candidates were further analyzed using immunofluorescence and in vitro binding studies to show a direct interaction with XPO1.

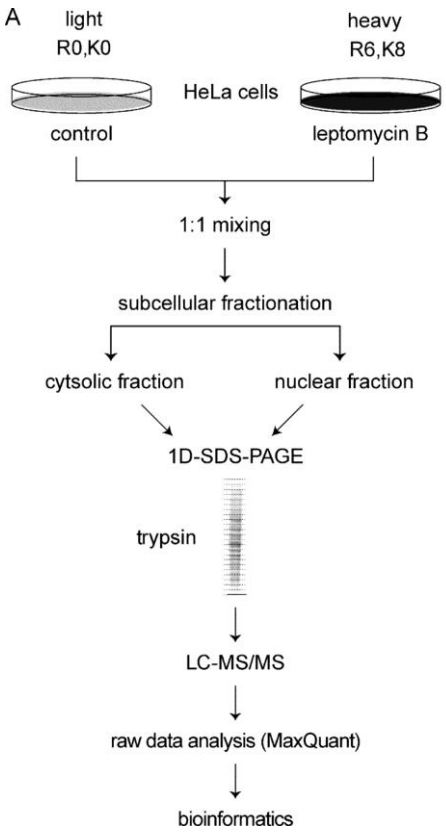


**Figure 18: Workflow of SILAC and quantitative MS (image reproduced from (Thakar et al., 2013))**

SILAC and *in vitro* transport in permeabilized cells combined with quantitative MS to identify cargos of importin βs

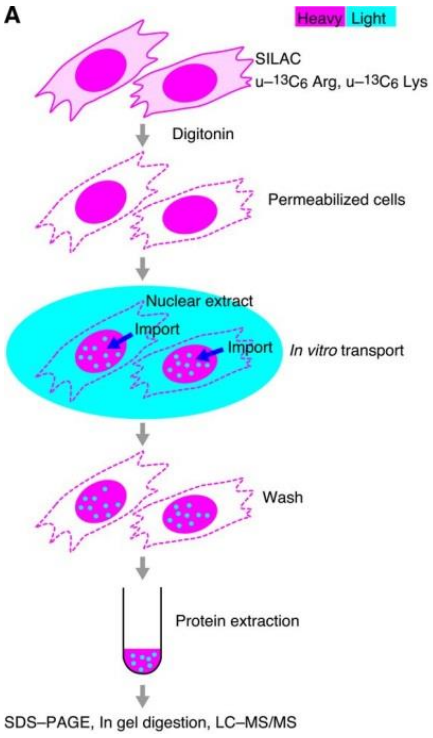


**Figure 19: Workflow of *in vitro* transport assays using SILAC and quantitative MS (image reproduced from (Kimura et al., 2013a))**

A similar approach using SILAC and quantitative mass spectrometry but in combination with an *in vitro* transport system was used by Kimura and colleagues to identify potential cargos of various importin βs (Kimura et al., 2013a; Kimura et al., 2017; Kimura et al., 2013b). HeLa S3 cells were grown in medium containing heavy labeled amino acids and then permeabilized using digitonin which keeps the NE intact. The permeabilized cells were incubated in light labeled nuclear extract, with or without the particular NTRs of interest. The nuclear extract was analyzed using quantitative MS. The complete workflow is depicted in Figure 19. Known cargos were identified as well as new ones, which were in part validated using an *in vitro* binding assay called Bead Halo. Bead Halo assays can detect interactions of affinity-tagged proteins immobilized on beads with fluorescent macromolecules (Patel and Rexach, 2008).

Inhibition of XPO1 using Leptomycin B and nucleocytoplasmic partitioning combined with quantitative MS

Wühr and colleagues used *Xenopus* oocytes to study the composition of the nuclear proteome in this particular cell type (Wühr et al., 2015). They used oocytes because of the easy and fast separation of nuclei from cytosol. The proteins localized within the cytoplasm, nucleoplasm or both were compared to changes in their distribution induced by inhibition of XPO1 with Leptomycin B for 24h. Only 3%, corresponding to 226 proteins, of all identified proteins showed a significant enrichment within the nucleus upon XPO1 inhibition. More than 80% were not previously characterized as potential cargos of XPO1.

Affinity purification of XPO1 cargos by immobilized XPO1 in a RAN-dependent manner in three different species

Cellular extracts of *Xenopus*, HeLa cells and *S. cerevisiae* were analyzed in their binding capability to immobilized XPO1 in a RanGTP dependent manner by Kırlı and colleagues (Kırlı et al., 2015). NTRs were depleted from the cell extracts and then proteins bound to XPO1 analyzed by quantitative MS. According to their binding strength, potential cargos were classified into different categories. Overall many new cargos of XPO1 were identified ranging from 700 in *S. cerevisiae* up to 1050 in HeLa cells. More than 90% of the potential new cargos were previously unknown to interact with XPO1. 29 candidates (23 validated positive) were further screened with confocal laser scanning microscopy by fusing the candidate cargo to GFP and incubating the cells with and without Leptomycin B.

All the previously mentioned studies provided an overview of transportins and XPO1 dependent transport pathways. New cargos were identified and cross quantified using complementary experimental approaches. However, these experiments were conducted under non-physiological conditions or *ex vivo* and more importantly, are not generalizable to all NTRs. Especially the use of inhibitors is difficult, since only few are known to specifically inhibit only a selected NTR and not to affect a whole class of transport receptors. The specificity of the cargos was largely not addressed and therefore it cannot be excluded that a cargo can also be transported by other NTRs. The cargo spectrum of the vast majority of NTRs thus remains incompletely charted.

### 3.3.2 Generation of stable cell lines with BirA fused to a protein of interest

To globally map the cargo spectrum of a wide range of different NTRs I applied the BioID system. The advantages over previous proteome wide methods are the broad applicability to, in principle, all the NTRs, the capability to capture transient interactions *in vivo* (Coyaud et al., 2015), and the independence from inhibitors. The promiscuous biotin ligase BirA was fused to NTRs (Figure 20A), and stable and inducible (Tet-ON) cell lines were generated using a HEK293-based Flp-In T-REx system. The choice to generate two independent cell lines for each NTR bearing BirA either at the N- or C-terminus was aimed at ruling out potential interferences with protein function. These interferences might arise from the fact that the biotin ligase is e.g., blocking an important protein-protein interaction interface that is necessary for NTR activity. Each construct also contains a FLAG tag that enables to assess the expression and localization of the fusion proteins by Western blotting (WB) and fluorescence microscopy.

**Table 1: Overview of all stable cell lines generated of NTRs to analyze the nucleocytoplasmic transport system. An overview of alternative names can be found in Table 7 in the supplement.**

| NTR | Class | Directionality of Transport | Position of BirA |
|---|---|---|---|
| **IMA1** | Importin α II | Import | C- and N-Terminus |
| **IMA5** | Importin α I | Import | C- and N-Terminus |
| **IMA6** | Importin α I | Import | C-Terminus |
| **KPNB1** | Importin β | Import | C- and N-Terminus |
| **IPO4** | Importin β | Import | C- and N-Terminus |
| **IPO5** | Importin β | Import | C- and N-Terminus |
| **IPO11** | Importin β | Import | C-Terminus |
| **IPO13** | Importin β | Import/Export | N-Terminus |
| **TNPO1** | Importin β | Import | N-Terminus |
| **TNPO2** | Importin β | Import | C- and N-Terminus |
| **XPO1** | Importin β | Export | C-Terminus |
| **XPO2** | Importin β | Export | C- and N-Terminus |
| **XPO7** | Importin β | Export | N-Terminus |
| **NTF2** | Other | Import | C- and N-Terminus |
| **NXT1** | Other | Export | C- and N-Terminus |
| **NXT2** | Other | Export | C- and N-Terminus |

**Table 2: Overview of stable cell lines for additional targets and controls.**

| Additional targets | Position of BirA |
|---|---|
| **RAN** | C- and N-Terminus |
| **NLS-NES-Dendra2** | C- and N-Terminus |
| **BirA alone** | - |

16 different NTRs were fused to BirA (Table 1). To cover a broad range of nucleocytoplasmic transport pathways three importin α from two different subclasses and ten importin β, functioning as importin, exportin or both, were chosen. Additionally NXT1, NXT2, the Ran recycling factor NTF2 and Ran itself were fused to BirA. As controls, cells were transfected with BirA alone and BirA fused to NLS-NES-Dendra2, a green-to-red photoswitchable fluorescent protein, engineered to continuously shuttle in and out of the nucleus. To validate the proper function of each stable cell line, the expression of the NTR-BirA construct and biotinylation in the corresponding cell lines were confirmed by Western blot. The overexpression of the construct was induced with tetracycline followed by addition of biotin for efficient labeling of interacting proteins.

In Figure 20B, exemplifying Western blots for Transportin 2 (TNPO2) and Importin 5 (IPO5), both with BirA fused to the C- or N-terminus, are shown. All constructs show overexpression upon induction with tetracycline and the band size corresponds to the expected molecular weight of the particular construct of the NTR fused to BirA and FLAG tag. A strong biotinylation across most of the molecular weight spectrum appears only after addition of biotin to the cell culture medium.

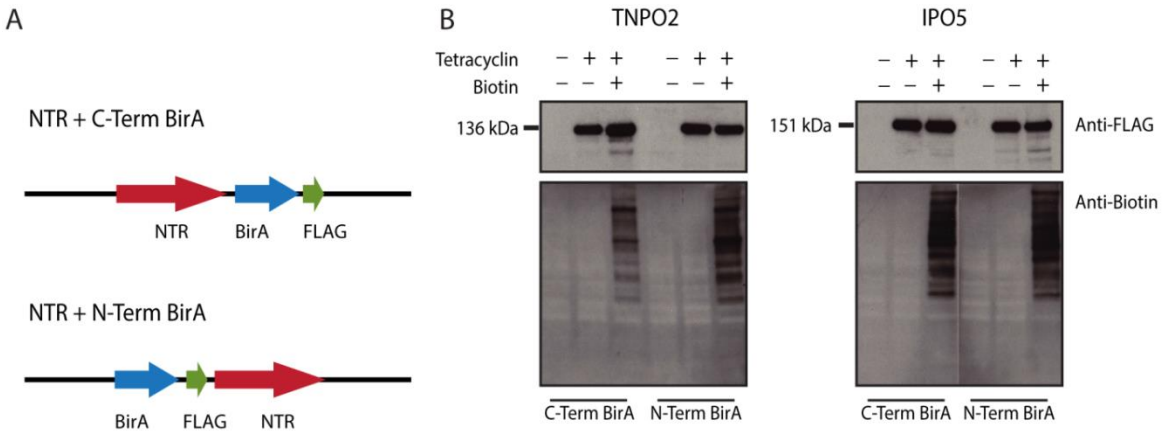


**Figure 20: (A) NTR-BirA construct for fusion of BirA to the C- or N-terminus of the NTR. (B) Western blot of total cell lysate from cells expressing Transportin 2 or Importin 5 with BirA, fused to the C- or N-terminus, treated with only tetracycline or with biotin in addition.**

To compare various different cell lines with each other, similar amounts of total cell lysates were analyzed by Western blot (Figure 21). The expression of the construct varied between different cell lines and partly also between C- and N-terminal versions. IMA1 and IMA5 both had a stronger signal for FLAG and the overall biotinylation level was higher when BirA was fused to the C-terminus. The class of importin alphas has one common feature, the IBB-domain at the N-terminus. This domain is important for the binding to KPNB1 and this might be a reason for the lower expression and the reduced biotinylation of proteins by these constructs. The two stable cell line variants of TNPO2 and IPO3 did not show any significant differences in the detected intensity for FLAG or the band pattern of the biotinylated proteins. Dendra showed a similar overexpression level, according to the FLAG tag, to the NTRs but the biotinylation signal is much weaker, since less potential interaction partners are available. Not only the overall biotinylation signal varies, but also various distributions of bands are observable. The pattern between importin α and importin β shows different intense bands and indicate distinct interactions partners within the cells.

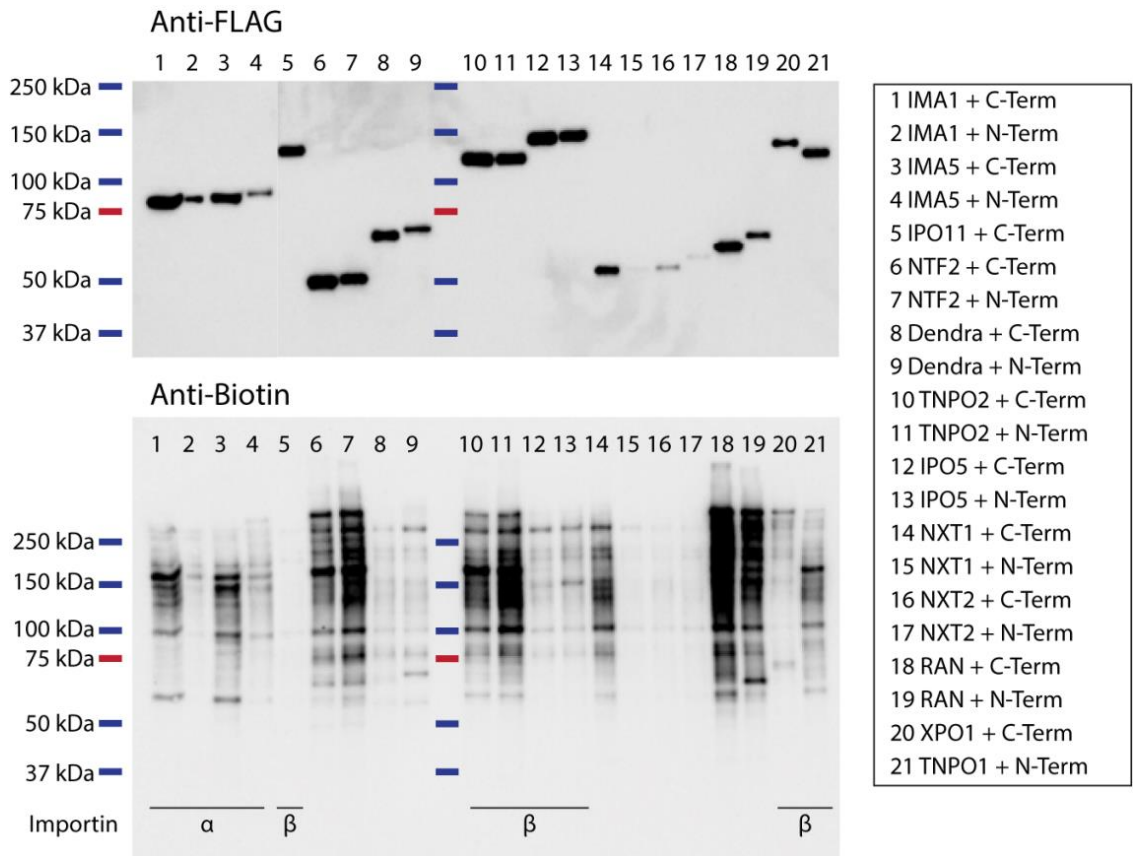


**Figure 21: Comparison of different overexpression and biotinylation levels across multiple stable cell lines. NTRs of the α or ß class are labeled.**

### 3.3.3 Staining of stable BirA-NTR cell lines over time in diverse cell lines

To assess the functionality of the NTR fusion proteins *in situ*, the overall distribution of biotinylated proteins was monitored using fluorescence microscopy. For importins a nuclear staining would be expected, for exportins a nucleocytoplasmic staining since almost all exported cargos had to be imported beforehand. The exact distribution is of course dependent on the transport kinetics of the fraction of labeled cargo proteins that shuttle in and out of the nucleus. For this experiment cell were grown on coverslips, expression of BirA fusion induced for 24 h and biotinylation allowed to proceed for additional 24h. Cells were fixed and stained with fluorescently labelled streptavidin to detect all biotinylated proteins. In Figure 22 an example of biotinylation kinetic (5 vs 28 h) for TNPO2 is shown. A strong enrichment of biotinylated proteins inside the nucleus over time is apparent as expected, and implies that the TNPO2-BirA protein actively facilitates nucleocytoplasmic transport.

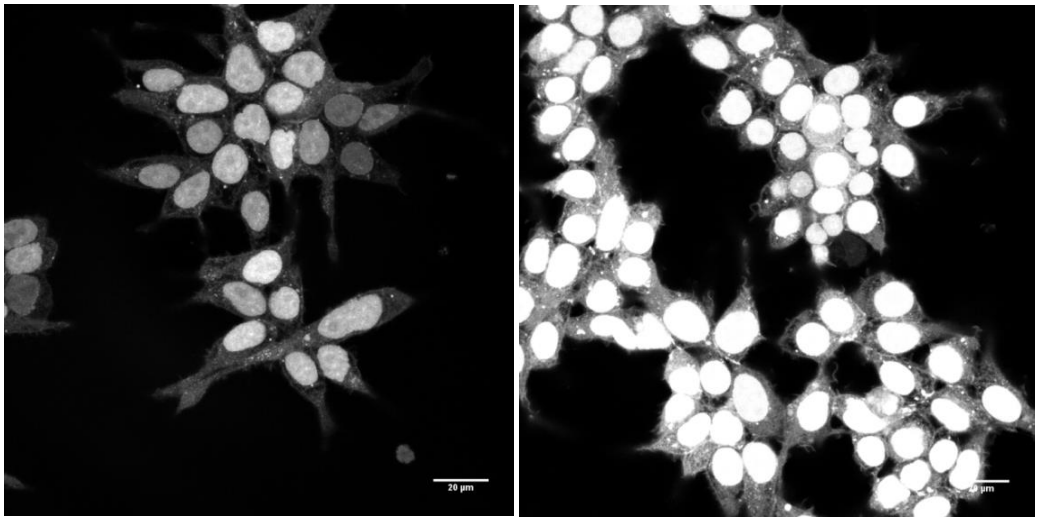

**Figure 22: TNPO2-BirA increase of biotinylated proteins over time. Treatment for with 50 µM biotin 5 h on the left and 28 h on the right. Scale bar, 20 µm.**

The analysis of different NTR fusion proteins, as shown in Figure 23 for TNPO2 and IPO5, revealed distinct staining patterns. Both cell lines show most biotinylated proteins inside the nucleus but for TNPO2 a distinct region, possibly the nucleolus, is excluded from the staining. In both cases cytoplasmic staining is observed, but in case of TNPO2 it is more disperse, while in case of IPO5 biotinylated proteins seem to accumulate at the plasma membrane and the cytoskeleton of the cells.

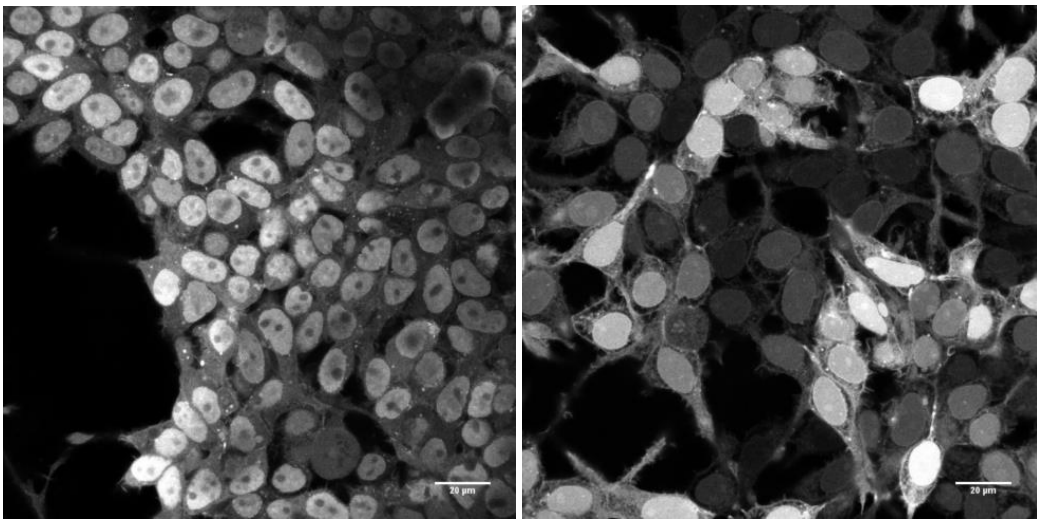

**Figure 23: Different staining pattern of biotinylated proteins for TNPO2-BirA (left) and BirA-IPO5 (right). Scale bar, 20 µm.**

To confirm this finding, I further analyzed the local distribution within the cells together with well-established markers. I used Hoechst dye to stain for the DNA and stained NPCs using the monoclonal

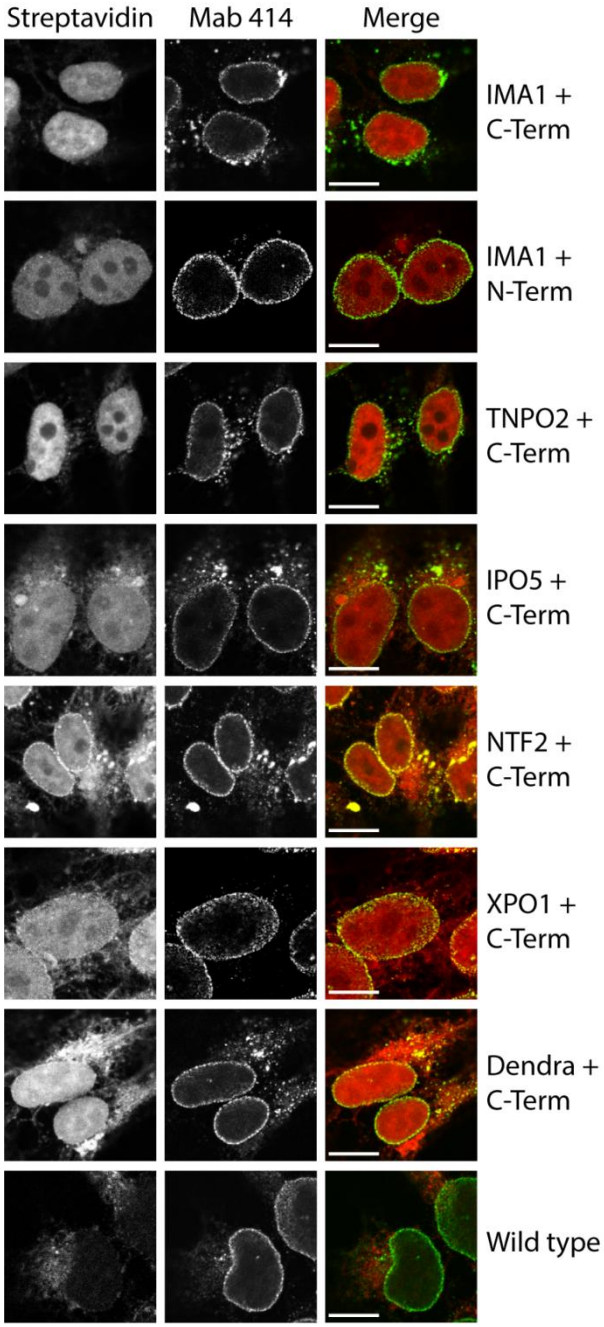


**Figure 24: Stable NTR-BirA cell lines fixed and stained with Streptavidin-Alexa647 (red) to visualize all biotinylated proteins and anti-Mab414 (green) to stain all FG-Nups as reference for the nucleus. See text for detailed explanation of different staining pattern. Scale bar, 10 µm.**

antibody Mab414. This antibody recognizes all FG-Nups (Sukegawa and Blobel, 1993) that interact with translocating cargo complexes and would enable me to answer the question if some NTRs more strongly interact with the NPC or just biotinylates more than others. In Figure 24 an overview of selected NTRs covering all clades of NTR and also 2 control cell lines are shown. IMA1-BirA is an importin α class NTR involved in the classical import pathway. Biotinylated proteins are enriched in the entire nucleus with no strong preference for any sub-nuclear compartment, neither the nucleolus nor the NE. The second cell line for IMA1 with an N-terminal BirA fusion showed an overall lower biotinylation signal and the nucleoli were considerably less stained. This fusion of BirA to the N-terminus might impair the transport capability at least to some extend since the N-terminal importin beta binding (IBB) to importin beta (KPNB1) is disturbed. Residual activity might be observed because transport in an KPNB1 independent manner is possible for some cargos (Miyamoto et al., 2002). Stable cell lines with the TNPO2-BirA show an exclusion of nucleoli and no enrichment for biotinylated protein at the NE. IPO5-BirA shows a stronger staining of biotinylated proteins also outside of the nucleus, which might be a hint for additional yet uncharacterized functions in other cellular processes. The biotinylation signal is enriched both, at the plasma membrane and at the NE (Figure 23). Also NTF2, the Ran recycling factor, shows a strong signal of biotinylated proteins at the NE. This might result either from a

strong accumulation of biotinylated RAN in proximity of the NPC or a stronger biotinylation of Nups because this NTR constantly shuttles through the NPC. XPO1-BirA shows a biotinylated protein in the cytoplasm and nucleus with a similar intensity. This is expected due to the fact that every protein, which is exported from the nucleus, has to be imported in the first place. Dendra2-NLS-NES does not display clear enrichment in the nucleus or cytoplasm demonstrating that neither import nor the export route dominates nucleocytoplasmic localization. This cell lines was used as control to eliminate background signal in the following MS experiments (see below). Non-engineered control cells only have a weak biotinylation signal that occurs exclusively in the cytoplasm and most probably arises from the naturally biotinylated proteins inside of mitochondria.

The visual inspection of stained cells implied different interaction spectra of the various NTRs. There are obvious differences in the subcellular localizations of the labeled proteins. Differential co-staining with FG-Nups are indicative of varying interactions with the NPC. These observations are consistent with the interaction patterns of the respective proteins (see below).

### 3.3.4 Affinity purification of proteins biotinylated by BirA and elution strategies

Subsequent to the initial characterization of the stable cell lines described above, I optimized the conditions for affinity purification (AP) of proteins that were biotinylated by BirA-fused NTRs. The general workflow is depicted in Figure 25. Stable cell lines were induced using tetracyclin for 24h to express the NTR-BirA fusion proteins in the cell. None of the cell lines had an obvious growth defect in cell culture nor any other apparent phenotype. In comparison to control cells, the survival rate or the number of apoptotic cells was similar. After 24 h of induction with tetracyclin, 50 µM of biotin was added to the cell culture medium to supply the biotin ligase with its substarte. Cells treated with tetracyclin only did not show an accumulation of biotinylated proteins implying that the naturally availability of biotin within the cells, even though they were cultured in a complete medium, is insufficient. The labeling with biotin was done for 24 h. This time period was also suggested by others (Roux et al., 2012). The rather long labeling time facilitates the accumulation of signal and the detection of even low abundant or rarely transported cargos.

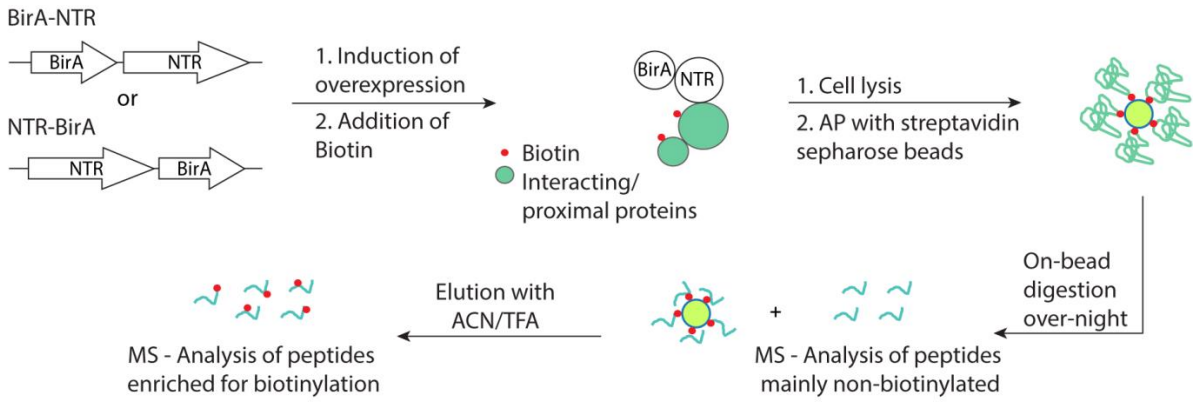


**Figure 25: General workflow for NTR-BirA affinity AP starting from cell line generation using two different constructs to rule out negative effects of the BirA on the proper function of the respective NTR. After induction and addition of biotin, cells were lysed and biotinylated proteins isolated using streptavidin sepharose beads. Isolated proteins were on-bead digested using trypsin and biotinylated peptides eluted an additional step using a mixture of ACN and TFA. Peptides from on-bead digestion and biotinylated peptides were analyzed by mass spectrometry (MS).**

Since biotinylation is a stable, covalent modification and complexes or protein structure do not have to be preserved, medium harsh lysis conditions were chosen. Streptavidin sepharose beads were used for the affinity purification. The interaction between streptavidin and biotin is one of the strongest interaction known in biology (~$10^{-14}$ mol/l (Green, 1975)) and facilitates extensive washes on beads to remove unspecific binding partners that might either bind to the beads directly or be part of a complex containing at least one biotinylated substrate. I tested different elution conditions to remove the entire protein bound, including high concentrations of competing biotin, alternative beads, low and high temperatures and combinations of these for variable durations. None of them resulted in a sufficient elution. Therefore, on-bead digestion using trypsin became my method of choice. MS analysis of the on-bead digested proteins revealed that only very few of the peptides were biotinylated. This finding illustrates the weakness of this commonly used approach. The targets are identified indirectly and one important layer of information is lost during this process because most peptides with a biotinylated lysine remain bound to the beads. This information could e.g. be used to indentify interacting domains within the protein structure or to asses its surface accessibility. In this particular case, the site of recognition is still unclear for many potential cargos and such information might facilitate the identification of new recognition motifs. I identified conditions that are sufficient to elute the remaining peptides using a mixture of ACN and TFA (Ori et al., 2009) - with only a minimal destruction of the beads. The on-bead and ACN/TFA elutions were analyzed by using separate shotgun proteomics runs.

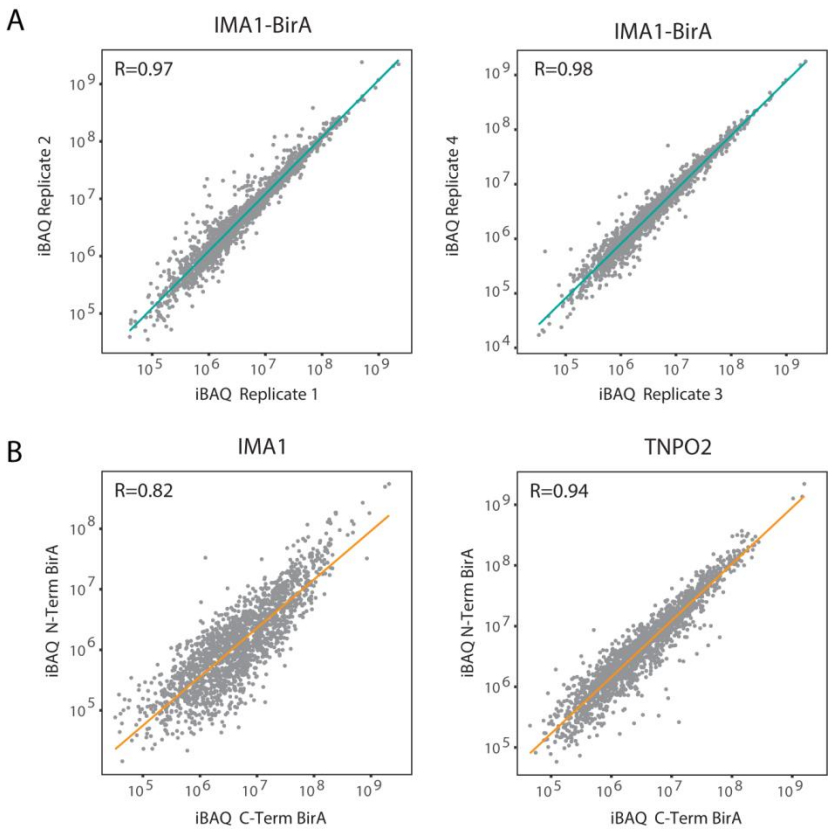


**Figure 26: Correlation of iBAQ scores (proxy for protein abundance bases on summed peptide intensity) of on-bead digested APs either between biological replicates (A) or between corresponding C- and N-terminally tagged NTRs (B). The correlation between biological replicates is high, but between cell lines more variations are detectable, as expected.**

In order to capture also low abundant proteins with high confidence, each AP was performed in quadruplicate (fully independent isolations performed using cells from different passages), and each sample was measured in technical duplicate (repeated injection into the mass spectrometer of the same sample). First, I compared the iBAQ scores for Streptavidin-enriched proteins across biological replicates to evaluate the performance of the biotin labeling and how consistent the results are within specific cell lines. Figure 26A represent two examples of the correlation of iBAQ scores between two biological replicates of IMA1-BirA. All combination of biological replicates had correlations above R=0.96. I obtained similar results for the other targets, demonstrating the high quality of the dataset. Next, I compared median iBAQ scores across cell lines tagged either at the C- or N-terminus. In some individual cases, I found strong differences. For example, the C and N-terminal IMA1 fusions correlated with only 0.82 in comparison to TNPO2 that had a correlation of 0.94 (Figure 26B). An overview of all correlations scores is shown in Figure 39 in the supplement. These results are in line with the above described Western blot analysis that already suggested that

N-terminal tags to IMA1 are limited functional. A dual N- and C-terminal tagging strategy is thus important to better understand the influence of BirA onto the proper function of each NTR.

### 3.3.5 Data analysis workflow

The main objective of the computational data analysis was to quantitatively and statistically analyze all APs although the identified proteins strongly varied for each NTR due to their different specificities. The following data analysis workflow was developed and applied on the dataset together with Dr. Alessandro Ori and with support of Dr. Bernd Klaus (Center for Statistical Analysis, EMBL). The challenge thereby was not to lose any information about highly specific interaction partners that were often not cross-quantified owning to the intrinsic limitations of data-dependent acquisition by mass spectrometry and the dynamic range of the analyzed samples.

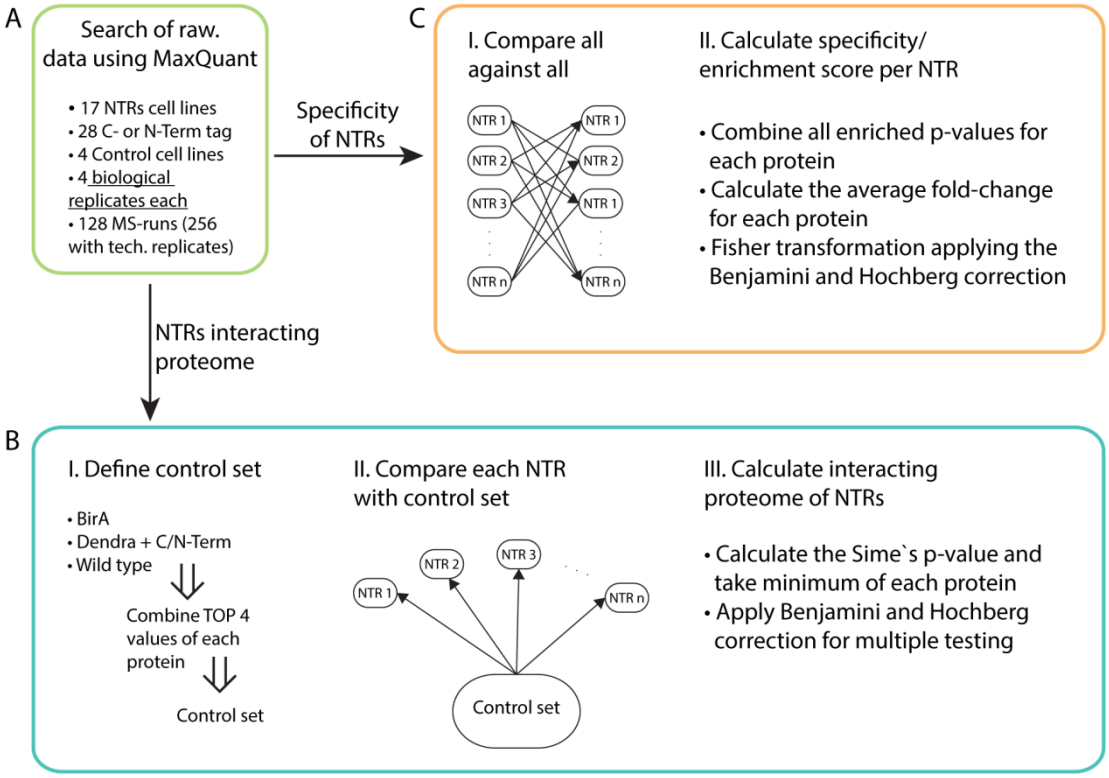


**Figure 27: Workflow for data analysis. (A) Dataset consisting of 17 different NTRs, in total 28 APs including cell lines with BirA fused to both sides, and 4 control APs. Each experiment was done in quadruplicate and technical replicate. (B Charting of the NTR interacting proteome (NIP) compared to the background proteome. (C) Procedure to define the specificity of each NTR by calculating enrichment scores (ES).**

Two independent workflows were developed to (i) globally define the NTR interacting proteome (NIP), and (ii) determine the specific interaction spectrum for each NTR. Both were based on label-fee quantification of peptides derived from on-bead digestion. The subsequent ACN/TFA elutions

containing mainly biotinylated peptides were used for a complementary analysis that is described below in section 3.3.8. In total 17 different NTR (including RAN) were fused to BirA and 28 AP datasets were generated, including cell lines with BirA fused to either terminus (Figure 27A). Additionally, 4 control APs were measured.

First I defined the NIP to identify general cargos and other common interactions partners of NTRSs, such as members of the NPC. In order to achieve this, I excluded background interactions by comparison with a newly defined control sample set (Figure 27B). As control APs I used stable cell lines expressing either a BirA construct alone or BirA fused to NLS-NES-Dendra2 either on the C- or N-terminus. The advantage of the "Dendra control" was the capability of detecting a background proteome both in the cytoplasm and in the nucleus as well as the NPC itself on account of its constant shuttling between these two compartments. Additionally I performed an AP on non-engineered HEK293 control cells exposing it to the same cell culture procedures and affinity purification conditions.

**Table 3: Strategy for the quantification of proteins that are not identified in specific biological samples. MNAR or MAR values based on the number of NA (not available) values in the four biological replicates per sample. The same applies for reciprocal numbers of NAs.**

| Sample 1 | Sample 2 | Data imputation strategy |
|---|---|---|
| 0 NA | 3 NA | MNAR |
| 0 NA | 4 NA | MNAR |
| 4 NA | 1 NA | MNAR |
| 3 NA | 1 NA | MAR |
| 2 NA | 1 NA | MAR |
| 2 NA | 0 NA | MAR |
| 1 NA | 1 NA | MAR |
| 1 NA | 0 NA | MAR |

With this control I could exclude proteins that bound unspecifically or due to their natural biotinylation sites to the beads and exclude side effects potentially arising from the excess of biotin in the cell culture medium. These four control APs were used to assemble a new control data set by selecting for each protein the four highest intensity values recorded in all control samples. Next, I performed pairwise comparisons of all NTR-BirA APs against the combined control matrix. To quantify proteins absent or poorly identified in either one of the two samples to be compared, these were set to background level using data imputation. Stringent criteria were used to identify values that are missing not at random (MNAR, e.g. because they specifically interact only with one of the two sample set and are thus not detectable in the other) or values missing at random (MAR, e.g.

because they are close to the limit of detection of the mass spectrometer). These two different types of missing values are expected to show different patterns over the multiple replicates of each sample group: MNAR are expected to be missing preferentially in one sample group while MAR should occur at similar frequency in both groups. An overview of different combinations of missing values (NA) per sample group is given in Table 3. Following data imputation and normalization, proteins were categorized into significantly enriched, depleted or unaffected groups as compared to the control dataset using pairwise t-test for each NTR. In order to combine all the pairwise comparisons and correct for multiple testing, I computed the Sime's adjusted p-values for every protein and then took their minimum to obtain one combined p-value for each protein (Lun and Smyth, 2014). The combined p-value obtained using the Sime's method indicates whether there is at least one condition (NTR) in which the given protein was enriched over the control dataset. With this approach I was able to define the NIP and the background proteome for the NTRs used in this study.

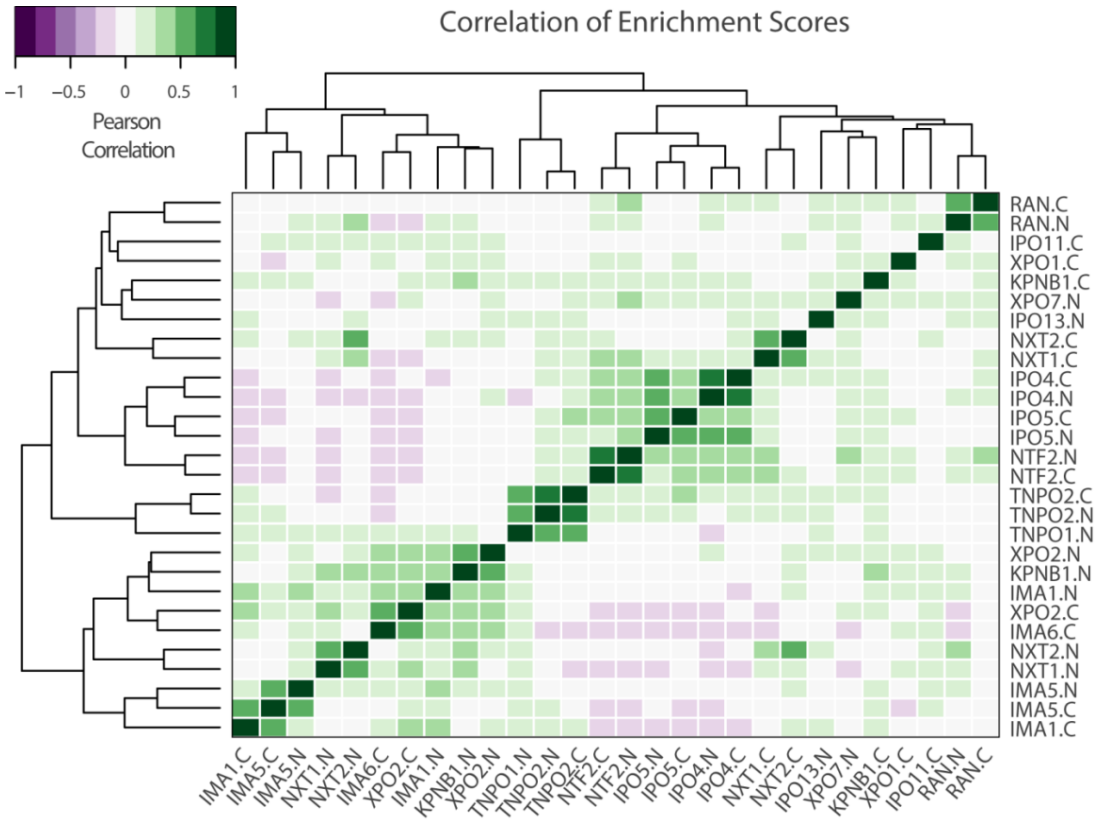


Figure 28: Pearson correlation of all 28 NTR-BirA AP ES scores.

The second part of the data analysis was focused to determine the specificity of the different NTRs. This approach enabled me to identify proteins that display preferential interactions with individual

or subsets of NTRs. Here, the control APs were excluded from the analysis and the data covering the 17 unique NTRs were directly compared to each other in a pairwise fashion (Figure 27C). Missing values were categorized into MAR and MNAR across biological replicates in the same fashion as described above. Following data normalization, differential protein enrichment was assessed using a one-sided t-test for all possible pairs of NTR samples (756 pairwise sample comparisons in total). An enrichment score (ES) was calculated for each protein-NTR combination by multiplying all enriched p-values obtained in all pairwise comparisons. An average fold change (FC) for each protein was also calculated across all experiments. To estimate global FDR (false discovery rate), I used the Fisher transformation applying the method of Benjamini and Hochberg for multiple testing to calculate a Fisher p-value. To confirm that the ES calculations represent the specificity of the different NTR-BirA APs, I performed a Pearson correlation of all ES scores. As shown in Figure 28 the correlation between C- and N-terminally tagged NTRs was high. The specificity of NTR was resolved but also their functional relationship was represented. For example, all three experiments of different Transportin samples clustered together, which is also reflected in similar results for enriched proteins (see below). Also here, the two differentially tagged versions of IMA1 do not cluster together, which was consistent with the apparent levels of biotinylation detected by Western blot (Figure 21) and the correlation of iBAQ scores (Figure 26). Taken together, this analysis comprises a solid validation for the quality of the data and reassures the usage of ES for all following data analysis steps.

### 3.3.6 NTR interacting proteome (NIP)

The statistical work flow described above allowed me (i) to categorize the identified proteins based on their enrichment versus an appropriate set of controls, and (ii) to evaluate NTRs experiments against each other on the other side. Specifically, for the first part of data analysis I focused on the discrimination of the NTR interacting proteome (NIP) from the background proteome. By defining a new set of control samples all the NTR-BirA APs could be compared to one reference background sample. This approach enabled me to define a subset of proteins enriched in NTR samples that are involved in nucleocytoplasmic transport including cargos and NPC components. A second subset of proteins that was enriched in the control samples includes the expected background proteins for BioID experiments such as naturally biotinylated proteins and unspecific binders to the streptavidin beads. In the following, I will describe the respective results from a biological point of view.

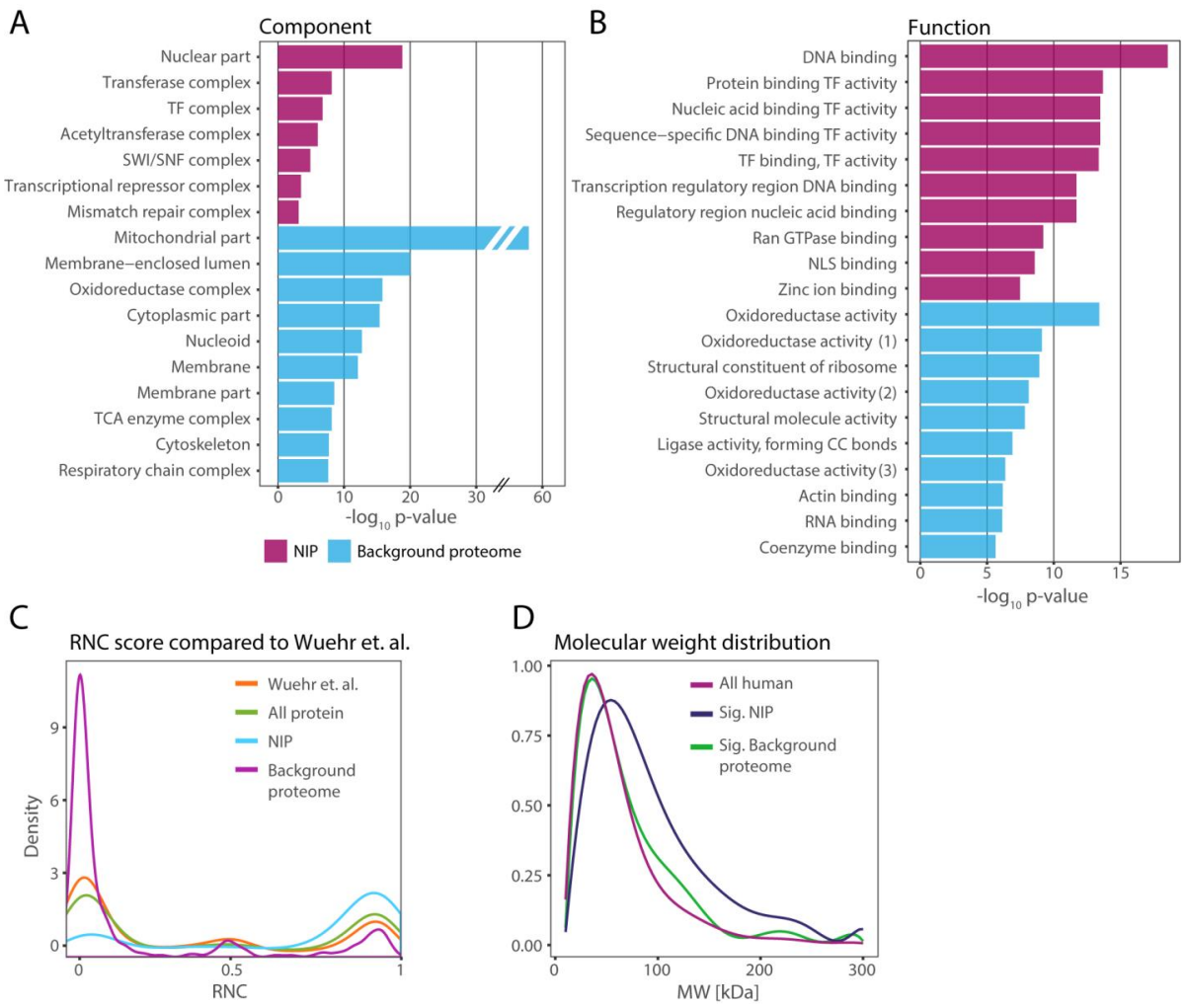


**Figure 29: GO-Enrichment for Component (A) and Function (B) a maximum of 10 GO-Term are plotted for the NIP or background proteome. (Oxidoreductase activity on (1) the aldehyde or oxo group of donors, (2) the aldehyde or oxo group of donors, NAD or NADP as acceptor, (3) NAD(P)H.) (C) RNC score from Wühr et al. compared to all significantly identified proteins in the NIP and NIP background. (D) MW of all protein in the NIP and background compared to all reviewed human proteins in the Uniprot database.**

The NIP and background proteome contained a total of 3754 proteins. In order to characterize the composition of these two set of proteins, I performed Gene Ontology (GO) analysis by ranking the two data sets separately according to the enrichment p-value (Eden et al., 2007; Eden et al., 2009). In a second step, redundancy in the obtained list of enriched GO terms was reduced by removing overlapping terms using REVIGO (Supek et al., 2011). A maximum of ten GO terms were plotted in Figure 29A and B using the $-\log_{10}$ p-value showing the differentially enriched GO-Terms in the categories 'Function' and 'Component' for the NIP and background proteome. The NIP displayed a strong enrichment for the nucleus as a compartment and various complexes localized to the nucleus. In contrast, the background proteome was enriched for mitochondria, the main localization

site for natively biotinylated proteins, membrane and cytoplasm related GO-Terms (Figure 29A). In case of the GO category 'Function', the NIP was strongly enriched for binding functions to DNA, RAN GTPase or zinc ions (Figure 29B). Probably related to the latter, also transcription factor activity and regulatory activities were enriched. The enriched GO-Term NLS binding is also observed, as one would have expected because of the high content of NTRs within the data set. GO-Terms enriched in the background proteome mainly represent mitochondrial functions. These findings are consistent with the identified 'Component' GO-Terms and the above described IF staining patterns.

Next, I compared the nucleocytoplasmic distribution of the identified proteins to nucleocytoplasmic partitioning experiments done by Wühr and colleagues (Wühr et al., 2015). For each of the 9262 identified proteins they defined a relative nuclear concentration (RNC) by estimating the nuclear and cytoplasmic abundance of each protein. In Figure 29C the distribution of the RNC score of these proteins is compared to all proteins identified across all my experiments and the proteins significantly enriched (adjusted Sime's p-value < 0.01) for the NIP (950) and the background proteome (556). The distribution of all identified proteins is almost evenly spread between high (enriched within the nucleus) and low (enriched within in the cytoplasm) RNC scores. The background proteome is enriched for lower RNC scores, consistent with primarily non-nuclear localization. As expected, the RNC scores for the NIP showed a strong enrichment for higher values.

Active nucleocytoplasmic transport by NTRs is expected to target larger molecular weight (MW) proteins, while smaller proteins can freely diffuse across the NPC. I therefore analyzed the molecular weight (MW) distribution of the NIP and background proteome. The analyzed MW is solely based on polypeptide length, while the native MW might be different due to posttranslational modifications or homo- or heterodimeric complex formations of proteins, of which only one subunit might carry an NLS or NES. Nevertheless, the MW distribution of proteins categorized into the NIP indeed showed a shift towards high MW as expected (Figure 29D).

### 3.3.7 Specificity of NTR

Validation of the workflow and data analysis with known cargos

In the following I will describe the analysis of specific interaction partners of each NTR. For this approach the data analysis was done as described previously in section 3.3.5. For each identified protein a fold change and an enrichment score (ES) in comparison to all other NTRs was determined and Fisher p-values were calculated to identify significant proteins. To validate the data analysis

workflow I first compared my dataset to around 90 previously reported cargos of Transportin (Chook and Süel, 2011; Kimura et al., 2013a; Kimura et al., 2017; Soniat and Chook, 2015; Twyffels et al., 2014). Many of them interact with both isoforms; TNPO1 as well as TNPO2 and therefore I combined the known cargos of both. In Figure 30 the $-\log_{10}$ enrichment scores of TNPO2-BirA and IPO5-BirA were plotted against the $\log_2$ average fold change. The known cargos of Transportin are more enriched in TNPO2 as compared to IPO5. Not all known cargos show enrichment. A small number on the contrary rather showed a trend towards higher ES in IPO5. This might be due to the fact that previous experiments validating certain cargos are done in a cargo-centric way, meaning the strongest or first identified interaction with a specific NTR is counted as the responsible NTR for a specific cargo but most often only a limited number or even just one NTRs were tested for their transport capability. It is however well known that certain cargos can be transported by multiple NTRs with varying interaction strength, which would have been missed by such experiments. Additionally a validated interaction, proven by classical affinity purification approaches, with a certain NTR is done *ex situ*. Therefore the method of proximity labeling *in vivo* of interacting proteins using NTRs of all classes is able to give a more complete picture of the landscape of NTR cargo relationships.

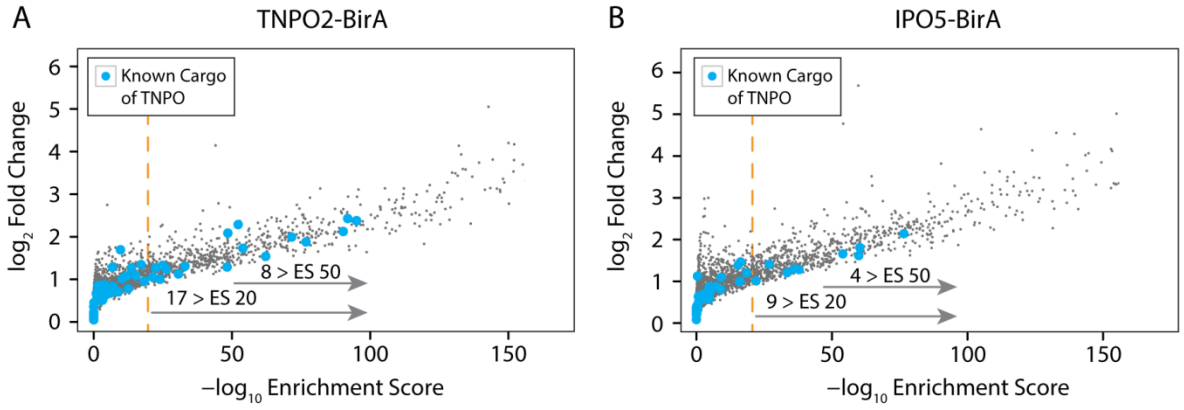


**Figure 30: Known cargos of Transportin mapped on the TNPO2 AP and on IPO5. Cargos are more enriched for TNPO2 than in IPO5. Arrows indicate the number of known cargos for transportins identified with an ES above 20 or 50. The orange dashed line indicates an ES of 20.**

Since many known cargos are not exclusively transported by a specific NTR the validation using the calculated ES might be too stringent, which is of course not the case for the above described NIP. Therefore, I mapped cargos recently identified by Kimura and colleagues (Kimura et al., 2017) using SILAC-Tp, a combination of partly labeling amino acids with stable isotopes, permeabilized cells and mass spectrometry. They identified cargos for various importin β categorizing them into two

different classes with a lower (Category A) and a more stringent (Category B) cut-off. In Figure 31 potential cargos are mapped for IPO4, IPO5, TNPO1, TNPO2 and KPNB1. The median of the distribution of quantile normalized iBAQ scores of all identified proteins for one specific NTR compared to the iBAQ scores of potential cargos of category A is significantly lower in all experiments (labeled with an asterisk). For KPNB1, IPO13 and both Transportins the interaction of potential cargos were tested using a bead halo assay (Patel and Rexach, 2008). Since KPNB1 mainly interacts as an adaptor between importin α bound to the cargo and the FG-Nups of the NPC, both the binding to KPNB1 itself and the binding dependent on importin α were tested for selected proteins. For TNPO2, TNPO1 and IPO13 various cargo proteins were also validated and many of them were identified in my APs for the respective NTRs.

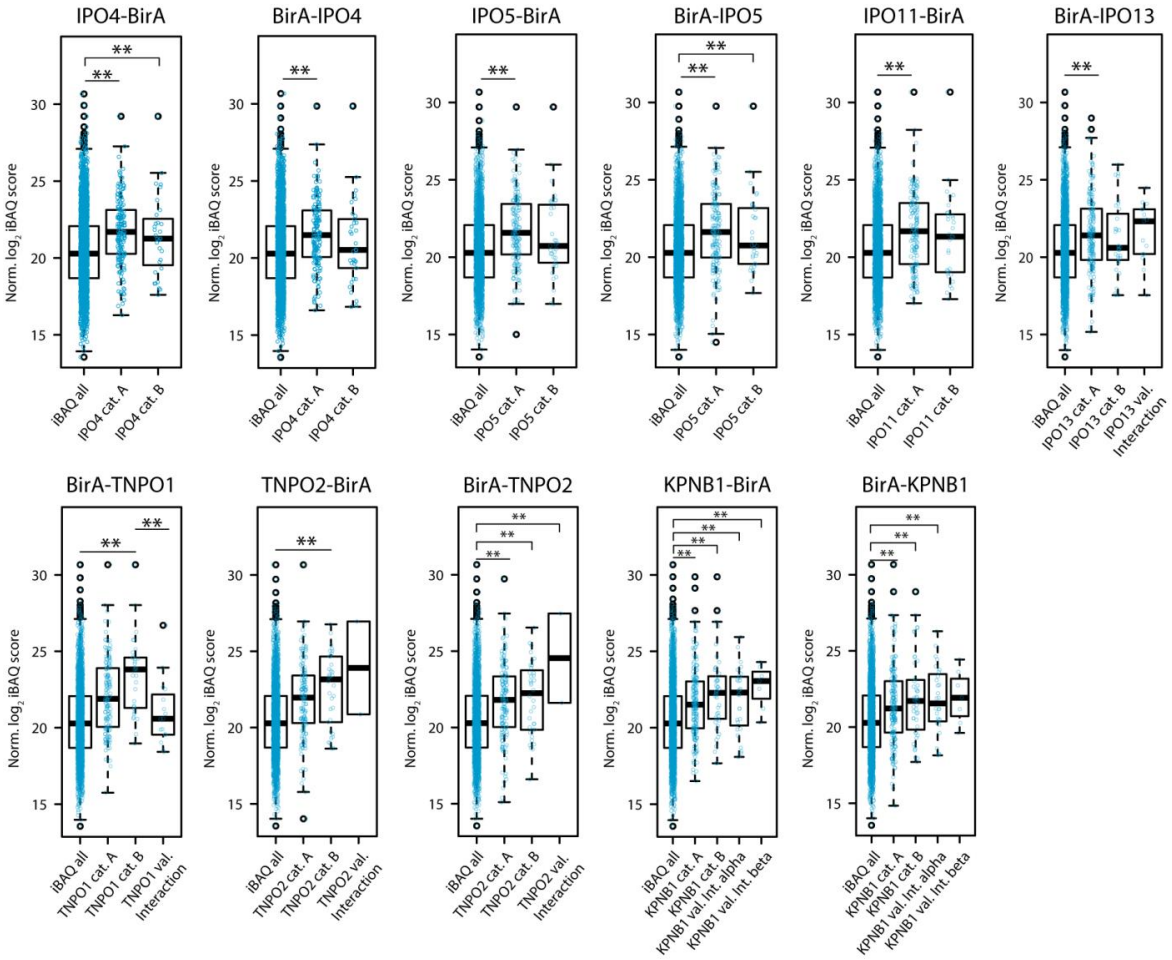


**Figure 31: iBAQ scores of identified cargos using SILAC-Tp by Kimura and colleagues (Kimura et al., 2017; Kimura et al., 2013b) in comparison to all iBAQ scores measured for one AP. The iBAQ scores for determined cargos are significantly higher and labeled with an asterisk. See text for more details.**

To define the specificity of each NTR all further data analysis was done using the ES. The major advantage of this score is that each NTR-BirA experiment can be considered separately and the specificity of each cargo for the respective NTR can be evaluated. Therefore, cargos with a tendency to use multiple NTR are penalized and their ES is lower (Figure 35A). The same is happening to proteins; especially members of the NPC, interacting with NTRs for example while transiting across the NE. Since I tested NTR of almost all classes I gained a broad overview of either specific or shared features of them. Some constructs like BirA-IMA1 as well as BirA-IMA5 might be restrained from being active participants in the nucleocytoplasmic transport system since their IBB-binding site, located at the N-terminus, might be blocked by the fusion to BirA. I decided to keep these APs within the data analysis as internal 'negative controls'.

GO-Term analysis of enriched proteins across all NTR-BirA APs of the alpha and beta class

First, I determined the enrichment of GO-Terms across all experiments to detect functions potentially shared between NTRs. GO-Term enrichments in the categories 'Biological process' and 'Component' were computed using single ranked list of the ES for each experiment. The $-\log_{10}$ p-values of the enriched GO-Term were used to assemble an overall correlation matrix (Figure 32). IMA6-BirA, XPO2-BirA and IMA1-BirA showed the strongest correlation within the category of biological processes. They are enriched amongst others in RNA related processes, like processing and metabolism, primary metabolic processes and chromatin organization. Even though most enriched GO-Terms are shared by them, the GO-Term related to RNA-processing and chromatin organization is more enriched in IMA6 in comparison to the other two importin α indicating a preference for specific cargo sets. XPO2, the recycling transport factor of importin α, correlated strongly with all NTRs of the alpha class. An interesting question imposed by this finding, which needs to be investigated further, is if XPO2 interacts with similar cargos as importin α or if the contact with XPO2 is made when the cargo is released of importin α by Ran-GDP. The biological process GO-Terms cell adhesion, vesicle-mediated transport or cell division are not directly connected to nucleocytoplasmic transport but might point to additional function of these NTRs. These enrichments also correlated with the more prominent cytoplasmic stainings of biotinylated proteins, besides the enrichment within the nucleus, especially for IPO5. Some enriched GO-terms are related to posttranslational modification like protein acetylation and acylation, particularly for IMA1-BirA and IMA5-BirA, or methylation. This could be an indication for various levels of posttranslational modifications, either at the cargos or the NTRs, to fine tune the transport process

either by restricting or enhancing interactions or the transport of proteins involved in posttranslational modifications.

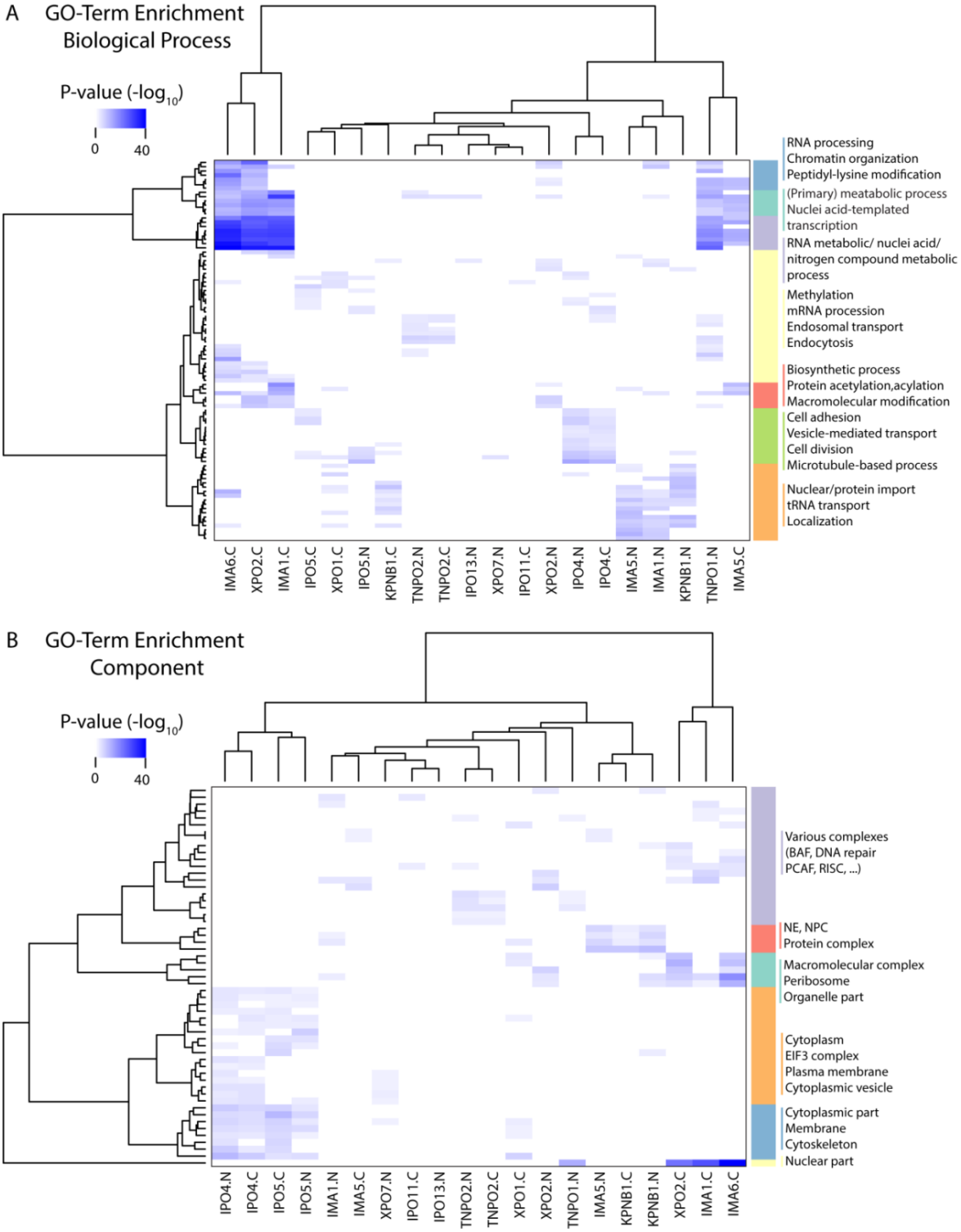


**Figure 32: GO-Term enrichment for single ranked ES of the IPs of the alpha and beta class of NTRs. The $-\log_{10}$ p-values of each GO-Term is used for correlation based on biological process (A) or component (B).**

Cargo spectrum and Network based analysis

In comparison to previous studies on the nucleocytoplasmic transport system, I analyzed a broad spectrum of NTRs and defined their specific cargo spectrum. The number of significantly enriched proteins varies between 888 for BirA-IPO5 and 421 for BirA-TNPO1 when a Fisher p-value cut-off of <0.01 was applied (Figure 33A). A more stringent cut-off decreased the number of significant identifications considerably (Figure 33B).

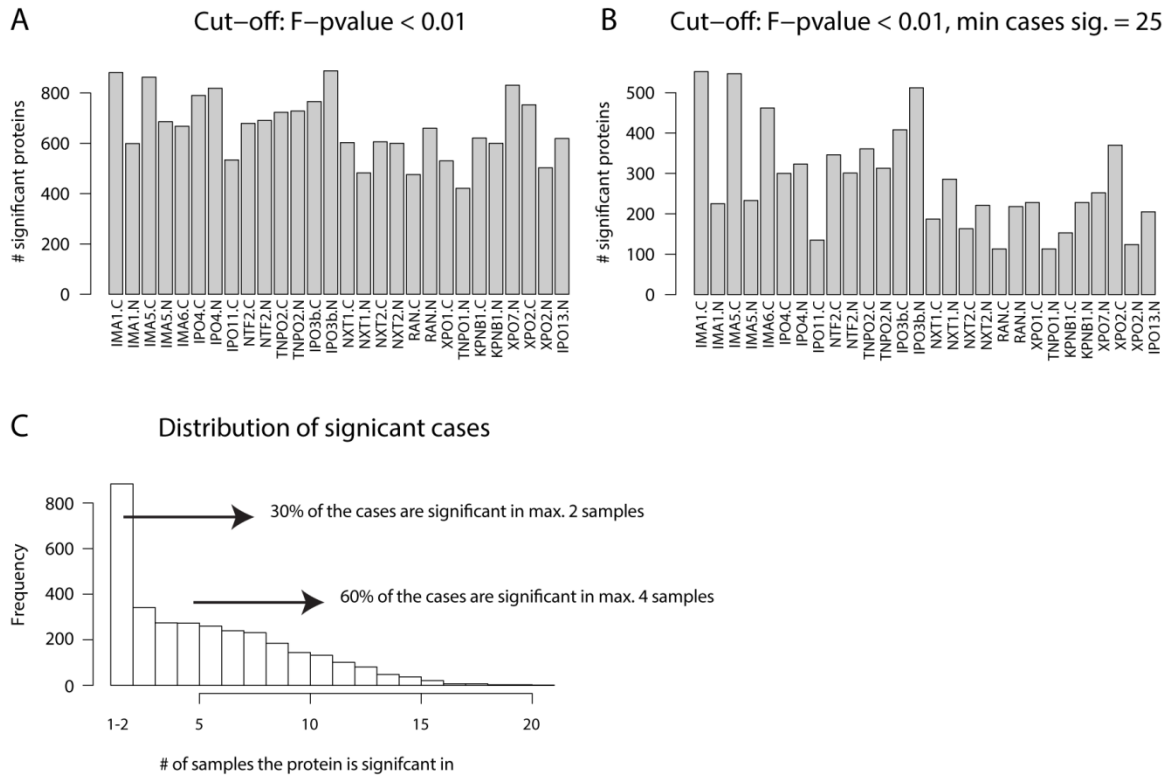


**Figure 33: Number of significant proteins per AP with a Fisher p-value cut-off of 0.01 (A) and minimal 25 significant enrichment cases compared to the other APs (B). (C) Distribution in how samples a protein is significantly enriched (Fisher p-value < 0.01).**

Another indication for the specificity was the number of samples a protein is enriched for (Figure 33C). Around 30% of the significant (Fisher p-value <0.01) protein are only enriched in a maximum of 2 APs. This is most often due to the fact that many NTRs are tagged with BirA on either site. In a maximum of 4 APs over 60% of the significantly enriched proteins fall into. This result clearly showed the advantage of this new data analysis method and its applicability to a broad but related dataset like the nucleocytoplasmic transport system.

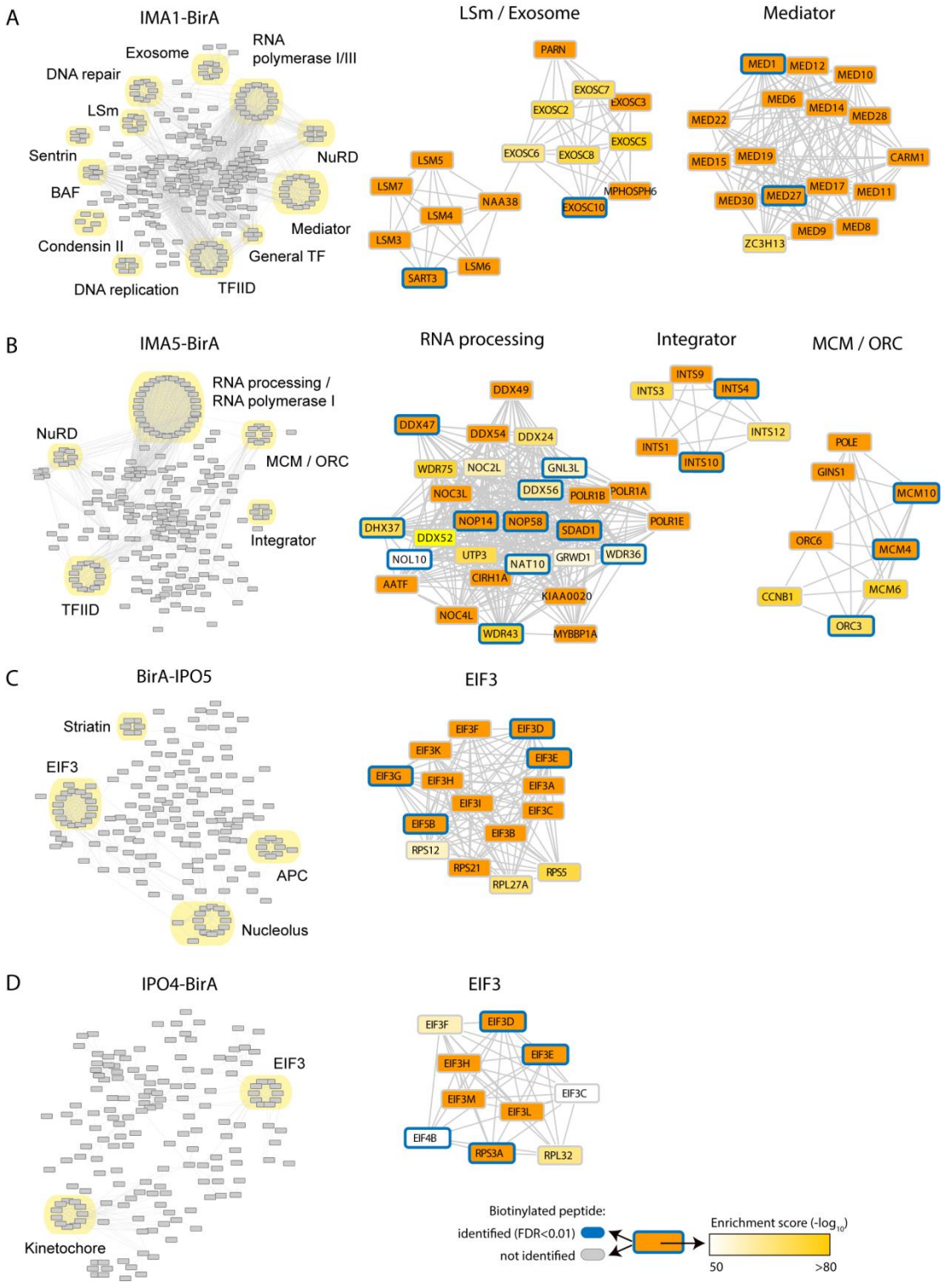


**Figure 34: Network analysis of the top 2 % enriched proteins for IMA1-BirA, IMA5-BirA, BirA-IPO3beta and IPO4-BirA. Selected sub networks with the enrichment score and detected biotinylated peptides are shown. LSm (like Sm), BAF (BRG1- or HBRM-associated factors), NuRD (Nucleosome Remodeling Deacetylase), TF (Transcription factor), TFIID (Transcription factor II D), MCM (Minichromosome maintenance), ORC (origin recognition complex), EIF3D (Eukaryotic translation initiation factor 3 subunit D), APC (Anaphase-promoting complex).**

Next I asked if there are any protein complexes or functionally related protein groups transported by one specific, or a subset of related NTRs. I therefore mapped the identified proteins to the complete STRING network (Szklarczyk et al., 2015), consisting of 15450 interactions of very different types (December 2016). Only high confidence interactions with a STRING interaction score above 0.7 were considered. Network analysis was performed on the most specific top 2 % proteins that were extracted as a sub network consisting of 309 proteins for each NTR-BirA experiment. Overviews of networks retrieved from selected NTRs are shown in Figure 34, and in Figure 40 and Figure 41 in the supplement. The network extracted for IMA1-BirA (Figure 34A) is characterized by many highly specific complexes, like mediator, TFIID (Transcription factor II D) and DNA-directed RNA polymerase 1 and 3 to name just a few. Many of these identified complexes have functions within the nucleolus, which again correlates with the IF staining pattern of biotinylated proteins within the whole nucleus (Figure 23) for this cell line. In case of IMA1-BirA, the Exosome and LSm (like Sm) protein complex, both involved in RNA related processes, and Mediator, involved in transcription, are identified. Many of their subunits were detected with a similar enrichment score but only one or two members had directly biotinylated peptides. This finding suggests that while whole complexes are subject to nucleocytoplasmic transport, only individual proteins contact the NTR.

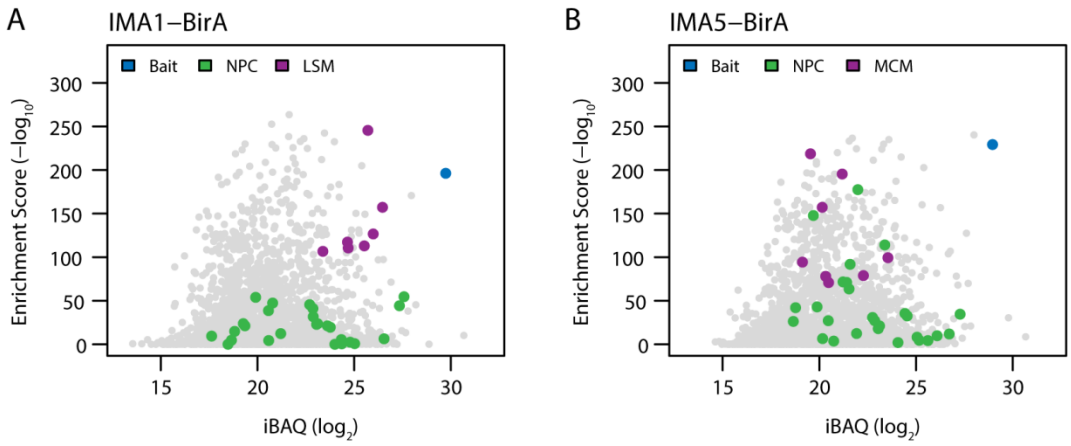


**Figure 35: Correlation between iBAQ and Enrichment Scores in different APs for selected protein complexes.**

In case of the functionally related IMA1 and IMA5, the network analysis further underlined that similar biological processes and functions are associated with both but it also clearly showed the specificity of them (Figure 34B). The TFIID, NuRD (Nucleosome Remodeling Deacetylase) and the DNA-directed RNA Polymerase I are significantly enriched in both APs. The integrator and the pre-replication complex, consisting of ORC (origin recognition complex) and MCM (mini chromosome

maintenance), are stronger enriched in IMA5. The LSm and mediator are exclusively enriched for IMA1-BirA. The specificity of these complexes could also be seen by plotting the $\log_2$ iBAQ scores versus the $-\log_{10}$ ES (Figure 35). Both complexes, the LSm for IMA1-BirA and MCM for IMA5-BirA, are highly enriched and therefore more significant in either AP in comparison to NPC components (common interaction partners of all NTRs).

BirA-IPO5 and IPO4-BirA (Figure 34C and D) both showed enrichments for the EIF3 (Eukaryotic translation initiation factor 3). Not only is this complex enriched but also 2 subunits, EIF3D and EIF3E, are consistently biotinylated in both APs. Well-established protein complexes and also functionally related proteins were detected as exemplified for IMA5-BirA. Many of the identified proteins are involved in various different aspects of RNA processing, including DNA-directed RNA Polymerase I and ATP-dependent RNA helicases. Many of these were also identified to be biotinylated. XPO1-BirA, depicted in Figure 40A in the supplement, did not have any highly enriched protein complexes. This might be due to the fact that Exportin 1 is the main export factor considered to transport the vast majority of proteins out of the nucleus to also compensate for nucleocytoplasmic leakage, and thus is less specific. Therefore the data analysis workflow tailored to identify specific cargos might have penalized many exportin cargos because also importins interact with these proteins. XPO2 so far had a limited known cargo spectrum mainly consisting of the recycling of importin α. The network analysis of XPO2-BirA (Figure 40B in the supplement) revealed the enrichment of the spliceosome and the PCAF (P300/CBP-associated factor) complex. Both are mainly acting within the nucleus but were also reported in the cytoplasm, e.g. depending on acetylation of PCAF. This might indicate a broader, tighter regulated cargo spectrum of XPO2 then anticipated. The cargo spectrum of XPO7 was so far almost completely unknown but with the network analysis (Figure 37C in the supplement) I was able to identify several specifically enriched complexes like the COP9 signalosome, the mediator and integrator complex. KPNB1 acts mainly as an adaptor protein for importin α to mediate the nucleocytoplasmic transport. This is also represented in the network analysis for BirA-KPNB1 (Figure 41A in the supplement). One strongly enriched cluster consisted of NTRs of the importin α class and various different Nups. NTF2-BirA shows a similar enrichment for Nups. This is probably due to the fact that both KPNB1 and NTF2 mainly act as recycling factors for importin α or RAN, respectively, and therefore have to pass the NPC constantly and many more contacts are made.

The above described network analysis gave an overview of the specificity of selected NTRs. Clusters of functionally related proteins or complexes could be identified and have to be further validated. Since only the top 2 % of all proteins were selected, potential interaction below this threshold will be missed. Each network also contained many non-clustered proteins which are all potential cargo proteins and need to be further inspected and matched with the biotinylation dataset. These sites might provide more information about new and so far unidentified localization sequences.

### 3.3.8 Interaction with the NPC

In the last step of the data analysis I had a closer look at the biotinylated peptides eluted from the beads subsequent to on bead digestion. Since the NPC and the passage through this gate is an essential part of the nucleocytoplasmic transport, I focused my analysis on biotinylation sites identified within Nups. I identified biotinylated peptides for several Nups, but the vast majority arose from Nups containing FG-repeats, the main interactors with the NTRs. Biotinylated peptides were also identified in Nups of the scaffold structure; those did not locate into structured regions but into flexible loops. Similar observations were made for FG-Nups. Structured regions like the Ran binding domain in Nup50 and Nup358, the beta propeller of Nup214 or the tetractricopeptide repeats (TPR) of Nup358, which anchors these two proteins into the scaffold structure of the NPC, are rarely biotinylated. This demonstrates that biotinylation preferably occurs in flexible region capable to interact with the NTRs. In Figure 36, Figure 37 and Figure 38 an overview of all biotinylation sites measured in at least 2 out of 4 biological replicates in Nup50, Nup153, Nup98, Nup214 and Nup358 are depicted. The measured intensities were quantile normalized and median centered across each AP.

In Figure 36A the biotinylation sites of Nup50 are shown. Some sites like position 59 and 62 as well as 229, 242 and 249 seemed to get biotinylated by almost all NTR fusion proteins. Especially the sites 59 and 62 were detected with the highest intensity across all APs. The N-terminus contains two Importin α binding sites (Binding site 1: 1-15; Binding site 2: 24 or 29-46) (Ogawa et al., 2010). The second binding site is close to the highly intense biotinylated sites, which might indicate a more specific interaction site for all classes of NTRs in this particular region. A direct binding site is unlikely to be biotinylated because it is normally occupied by the binding partner. The other biotinylation sites cluster around the FG-Repeats of Nup50. Varying preferences for different classes of NTRs can be monitored. The sites 110, 127 and 229 seemed to be highly favored by importin alphas and

Transportin. IPO4, IPO5, NXT1, NXT2 and the exportins, except XPO2-BirA, in comparison did almost never biotinylate any of these sites.

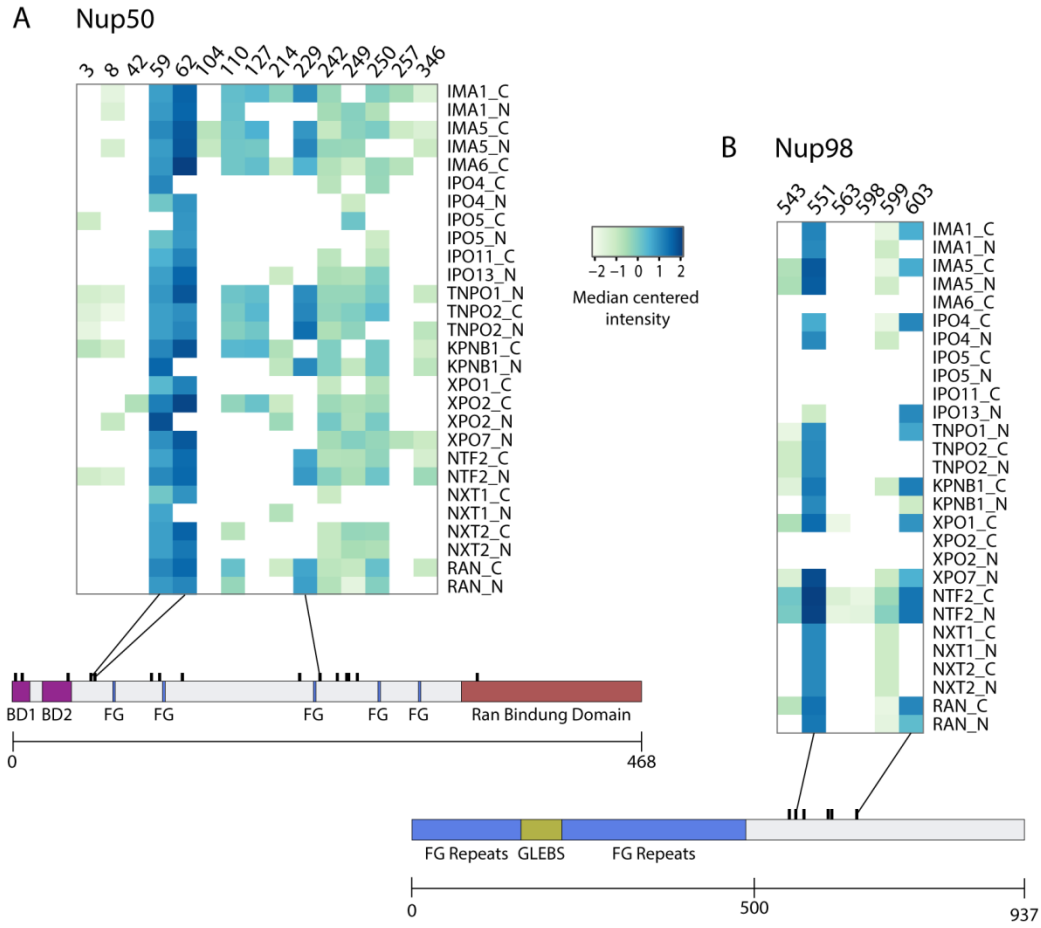


**Figure 36: Normalized and median centered intensities of biotinylated peptides of Nup50 (A) and Nup98 (B). Structural domains like Importin α binding site (BD1 and BD2) and Ran binding domain are highlighted as well as FG-Repeats.**

Nup98, as shown in Figure 37Figure 36B, has only few biotinylated sites. The first half of Nup98 is covered by many FG-Repeats indicating a high interaction rate with NTRs. Unfortunately I could not detect a single biotinylated peptide arising from this region. The reason for this can be explained by the enzyme used for digestion prior to MS analysis. In the first 500 AA almost no arginines and only few lysines are present which get recognized by Trypsin. Thus, the generated tryptic peptides are too long to be measured in MS. The GLEBS domain, dividing the FG repeats, has potentially detectable peptides, but this region is interacting with RAE1 and therefore might be blocked from biotinylation. Position 551 showed the highest biotinylated peptide intensity across many APs. Since Nup98 has multiple locations within the NPC and its function at different positions still needs to be investigated, an interpretation of the obtained data is currently challenging.

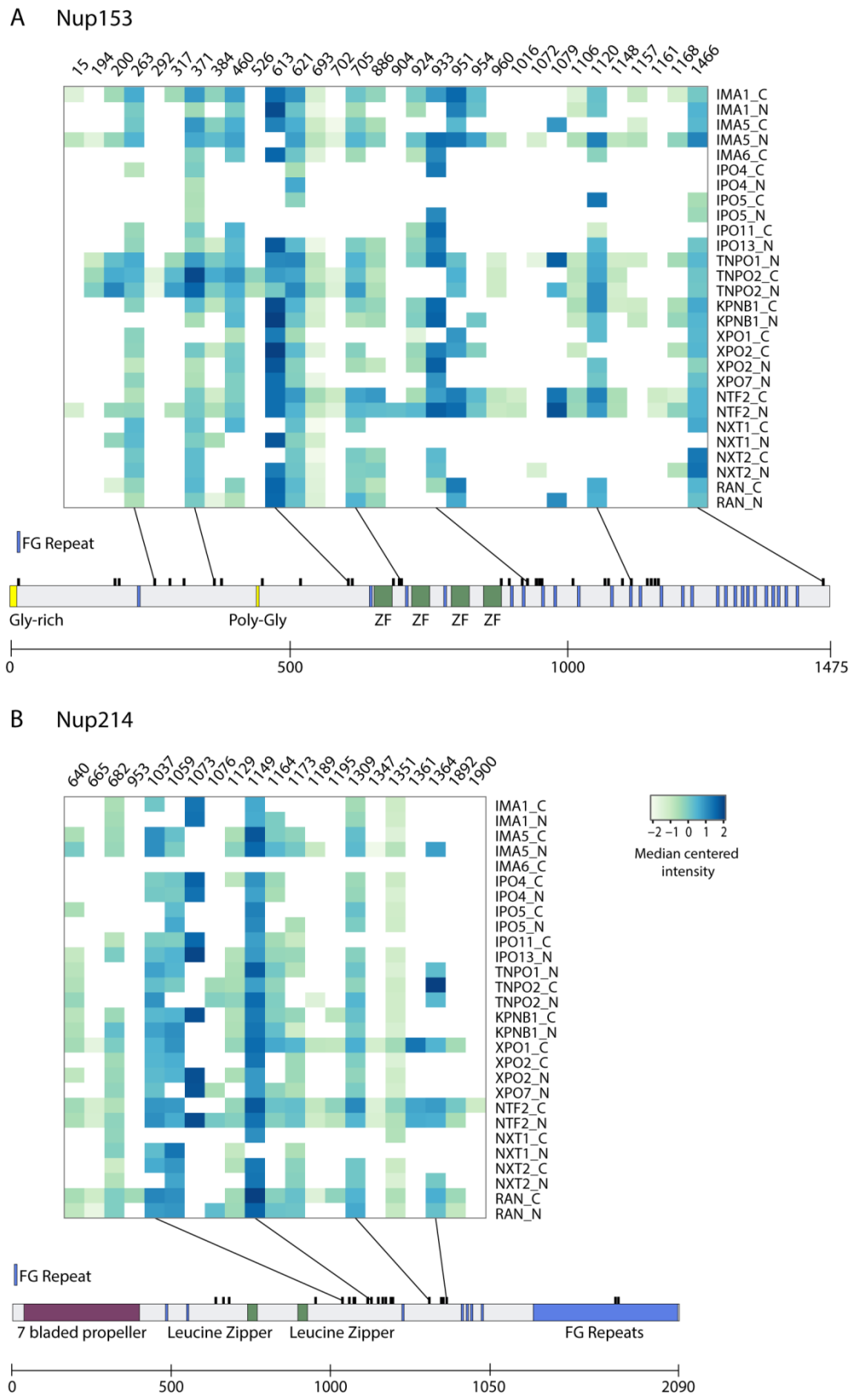


**Figure 37: Normalized and median centered intensities of biotinylated peptides of Nup153 (A) and Nup214 (B). Structural domains like zinc fingers (ZF), leucine zippers and 7 bladed propeller are highlighted as well as FG-Repeats.**

The biotinylation sites of Nup153, located within the nuclear basket, are shown in Figure 37A. Except for the four zinc fingers located in the middle of the protein most of the structure of Nup153 is disordered. The majority of the FG-Nups is accumulated in the last third of the protein, but a lack of tryptic peptides makes it impossible to detect biotinylation sites between the amino acids 1187 up to 1452. The distribution of biotinylation sites seemed to be all over the protein and not related to specific features. A couple of common sites can be detected, but similarly to Nup50, which is also located on the nucleoplasmic site of the NPC, specific positions seemed to be avoided by IPO4, IPO5, NXT1 and NXT2. NTF2 shows the highest amount of biotinylation sites across the whole protein. Similar observations could be made with Nup98, Nup214 and Nup358. Since NTF2 is exclusively shuttling in and out of the nucleus to recycle Ran the contact with the NPC might be more intense.

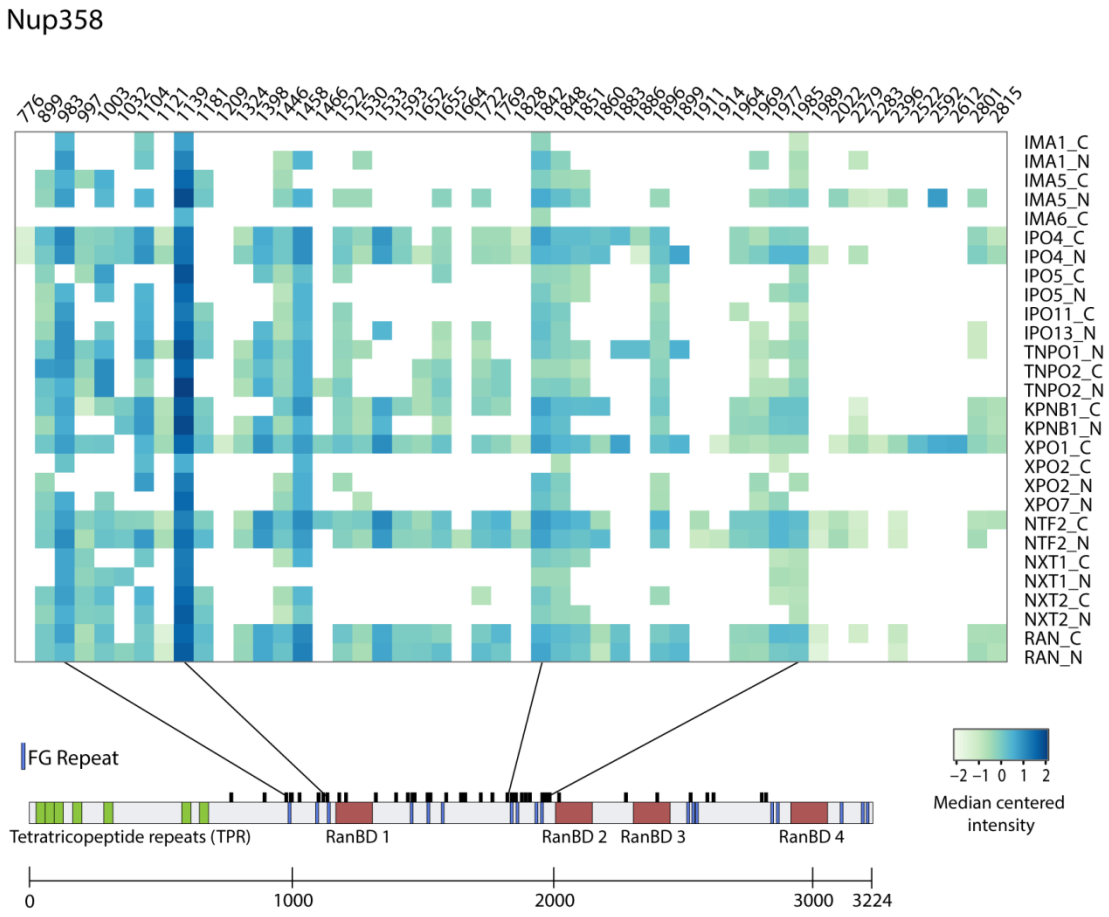


**Figure 38: Normalized and median centered intensities of biotinylated peptides of Nup358. Structural domains like tetratricopeptide repeats (TPR) and Ran binding domains (RanBD) are highlighted as well as FG-Repeats.**

Nup214, a nucleoporin located at the cytoplasmic site of the NPC, and its biotinylation sites are shown in Figure 37B. The structure is characterized by a seven bladed beta propeller anchoring it

into the scaffold structure of the NPC and two leucine zippers. In comparison to Nup50 and Nup153 the biotinylation sites seem to cluster more into a specific subregion of the protein, even though most FG-repeats are located within the C-terminal part of Nup214, which also lacks tryptic peptides to some extent. There were no obvious preferences of any NTR.

The last nucleoporin investigated for biotinylation sites was Nup358, located also at the cytoplasmic site of the NPC (Figure 38). The tetratricopeptide repeats (TPR) most probably anchor it to the NPC. Additionally there are four Ran binding domain (RanBD) spread across the nucleoporins. The most intense biotinylated peptides cluster close to different FG-repeats and also in close proximity to the first and second RanBD. Within the class of importin α only few biotinylated peptides were identified in contrast to Nup50 and Nup153. IPO4, XPO1, NTFs and RAN APs are most enriched for biotinylation sites.

These different patterns of biotinylation sites across various FG-Nups highlight preferred interaction sites and possibly hint towards distinct and shared routes through the NPC taken by various classes of NTRs. The distribution of biotinylation sites seemed to be restricted to unstructured and FG-repeat containing stretches of each protein. Structured regions are almost completely excluded and might not be able to participate in an active nucleocytoplasmic transport. Further investigations are required to ultimately proof specific roles and possibly even regulation of FG-repeat interactions with NTRs.

# 4 Discussion

The NPC and the nucleocytoplasmic transport system are an essential part of a eukaryotic cell. So far both are not fully understood and our knowledge of how the composition of the cytoplasmic and nuclear compartments is maintained is limited. My work on NPC composition during development and on the cargo spectrum of a broad range of NTRs provides a base for understanding how nucleocytoplasmic transport activity might be regulated across different biological conditions.

## 4.1 NPC composition variations in time (development) and space (NE to AL)

NPC composition during neuronal differentiation in Zebrafish

I chose zebrafish as a model to study neuronal differentiation and focused mainly on the compositional variances in the NPC and the abundances changes of NTRs. A protocol had been developed to isolate undifferentiated cells (6 hpf) and enrich for neuronal cells (24 hpf) using FACS sorting on cells expressing a fluorophore under control of a neuronal specific promotor. This protocol enabled me to efficiently remove yolk proteins to avoid highly intense peptides during MS. Also the complete cell dissociation of 24 hpf embryos was crucial to obtain highly enriched single neuronal cells after FACS sorting. The shotgun proteomics with dimethyl labeling and fractionation might be further explored to gain insights into the proteomic changes during development. Several NTRs were significantly changed during neuronal development, most prominently two members of the importin α family with opposing fold changes (Figure 12). Likewise, for mouse embryonic stem cell development into neuronal cells a switch in expressed importin alpha types had been observed (Yasuhara et al., 2013). My observations were in line with these findings indicating a switch or change of abundance of selected NTRs to direct the differentiations processes. The changes within Nups were much milder and only four Nups, namely ALADIN, Nup50, Ndc1 and Nlp1 had a significant fold change in either direction. These four Nups are all part of the dynamic part of the NPC, and probably more exposed to its interaction partners, offering the possibility of influencing the nucleocytoplasmic transport by their presence or absence at the NPC. The nucleoporin Nlp1 and Pom121 showed similar results in experiments carried out in our laboratory by Dr. Alessandro Ori, who compared rat liver and brain tissues. This is additional evidence of these Nups to either direct cell differentiations or to at least be differentially expressed across cell types and probably to be important for their homeostasis. Further work is certainly required to understand underlying

mechanisms, which might e.g. lead to the changes of the subcellular localization of transcription factors and other regulatory elements.

Effects of ALADIN knockdown on zebrafish development has to be further investigated

Since ALADIN showed a significant down regulation during neuronal development and has been targeted by very few studies so far I decided to analyze it in more detail. ALADIN mutations cause the Triple A syndrome that results in neuronal defects. I used the CRISPR/Cas9 system to induce a knockdown of ALADIN in the first cell stage after fertilization to potentially affect the whole embryo. Since heterozygous mutations, at least in humans do not lead to symptoms of the Triple A syndrome, I generated a $F_1$ generation, passing on the mutation in the germ line, to cross suitable fish to gain a homozygous $F_2$ generation. So far only few fish laid eggs with potential homozygous mutations. Amongst the successfully fertilized eggs, some embryos had severe obvious defects (Figure 15). So far I could not link these defects to homozygous or compound heterozygous mutations within the ALADIN gene. The embryos also had a limited survival rate ranging between 6 to 10 dpf. The next steps would involve linking these defects to the mutation in ALADIN, investigating the contribution of maternal mRNA as well as protein in the first days of development and experiments to rescue this knockdown. Possible rescue strategies would include injection of mRNA and comparing the survival rate to not injected embryos. Another option would be rescue by using a BAC (bacterial artificial chromosomes) clone. Since it is a genomic clone with its own regulators; the proper expression in space and time is more likely. BAC clones could also be used as a reporter line of expression in combination with GFP to identify tissues and developmental stages in which ALADIN has an important contribution. This would allow evaluating phenotypic changes in various organs and validating the nucleocytoplasmic transport system with and without ALADINs presence at the NPC.

Annulate lamellae: A repository for NPCs in a more simple design

Annulate lamellae are a pool of NPCs located in the ER. The question in the beginning was how they differ to NPC located in the NE and if and how they might contribute to the rapid growth of the nuclei in the first 13 divisions during *Drosophila* embryogenesis. By subcellular fractionation and subsequent shotgun proteomics I was able to detect compositional changes of the NPC located at the NE, the AL and the soluble Nups in the cytoplasm (Figure 17). Nup358 and ELYS, the only asymmetrically Nups present in AL-NPC, might function as important Nups to keep the scaffold

structure. It has been shown that in Hek cells the loss of Nup358 also leads to the loss of one of the Nup107 subcomplex rings located on the cytoplasmic site of the NPC (von Appen et al., 2015). Knockdown of Nup358 resulted in reduced intensity of all the Nup107-subcomplex members by about 25 %, which is consistent with the loss of one Nup107 subcomplex ring. Since Nup358 has this important scaffolding function it might be possible that Nup358 fulfills this task in AL-NPC also at the "nucleoplasmic side". If Nup358 is actually present on both sites, and therefore also symmetrically distributed in AL-NPC, needs to be investigated further. The missing of other asymmetrically Nups, like the Nup62 subcomplex, Nup214, Mbo and Mtor, could occur for two reasons. First, it might be not possible to build fully asymmetric NPCs in the cytoplasm because specific environmental requirements only present in the nucleoplasm are necessary for successful assembly. Second, the missing asymmetrical Nups could be a huge advantage for insertion into the NE, because the orientation cannot be violated. The necessity of asymmetrically distributed Nups, as soon as the AL-NPCs become part of the NE is most probably covered by the high cytoplasmic concentration of these Nups. This allows building a functional asymmetric NPC with competence for transporting cargos in either direction.

## 4.2 The highly specific cargo spectra of NTRs

The NTRs-cargo interactions and transport in and out of the nucleus are an important process in all eukaryotic cells. Thousands of different proteins need to be transported and only a limited number of different NTRs are present. Most of the NTRs are studied poorly since the interaction is transient and the passage through the pore ultrafast. By applying proximity labeling I was able to overcome many of the previous challenges and to measure interactions *in situ*. It allowed me to test a broad range of NTRs, covering the class of importin α and β, with predominant functions in import, export or bidirectional cargo transport, NTRs with more specific functions like RNA related export or recycling of Ran back into the nucleus to maintain the essential Ran gradient across the different compartments.

Data analysis workflow to detect and validate cargos

Data analysis of various affinity purified samples often had the disadvantage to detect highly specific interactions partners on one side but not within the samples on the other side. Therefore it was not possible to quantify and statistically analyze them in comparison to other APs. I developed a strategy to overcome these weaknesses by imputing missing values in a stringent and selective way.

Data can be missing due to complexity of measurements, meaning overlaying spectra cannot be assigned correctly and no intensities are measured. Due to the stringent definition of cases missing at random (MAR) and four biological replicates used, these data points missing by chance can be distinguished from those being below the detection limit and the missing data at least to some extent could be restored for quantification based on values of the same protein in different biological replicates. I demonstrated the effectiveness and reliability of this workflow for MAR cases in comparison to NA values and none of the tested proteins received a higher fold change and the assigned p-values were even more conservative as compared to classical workflows.

The NIP defined the transported proteome of a cell

Around 25 % of the quantified proteins in the NIP were significantly enriched. These proteins include NTRs, nucleoporins and potential cargos. Many cargos perhaps still were not detected even though a broad spectrum of NTRs was investigated, not all are covered. Some cargos might also use highly specific NTRs or have no need to be transported due to cell type and stage. By comparing the identified proteins to previous studies of subcellular distribution most proteins of the NIP are more enriched within the nucleus (Figure 29C). The exportins used in this study did not seem to shift the distribution towards cytoplasmic proteins. Exportins at the first glance seem to mainly export proteins which were imported once, since no significant enrichment for lower RNC, equal to a cytoplasmic localization, was detectable. That indicated that export is not exclusively used for proteins which end up accidentally inside the nucleus by for example leakage of NPCs and uncontrolled diffusion in the nucleus or entrapment of cytoplasmic proteins during cell divisions. This approach was the first step in data analysis. I identified many proteins highly enriched and significant for interaction with NTRs but I was not able to distinguish between common interactions partners and highly specific cargo proteins for selected NTR.

Enrichment Score calculations define specific cargo spectra

So far most of the cargo spectra mapping approaches were limited to XPO1 or importin β. Previous experiments used irreversible inhibitors, like leptomycin B, or massive interventions like digitonized cells. Applying the BioID method to map the landscape of the NTR cargo relationship was the first approach covering many NTRs and potential interactions were labeled *in situ* with minimal interference.

The calculations of enrichment scores for each NTR separately in respect to all others allowed the discrimination between common interaction partners and specific cargos. NPC members were highly specific for the NIP but got a penalized enrichment score since it is a common interaction partner for all NTRs, because each transport event and the recycling of the NTRs afterwards require the passage and interaction with the NPC.

Position of BirA fused to NTRs partly affects the transport competence

Most of the NTR-BirA cell lines were generated by fusing the biotin ligase either to the C- or N-terminus of a given NTR. Some NTRs like, NTF2, TNPO2 or IPO5 seemed to be unaffected according to WB or IF stainings. For IMA1 and IMA5 the expression of the FLAG-tag measured in WB was lower and therefore also the amount of biotinylated peptides. A possible explanation might be the location of the IBB domain at the N-Terminus of importin α. The IBB domain is required for binding to the adaptor KPNB1. The fusion of BirA there could make the binding impossible or weaker. In addition to the IBB domain, an autoinhibitory sequence, with high similarities to an NLS is present which blocks the NLS binding site (Goldfarb et al., 2004). The fusion of BirA might weaken this interaction and potential cargos are capable of binding but the formation of the complex with KPNB1 is hampered. This could explain the decreased level of biotinylation since a complex of importin α and cargo can be formed and biotinylation occurs but no active transport is possible. These cargos have to be released again to bind to functional NTRs for efficient import. To my knowledge there is no mechanism known for a cargo release from importin α within the cytoplasm for cargos bound by mistake. The IF staining for biotinylated proteins was weaker within the nucleus and the nucleoli were almost completely excluded for BirA-IMA1 and BirA-IMA5 (Figure 24). The missing signal within the nucleoli might be an indication for an extended function of IMA1 in not only enabling a transport across the NE but also the transport to its final destination and place of action. IMA1-BirA had a strong enrichment for complexes, which are localized within the nucleoli like the mediator, DNA-directed RNA polymerase I and III and TFIID. For BirA-IMA1 I identified only few and much less significantly enriched members of the aforementioned complexes.

My results showed the importance of fusing both termini to BirA. Even for cell lines like TNPO2 or IPO5 with highly similar enriched proteins for BirA fused either to the C- or N-terminus; the biotinylation sites might provide additional information about binding sites or additional proteins interacting with the actual cargo protein.

Known cargos showed a trend to be enriched for the respective APs

To validate the cargo candidates, I compared them to a previously published dataset (Kimura et al., 2017). They applied a method called SILAC-Tp, which combined heavy labeled cytosolic extracts, digitonized cells and only a selected number of NTRs to monitor the import of cargos depending on these NTRs. All iBAQ scores of identified proteins were compared with a subset of iBAQ scores corresponding to potential cargos identified by Kimura and colleagues (Figure 31). For each NTR the proposed cargos had a significant higher median iBAQ score but still the distribution was huge. These findings might have different reasons. One major difference was the experimental set-up *in situ* for BioID and *in situ*, for SILAC-Tp. It is known that cargos can bind to different NTRs, but *in vivo*, cargos might have preferred NTRs. A second explanation for the observed differences might be the necessary exchange of the cytoplasm for SILAC-Tp which might disrupts cytoplasmic organizations and regulatory networks. Additionally the altered protein concentration might interfere with transport processes.

The similar trend for cargos identified by another method made me confident about the applicability of BioID in large scale and revealed additional benefits compared to other methods. BioID can be used *in situ*, that means all intracellular regulation networks are present. Since the normal set of NTRs are present, cargos only bind and get biotinylated by the NTR-BirA construct if it is the favored transport pathway. Another advantage is also the possibility to monitor exporting and bidirectional NTRs.

Cargo spectra overlap but also show specific enrichments between related NTRs

I was able to map highly specific cargo spectra by systematically mapping NTR-cargo relationships *in situ* in an unbiased way using BioID. The use of an enrichment score calculated for each NTRs penalized common interaction partners like the NPC and boosted NTR specific cargos. I also showed that the overall intensity measured for proteins does not correlate with and affects their degree of specificity (Figure 35). Already significantly enriched GO-Terms revealed clustering between different preferences in cargos (Figure 32). With the network analysis I was able to identify whole complexes specifically enriched for a certain NTR (Figure 34, Figure 40, Figure 41). Noteworthy were the biotinylated proteins identified in each complex. For most complexes only a minor fraction of the members had a biotinylation mark. It is possible that only some subunits are transported by the selected NTRs and the assembly of the complex occurs within the nucleus or that a preassembled

complex or subcomplex is transported and only some proteins are responsible for the contact to a NTR. Since most of the identified complex member had similar enrichment scores the second option might be more likely. Additionally almost no other subunits were identified in any other NTR AP sample, despite the fact that a broad range of NTRs was covered. Many other proteins that are not part of a prominent complex, were also significantly enriched and in part biotinylated. The specific biotinylation sites have a potential to elucidate the interaction of cargos with NTRs better. Do biotinylation sites tend to be closer to potential NLS/NES recognition sites? Are specific sites preferred or is the distribution across a protein random? What are the different biotinylations patterns when comparing C- and N-terminally tagged NTRs? Since I detected specifically enriched biotinylation sites for FG-Nups, which are better understood in terms of their interaction with NTRs, the biotinylation seemed to be directed to certain sites and not randomly distributed. Therefore biotinylation sites could be used to predict recognition sequences such as e.g. NLS and NES.

My dataset using BioID and mapping the cargo spectrum with the additional information of the biotinylation sites is a rich resource and starting point for further investigations and validation experiments for selected complexes and single proteins. Additional levels of regulation like posttranslational modifications have to be considered as well. The LSm complex, highly specific for IMA1-BirA, for example is known to function inside the nucleus but also within the cytoplasm related to RNA processing (Spiller et al., 2007). The complex consists of a core with 6 proteins and 2 variable members depending on the localization. The exact import, depending on a complex formation, and export routes, probably involving XPO1, and underlying signals are incompletely understood and perhaps can be investigated further based on this dataset. For XPO1 no obvious complexes were detected to be significantly enriched; in line with the fact that it interacts with a huge fraction of the nuclear proteome. This wide cargo spectrum can be explained by the NES sequence recognized by XPO1 which is present in many proteins (Güttler and Görlich, 2011).

Potential function of NTRs beyond the scope of nucleocytoplasmic transport

The network analysis also highlighted some at the first glance unexpected identifications. Complexes or protein groups like AP2-complex for IPO5 and IPO4, centriole related proteins for KPNB1 or ArfGAPs for NTF2 are not directly linked to nucleocytoplasmic transport. The AP2 complex, normally involved in clathrin-mediated endocytosis, has been shown to colocalize with importin β and Nup153 (Kitagawa et al., 2008). The staining pattern for IPO5 also showed an enrichment of biotinylated proteins within the cytoplasm. Further experiments are necessary to prove an

interaction of IPO4 or IPO5 with AP2 either at the NE or within the cytoplasm and identify the potential functions. The involvement of importin β and Ran in the transport within cilia has been shown previously (Fan and Margolis, 2011; Lim and Tang, 2015). The exact mechanism, directions of transport (within the cilia, to and from the nucleus) and NTRs involved needs to be investigated further. These examples show the great potential of the BioID method to capture additional functions of NTRs within the whole cell.

Different NTRs prefer different routes through the NPC

So far it was extremely difficult to map interaction of NTR with Nups *in vivo* in a fully functional system. Some experiments were done with artificial pores or selected NTRs which were recombinantly overexpressed, purified and followed by interaction studies with selected FG-Nups. Since FG-Nups are completely disordered no structures can be obtained. Nucleocytoplasmic transport relevant features are completely missing in structural studies of the NPC. Simulations of FG-Nups interacting with NTRs are partially possible but far away from modeling a whole NPC-NTR-cargo interaction network. In addition, most of the interactions are transient and cannot be captured using any kind of affinity purification or crosslinking strategies. Since the BioID system does not require any stable interaction and just leaves a biotinylation site like a footprint, the trace can be detected later. The elution was done in two consecutive steps first by an on-bead digestion followed by an elution using ACN/TFA. The second fraction was highly enriched for biotinylated peptides and many of them were assigned to Nups. The biotinylation sites on Nups occurred exclusively on FG-Nups or flexible loops of scaffold Nups. Also the structural domains of FG-Nups were practically excluded from biotinylation sites. By mapping all measured intensities on the domain structure of the respective Nups common sites with high intensities for all NTRs but also sites which only interact with a subset of NTRs could be detected. Nup50 for example has two stretches binding to importin α at the N-terminus and exists in a long and a short isoform (Figure 36). In the long isoform both binding sites are present and promote the release of the cargo, the short version with only one binding site stabilized the NTR-cargo interaction (Ogawa et al., 2010). Two positions close to the second importin α binding site are highly biotinylated by importin α but all other NTRs showed the same trend. One disadvantage of the BioID labeling procedure is that the actual interaction sites cannot get biotinylated because the interaction makes them inaccessible for labeling. The distance to the actual binding site is difficult to predict because it depends on the position of BirA in relation to the binding site, the protein structure of the interaction partner,

potential additional proteins bound and the accessibility of free lysines on the surface. Since FG-Nups are mostly unstructured the likelihood that the biotinylation site occurs close to the actual binding site is higher. If the strongly biotinylated positions 59 and 62 within Nup50, which was also detected for importin β, has something to do with the importin α binding domains need to be investigated further. Perhaps these binding domains can also interact with other NTRs. Importin α and the Transportins have a similar biotinylation pattern for Nup50, but IPO4, IPO5, IPO11 and IPO13 are missing most of them and might point to different transport rotes. At first sight Nup98 seemed to be less biotinylated even though many FG repeats cover the first half of the protein. By a closer look at the amino acid distribution it became apparent that only few lysines are present. Biotinylation of these might prevent tryptic digestion. To a lesser extent these difficulties also occurred for other FG-Nups. To further investigate the preferred transport routes it might thus be necessary to use different digestion enzymes. Nup214 had a similar biotinylation pattern for all importin α and β and it also had been shown before that Nup214 depleted cells had no import defects (Hutten and Kehlenbach, 2006) but export mediated by XPO1 required the presence of Nup214. This might explain the stronger biotinylation of Nup214 for XPO1. Nup358 was the only FG-Nup tested to have increased biotinylation site coverage by IPO4. The reason and perhaps function of this has to be investigated further.

The mapping of the biotinylation sites and the comparison of patterns of different NTRs definitely opened a new perspective on transport routes. I could show that classes, like importin α, importin β or transportins, have class specific preferred interactions and some like IPO4 have also an even more defined interaction pattern. The investigations of mutations at specific FG-repeats might advance the understanding of transport routes. Additionally, knock-down cell lines of selected FG-Nups could be used and the effects on the nucleocytoplasmic transport monitored. The question to answer would be if the transport is less efficient or another route through the NPC is taken.

The application of the BioID method on the nucleocytoplasmic transport system opened up a new perspective. It allowed me to identify specific functionally related proteins and complexes and assign them as cargos to selected NTRs. The detected biotinylation sites on cargos can be used for recognition sequence prediction and the ones identified on the FG-Nups to better understand the transport pathways taken. This extensive dataset can serve as a rich resource for follow-up experiments to validate selected interactions in various conditions.

# 5 Material and Methods

## 5.1 Zebrafish

### 5.1.1 Fish strains

Zebrafish (*Danio rerio*) strains were maintained following standard protocols (Westerfield, 2007). Embryos were raised in E3 buffer (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$) at 26 to 30 °C. All zebrafish experiments were conducted on embryos younger than 3 dpf in collaboration with Darren Gilmour's Lab at EMBL, Heidelberg. For isolation of undifferentiated cells a wild type strain (golden) and for neuronal cells the NBT-DsRed strain was used. In this fish strain the red-fluorescent-protein is under the control of a neural-specific tubulin promoter of *Xenopus* (Peri and Nüsslein-Volhard, 2008).

### 5.1.2 Isolation of undifferentiated and neuronal cells from zebrafish embryos

Early embryos (6 hours post fertilization (hpf))

Wild type embryos were removed from their chorions using 1 ml of pronase (stock 30 mg/ml) in 40ml buffer E3 and incubated for 10-15 min with gentle shaking every 2 min in a small beaker. The supernatant was removed and the embryos were washed 4-5 times using buffer E3. The embryos were splitted into batches of around 250-300 per 1.5 ml tube. 1 ml of deyolking buffer (55 mM NaCl, 1.8 mM KCl, 1.25 mM $NaHCO_3$) was added per tube und everything passed twice through a 200 µl pipet tip. The tubes were incubated at RT at 1100 rpm for 5 min and afterwards spun at 300 g for 30 sec to remove the supernatant. The embryos were washed using 1 ml of wash buffer (110 mM NaCl, 3.5 mM KCl, 2.7 mM $CaCl_2$ and 10 mM Tris/HCl pH 8.5), shaken at 1100 rpm at RT for 2 min and spun as above to remove the supernatant. The wash step was repeated twice. The deyolked and dissociated embryos were resuspended in 400 µl wash buffer and passed through a 40 µm cell strainer to remove undissociated cells. The merged cells were washed as above and resuspended in 110 µl PBS and counted using a hemocytometer.

Late embryos (24 hpf)

NBT-DsRed: After 24 hpf the embryos were sorted for NBT dsRed positive embryos. Up to the addition of deyolking buffer all steps were the same as for early embryos. After the addition of 1 ml

deyolking buffer per tube the embryos were passed 10 times through a 1000 µl pipet tip, followed by washing twice with deyolking buffer and four times with washing buffer. For better cell dissociation the embryos were rinsed once with Accumax (Millipore) and then resuspended in 1ml Accumax and transferred to a 15 ml tube. The embryos were incubated at RT for 5 min at the lowest speed of the vortex mixer. The embryos were dissociated by pipetting for 2 min using a 1000 µl pipet tip, 2 min incubation on the vortex mixer and 1 min of additional pipetting. The cells were spun for 1 min at 300 g at RT and washed twice using 1 ml of PBS with 0.5 % BSA. 400 µl of PBS with 0.5 % BSA was used per tube to resuspend the cells afterwards passed through a 40 µm cell strainer and merged. DNAse I (Roche, Cat. No 11284932001, stock 20.000 U/mL, 10 mg/ml in water) 170 U/ml and 10 mM $MgCl_2$ was added. Cells expressing the DsRed fluorescent protein were FACS sorted to obtain a highly enriched fraction for neuronal cells.

### 5.1.3 MS sample preparation and dimethyl labeling

Cells were lysed by addition of Rapigest (Waters) and urea to a final concentration of 0.2 % and 4 M, respectively, and sonicated for 3 x 30 sec to shear chromatin. Before protein digestion, samples were stored at -80 °C. Samples were quickly thawed and sonicated for 1 min. DTT was added to a final concentration of 10 mM and incubated for 30 min with mixing at 800 rpm to reduce cysteines. Then 15 mM of freshly prepared iodoacetamide (IAA) was added and samples were incubated for 30 min at room temperature in the dark to alkylate cysteines. Afterwards, 1:100 (w/w) LysC (Wako Chemicals GmbH) was added for 4 h at 37 °C with mixing at 800 rpm. Then urea concentration was diluted to 1.5 M with HPLC water and 1:50 (w/w) trypsin (Promega GmbH) was added for 12 h at 37 °C with mixing at 700 rpm. Afterwards the samples were acidified with 10 % TFA and the cleavage of Rapigest was allowed to proceed for 30 min at 37 °C. After spinning the sample for 5 min at 13000 g at room temperature the supernatant was transferred to a new tube to proceed with peptide desalting.

For desalting and cleaning-up of the digested sample, C-18 spin columns (Sep-Pak C18 Classic Cartridge, Waters) were used. A vacuum manifold was used for all washing and eluting steps. First the columns were equilibrated with 100 % methanol and then washed twice with 5 % (v/v) acetonitrile (ACN) and 0.1 % (v/v) formic acid (FA). The sample was loaded two times and then the column was washed 2 times with 5 % (v/v) ACN and 0.1 % (v/v) FA. The undifferentiated cell samples were labeled using an ‘light’ labeling reagent and the FACS sorted neuronal cells were labeled using a ‘intermediate’ labeling reagent inducing a mass shift of 28 or 32 Da respectively (Boersema et al.,

2009). Formaldehyde and the D-isotopomer of formaldehyde react with primary amines of peptides (N-terminus and side chains of lysines) and generate a mass shift of 4 Da. The labeling reagents consisted of 4.5 ml 50 mM sodium phosphate buffer (mixture of 100 mM $NaH_2PO_4$ and 100 mM $Na_2HPO_4$), pH 7.5, 250 µl 600 mM $NaBH_3CN$ and 250 µl 4 % formaldehyde for light or 4 % deuterated formaldehyde for intermediate labeling reagent, per sample. After the labeling procedure the column was washed 2 times with 5 % (v/v) ACN and 0.1 % (v/v) FA. For elution 50 % (v/v) ACN and 0.1 % (v/v) FA was used. The sample was then dried with the help of a speed-vac, and resuspended in 20 mM ammonium formate (pH 10.0), to be ready for high pH reverse-phase fractionation directly. To dissolve the dried samples, they were vortexed, mixed for 5 min at maximum speed in a thermomixer and sonicated for 90 s. The samples were stored at -20 °C.

### 5.1.4 High pH reverse-phase fractionation

Offline high pH reverse-phase fractionation was performed using an Agilent 1200 Infinity HPLC System equipped with a quaternary pump, degasser, variable wavelength UV detector (set to 254 nm), peltier-cooled autosampler, and fraction collector (both set at 10°C). The column was a Gemini C18 column (3 µm, 110 Å, 100 x 1.0 mm, Phenomenex) with a Gemini C18, 4 x 2.0 mm SecurityGuard (Phenomenex) cartridge as a guard column. The solvent system consisted of 20 mM ammonium formate (pH 10.0) as mobile phase A and 100 % acetonitrile as mobile phase B. The separation was accomplished at a mobile phase flow rate of 0.1 ml/min using the following linear gradient: 99 % A for 2 min, from 99 % A to 37.5 % B in 61 min, to 85 % B in a further 1 min, and held at 85 % B for an additional 5 min, before returning to 99 % A and re-equlibration for 18 min. Thirty two fractions were collected along with the LC separation that were subsequently pooled into 10 fractions. Pooled fractions were dried in a speed-vac and resuspended in 5 % (v/v) ACN and 0.1 % (v/v) FA and then stored at -80 °C until LC-MS/MS analysis.

### 5.1.5 Shotgun proteomics

For shot-gun experiments, samples were analyzed using a nanoAcquity UPLC system (Waters GmbH) connected online to a LTQ-Orbitrap Velos Pro instrument (Thermo Fisher Scientific GmbH). Peptides were separated on a BEH300 C18 (75 µm x 250 mm, 1.7 µm) nanoAcquity UPLC column (Waters GmbH) using a stepwise 145 min gradient between 3 and 85% (v/v) ACN in 0.1% (v/v) FA. Data acquisition was performed using a TOP-20 strategy where survey MS scans (m/z range 375-1600) were acquired in the orbitrap (R = 30000 FWHM) and up to 20 of the most abundant ions per full

scan were fragmented by collision-induced dissociation (normalized collision energy = 35, activation Q = 0.250) and analyzed in the LTQ. Ion target values were 1000000 (or 500 ms maximum fill time) for full scans and 10000 (or 50 ms maximum fill time) for MS/MS scans. Charge states 1 and unknown were rejected. Dynamic exclusion was enabled with repeat count = 1, exclusion duration = 60 s, list size = 500 and mass window ± 15 ppm.

### 5.1.6 Quantitative data analysis

For the quantitative label-free analysis, raw files from the Orbitrap were analyzed using MaxQuant (version 1.3.0.5) (Cox and Mann, 2008). MS/MS spectra were searched against the *Zebrafish* entries of the ENSEMBLE entries (database downloaded 2013_03, 42154 entries) using the Andromeda search engine (Cox et al., 2011). The search criteria were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues, unless followed by proline); 2 missed cleavages were allowed; carbamidomethylation (C) was set as fixed modification; oxidation (M) and acetylation (protein N-term) were applied as variable modifications; dimethyl was set as a label either light or intermediated (K and peptide N-term), if applicable; mass tolerance of 20 ppm (precursor) and 0.5 Da (fragments). The retention times were matched between runs, using a time window of 2 minutes. The reversed sequences of the target database were used as decoy database. Peptide and protein hits were filtered at a false discovery rate of 1% using a target-decoy strategy (Elias and Gygi, 2007). The “Ratio H/L” of the proteinGroups.txt output of MaxQuant was used for further analysis. All comparative analyses were performed using R version 3.0.1. (R Core Team, 2012). Only ratios identified in at least 2 replicates were considered. The fdrtool R package (Strimmer, 2008) was used to fit a two components model on centered ratio distributions. Protein groups with a ratio belonging to the non-null component (q-value < 0.1) were considered as differential expressed.

### 5.1.7 gDNA design and generation

To design suitable target sequences upstream of available PAM sequences (= NGG) the web based software ZiFiT Targeter (http://zifit.partners.org/ZiFiT/CSquare9Nuclease.aspx) was used. The target length was set to “20” and the “T7 promotor” was selected. Two target sites within the first exon with only few potential off target were selected. Oligos with the corresponding linker were ordered, for target site 1 (GGCTGTGTTTCCTCCTCCGC) TA**GGC**TGTGTTTCCTCCTCCGC and AAACGCGGAGGAGGAAACACAG and for target site 2 (GGAGCAGCACAACGAGCTGC)

TA**GGA**GCAGCACAACGAGCTGC and AAACGCAGCTCGTTGTGCTGCT (PAM sequences in bold, linker sequences underlined). Oligos were annealed by mixing 10 µl of each oligo pair (100 µM) and 20 µl of $H_2O$ and the following PCR program: 5 min 99 °C, 30 sec 98 °C, cool down to 8 °C with a ramp speed of 0.1 °C/s. The oligo with the annealed linker was ligated into the predigested (BsaI) plasmid pDR274 by using a Ligation Kit (Lucigen). 1 µl of annealed linker, 1 µl (60 ng/µl) digested pDR274 plasmid, 1 µl 10x T4 DNA ligase buffer, 1 µl T4 DNA ligase were mixed and filled up to 10 µl with $H_2O$. This mix was incubated first 2 h at RT followed by 15 min at 70 °C. 5 µl of the ligation mix was used for the following transformation into DH5α competent cells. The transformation mix was plated on kanamycin (30 µg/ml) containing plates and incubated at 37 °C. After 24 h single colonies were picked and grown in liquid cultures for 24 h at 37 °C while shaking. The plasmids were isolated by using the QuickLyse Miniprep Kit (Qiagen) according to the manufacturer's protocol. The plasmids were sequenced using the M13 forward primer and compared to the expected sequences to check for correct ligation of the annealed oligos. To obtain a higher amount if DNA, the plasmids containing the correct sequences were re-amplified in DH5α cells and isolated using the Plasmid Midi Kit (Qiagen). 25 µg of the target site containing plasmid was digested for 5 h at 37 °C using 10 µl NEB cut smart buffer, 5 µl of the restriction enzyme DraI and filled up with $H_2O$ to 100 µl. The heat inactivation was carried out at 65 °C for 20 min. The target sites were purified from an agarose gel by loading the whole digestion mix and using the MinElute Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. For the in vitro transcription the MEGAshortscript T7 Transcription Kit (Ambion) was used and then purified by using the MEGAclear Transcription Clean-Up Kit (Ambion) according to the manufacturer's protocol.

### 5.1.8 Injection of sgRNA and Cas9

Embryos were injected at one-cell stage using a mix of 100 ng/µl sgRNA for the first and second target site, 0.5 µl KCl (2 M) and filled up to 5 µl with $H_2O$. 1 µl of 5 µg/µl Cas9 protein was added. Cas9 protein was purchased from the Protein Expression and Purification Core Facility of EMBL Heidelberg. To allow proper binding of the Cas9 protein with the sgRNA the mixture was incubated at 37 °C for 15 min. The injection was carried out by using a glass needle connected to a microinjector (Pneumatic PicoPump, World Precision Instruments). The mixture was injected by piercing the chorion and yolk and injecting directly inside the one-cell state.

### 5.1.9 Raising of the fish and screening for mutation

Zebrafish were raised in E3 buffer at 26 - 30 °C to adulthood. Screening of mutations was either carried out using 24 hpf embryos or amputated fins for adult fish. 24 hpf embryos were removed from their chorion and transferred to 0.2 ml PCR tubes. Adult fish were anesthetized by incubation in water containing tricaine. The anesthetized fish were transferred to a cutting mat and part of the caudal fin was cut using a scalpel and transferred to a 0.2 ml PCR tube. Adult fish were put back into clean tank water to recover. The DNA of the embryos or fin was extracted using 50 µl QuickExtract DNA Extraction Solution (epicenter) according to the manufacturer's protocol. The region of interest was amplified using a touchdown PCR program and the following reaction mix: 0.75 µl forward primer (gtgaaatatcaggacacat), 0.75 µl reverse primer (gtgaaggtcgtgagtgt), 1 µl genomic DNA, 12.5 µl KAPA HiFi HotStart ReadyMix PCR Kit (KAPABiosystems) and 10 µl $H_2O$. The PCR program was the following:

| | | |
|---|---|---|
| | 95 °C | 4 min |
| 15 cycles: | 98 °C | 40 sec |
| | 65 °C | 45 sec (decrease by 0.5 °C per cycle) |
| | 72 °C | 1 min |
| 25 cycles: | 98 °C | 40 sec |
| | 56.7 °C | 30 sec |
| | 72 °C | 30 sec |
| | 72 °C | 3 min |
| | 4 °C | forever |

The PCR product was purified by either using Diffinity RapidTip 2 (Sigma-Aldrich) or columns of the MinElute Gel extraction kit (Qiagen) with modified conditions. 250 µl of binding buffer (PB) was added, mixed and transferred to a column. The column was spun and washed once with 700 µl wash buffer (PE). The column was spun and the flow through removed. A second spinning step was added to remove everything and the column was air-dried for 60 s. The PCR product was eluted by adding 18 µl of water heated to 65 °C. All sequencing steps were carried out using tgagggcaatacaatatac as a forward primer and gtgaaggtcgtgagtgt as a reverse primer.

## 5.2 Drosophila

### 5.2.1 Fly strains

For all experiment yw flies were used. Flies were grown on standard medium at 22 °C. Embryos were collected from apple juice plates, dechorionated and subjected to subcellular fractionation.

### 5.2.2 Western blot

For SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) 4-20 % precast polyacrylamide-minigels (BIO-RAD) and prestained markers (BIO-RAD) were used. The gels were run at a constant voltage of 300 V and the ampere limit was set to 20 mA per gel. Before transfer onto the nitrocellulose membrane (Sigma-Aldrich) the gel, membrane and the blot filter paper were soaked in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol, 0.04 % SDS, pH 8.3) for 15-30 minutes. For the transfer the semi-dry transfer machine (BIO-RAD) was used according to the manufacturer's protocol. As a loading control the membrane was stained for 10 min using Amido Black and destained with water. The membrane was blocked for 1 h at RT or overnight at 4 °C in 5% milk in PBST. For Western blot analysis antibodies were used in the following concentrations: mouse-anti Dm0 (1:100, DSHB, #ADL101), mouse-anti Ab414 (1:1000, Covance, #MMS-120P), mouse-anti α-Tubulin (1:5000, Sigma). Incubation with the primary antibody was done overnight and with the secondary antibody (1:5000, Jackson ImmunoResearch, # 115-035-044) for 2 h at RT. The Western blot was developed using ECL Plus (PerkinElmer) as a chemiluminiscent substrate, and the signal was imaged in a ChemiDoc Touch Imaging System (BIO-RAD).

### 5.2.3 Subcellular fractionation

Approximately 300 mg wet staged *Drosophila* embryos were dechorionated washed and lysed in Lysis buffer (200 mM Sucrose, 10 mM Tris pH 7.5, 25 mM NaCl, 5 mM $MgCl_2$). Nuclei were isolated by centrifugation at 5000 rpm for 13 min and stripped from attached membranes by centrifugation (45 min, 12000 rpm) through a 1 M Sucrose cushion. Microsomal membranes were isolated by spinning the supernatant from the nuclear precipitate at 40000 rpm for 45 min. The remaining supernatant was used as cytosolic fraction. Subsequently, fractions were lysed or supplemented by addition of Rapigest (Waters) and urea to a final concentration of 0.2 % (v/v) and 4 M, respectively, and sonicated for 3 x 30 s. Samples were stored at -80 °C before being further processed.

### 5.2.4 MS sample preparation and shotgun proteomics

The samples for mass spectrometry were prepared as described in section 5.1.3.

For desalting and cleaning-up of the digested sample, C-18 spin columns (Macro SpinColumn™, Harvard Apparatus) with a capacity of 30-300µg were used. The columns were always spun at 1000 x g for 1 min and 2 min when the sample was loaded. First, the columns were equilibrated with 100 % MetOH and then washed twice with 5 % (v/v) acetonitril (ACN) and 0.1 % (v/v) formic acid (FA). The sample was loaded two times and then the column was washed 4 times with 5 % (v/v) ACN and 0.1 % (v/v) FA. For elution, 50 % (v/v) ACN and 0.1 % (v/v) FA was used. The sample was then with the help of a speed-vac dried and resuspended in 50-100 µl 5 % (v/v) ACN and 0.1 % (v/v) FA. For dissolving, the sample was vortexed, mixed for 5 min at maximum speed in a thermomixer and sonicated for 90 s. The samples were stored at -20°C.

The settings for shotgun MS were the same as previously described in section 5.1.5.

### 5.2.5 Quantitative label-free data analysis

For the quantitative label-free analysis, raw files from the Orbitrap were analyzed using MaxQuant (version 1.2.2.5) (Cox and Mann, 2008). MS/MS spectra were searched against the *Drosophila* Swiss-Prot entries of the Uniprot KB (database release 2013_05, 26951 entries) using the Andromeda search engine (Cox et al., 2011). The search criteria were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues, unless followed by proline); 2 missed cleavages were allowed; carbamidomethylation (C) was set as fixed modification; oxidation (M) and acetylation (protein N-term) were applied as variable modifications, if applicable; mass tolerance of 20 ppm (precursor) and 0.5 Da (fragments). The retention times were matched between runs, using a time window of 2 minutes. The reversed sequences of the target database were used as decoy database. Peptide and protein hits were filtered at a false discovery rate of 1% using a target-decoy strategy (Elias and Gygi, 2007). Additionally, only proteins identified by at least 2 unique peptides were retained. The intensity-based absolute quantification (iBAQ) score, normalized by molecular weight, of the proteinGroups.txt output of MaxQuant was used for further analysis. All comparative analyses were performed using R version 3.0.1. (R Core Team, 2012). Only proteins identified in at least 2 replicates were considered when comparing protein abundances between different fractions. To reduce technical variation, the data was quantile-normalized using the preprocessCore library (Gentleman et al., 2004). Protein differential expression was evaluated using the Limma package

(Smyth et al., 2005). Differences in protein abundances were statistically determined using the Student's t-test moderated by the empirical Bayes method. Significant regulated proteins were defined by a cut-off of log2 fold change ≤ -1 or ≥ 1 and p-value ≤ 0.01.

## 5.3 Human cells – NTR

### 5.3.1 Human cell culture

Flp-In 293 T-REx (Thermo Fisher Scientific) HEK (human embryonic kidney) cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) high glucose 5 g/l (Sigma-Aldrich) supplemented with 10% heat inactivated fetal bovine serum (FBS). The parental cell line was grown with the addition of 100 µg/ml Zeocin (Invitrogen) and 15 µg/ml blasticidin (Thermo Fisher Scientific). After generation of stable cell lines Zeocin™ was replaced by 100 µg/ml hygromycin. Stable cell lines were seeded at a density of 1.6e04 and allowed to attach to the culture dish for 24 h and then induced by adding 1 µg/ml tetracyclin (Sigma-Aldrich) in ethanol directly to the medium. After 24 additional hours 50 µM Biotin (Sigma-Aldrich) of a 10 mM stock solution in water was added. The cells were washed twice with PBS with magnesium and calcium at RT for collection, while still attached to the culture dish, to exclude floating cells and the excess of biotin for all further experiments. To keep the cells intact, trypsin (Sigma-Aldrich) was used to detach the cells and inactivated after 7-10 minutes by addition of complete medium.

### 5.3.2 NTR-BirA Plasmid generation

For the generations of plasmids for genomic integration with the Flp-In T-Rex the Gateway Technology (Invitrogen) was used. Expression clones were generated by combining a destination vector (pcDNA5-pDEST-BirA-FLAG-N-term or pcDNA5-pDEST-BirA-FLAG-C-term) with an entry clone. An overview of all vectors used, can be found in Table 4. Some NTRs were purchased directly as entry clones, while I got other as cDNA without the necessary att sites for Gateway cloning. In the latter case, the required attB sites to generate an entry clone were added by PCR using the Phusion High-Fidelity DNA polymerase with primers listed in Table 5 according to recommendations of the manufacturer. The PCR product was extracted from an agarose gel using the QIAEX II Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. Entry Clones were generated using the donor vector pDONR 221 (Thermo Fisher Scientific) and the attB-site PCR product by performing the BP recombination reaction. The expression clones were created by the LR recombination reaction of entry clones and the destination vectors. Both reactions and the following transformation were

done according to the manufacturer's protocol. Two different BirA destination vectors were used to generate the fusion to the protein of interest either at the C- or N-terminus. The expression clone plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen)

**Table 4: Overview of cDNA or destination and entry vectors used to generate fusion proteins**

| | NTRs and other targets | Vector/ cDNA type | Antibiotic resistance | GenBank accession for cDNA | Source |
|---|---|---|---|---|---|
| **1** | BirA | Destination vector | Ampicillin | - | Gingras Lab, Lunenfeld-Tanenbaum Research Institute, Toronto |
| **2** | NLS-NES-Dendra2 | cDNA | Kanamycin | - | Ellenberg Lab, EMBL, Heidelberg |
| **3** | IMA1 | Entry clone | Spectinomycin | BC005978 | Human ORFeome |
| **4** | IMA5 | Entry clone | Spectinomycin | BC002374 | Human ORFeome |
| **5** | IMA6 | Entry clone | Spectinomycin | BC047409 | Human ORFeome |
| **6** | KPNB1 | cDNA | Chloramphenicol | BC003572 | Human ORFeome |
| **7** | IPO4 | cDNA | Kanamycin | BC136759 | Human ORFeome |
| **8** | IPO5 | Entry clone | Spectinomycin | BC001497 | Human ORFeome |
| **9** | IPO11 | Entry clone | Spectinomycin | BC033776 | Human ORFeome |
| **10** | IPO13 | Entry clone | Spectinomycin | BC008194 | Human ORFeome |
| **11** | TNPO1 | cDNA | Kanamycin | BC040340.1 | Human ORFeome |
| **12** | TNPO2 | Entry clone | Spectinomycin | BC072420 | Human ORFeome |
| **13** | XPO1 | Entry clone | Spectinomycin | BC032847 | Human ORFeome |
| **14** | XPO2 | cDNA | Kanamycin | BC109313 | Human ORFeome |
| **15** | XPO7 | Entry clone | Spectinomycin | BC030785 | Human ORFeome |
| **16** | RAN | Entry clone | Spectinomycin | BC014901 | Human ORFeome |
| **17** | NTF2 | Entry clone | Spectinomycin | BC002348.2 | Human ORFeome |
| **18** | NXT1 | Entry clone | Spectinomycin | BC003410 | Human ORFeome |
| **19** | NXT2 | Entry clone | Spectinomycin | BC014888.1 | Human ORFeome |

**Table 5: Primers to generate attB-PCR products for entry clones**

| Target protein | attB Primer | attB Primer sequence (5' to 3') |
|---|---|---|
| **NLS-NES-Dendra2** | attB1 | ggggacaagtttgtacaaaaaagcaggctatgttatgggggatccaaagagacc |
| **NLS-NES-Dendra2** | attB2 | ggggaccactttgtacaagaaagctgggtcagatctaccaccacccttgtac |
| **NLS-NES-Dendra2** | attB2 (with stop codon) | ggggaccactttgtacaagaaagctgggtcttaagatctaccaccacccttg |
| **TNPO1** | attB1 | ggggacaagtttgtacaaaaaagcaggctatgccatggagtatgagtggaaacc |
| **TNPO1** | attB2 | ggggaccactttgtacaagaaagctgggtcaacaccataaaaagctgcaaga |
| **TNPO1** | attB2 (with stop codon) | ggggaccactttgtacaagaaagctgggtcttaaacaccataaaaagctgca |
| **KPNB1** | attB1 | ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggagctgatcaccattct |
| **KPNB1** | attB2 | ggggaccactttgtacaagaaagctgggtcagcttggttcttcagtttcc |

| | | |
|---|---|---|
| **IPO4** | attB1 | ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggagtcagccgggctaga |
| **IPO4** | attB2 | ggggaccactttgtacaagaaagctgggtcggagaggcccagtacagcct |
| **XPO2** | attB1 | ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatggaactcagcgatgcaaatctg |
| **XPO2** | attB2 | ggggaccactttgtacaagaaagctgggtcaagcagtgtcacactggctgcc |

### 5.3.3 Site directed mutagenesis of cDNA

Several purchased entry clones or cDNAs had sequence disagreements with current sequences available at ENSEMBL. Sequence conflicts causing amino acid variances were changed using the QuickChange II Site-Directed Mutagenesis Kit (Agilent) according to the manufactures protocol. Primers designed by the web-based QuikChange Primer Design Program are listed in Table 6.

**Table 6: Primers for site directed mutagenesis**

| Target Protein | Expected mutation | Sequence primer forward (5' to 3') | Sequence primer reverse (5' to 3') |
|---|---|---|---|
| **IMA1** | T470C | cacctccatctaccacagccttggtttgttctgat | atcagaacaaaccaaggctgtggtagatggaggtg |
| **IMA1** | C1152T | gtcccattccttgtcagtgttctctctaaggcaga | tctgccttagagagaacactgacaaggaatgggac |
| **IMA1** | C1359A | tttctcagtttcacctagtttctcagcagcctgaaag | ctttcaggctgctgagaaactaggtgaaactgagaaa |
| **IMA5** | A218G | ggtgccatctccatgttactaatctgagcctcatgaaa | tttcatgaggctcagattagtaacatggagatggcacc |
| **IPO13** | T411C | catggccaagcgtcaggcatcatgctgagagc | gctctcagcatgatgcctgacgcttggccatg |
| **IPO13** | T2449G | agcacactggaacacagctttgacatccaatcgttc | gaacgattggatgtcaaagctgtgttccagtgtgct |
| **IPO13** | C2888A | tacaagaaagttgggcagtagtcagctgtgtaatctg | cagattacacagctgactactgcccaactttcttgta |
| **IPO11** | T783G | cagtctatctctacacacattatctgtacctatacttctactgca | tgcagtagaagtataggtacagataatgtgtgtagagatagactg |
| **XPO1** | C1A and A2T | gctaacattgtcataattgctggcatccaacttttttgtacaaagttggc | gccaactttgtacaaaaaagttggatgccagcaattatgacaatgttagc |
| **XPO7** | A2504G | ccttcagcatggagaagcagatggagatgcccttg | caagggcatctccatctgcttctccatgctgaagg |
| **XPO7** | C1194G | gcatgtggggctctgtggctttgacatacg | cgtatgtcaaagccacagagccccacatgc |

### 5.3.4 Stable cell line generation

On the first day 200000 cells were seeded per well of a 6 well plate without any antibiotics in the cell culture medium. On the second day the cells were transfected using 100 µl Opti-MEM (Gibco) with 3 µl of X-tremeGENE 9 DNA Transfection Reagent (Sigma-Aldrich). To this mix 1 µg of DNA, consisting of 100 ng of plasmid of interest and 900 ng of pOG44, was added and mixed well. After 15 minutes of incubation at RT this mix was added dropwise under constant shaking. The following day, each well was, after washing off detached cells, trypsinized onto a 150 mm plate. On the fourth day, hygromycin and blasticidin were added for selection. The following weeks, the medium was replaced twice per week with addition of new antibiotics, and the cells were trypsinized onto the same plate as soon as bigger colonies were formed. As soon as the plate was 80% confluent the cells

were prepared for long time storage at -180 °C in liquid nitrogen. The cells were washed, trypsinized, spun at 500 g and each quarter of a 150 mm plate resuspended in 500 µl FBS with 10 % DMSO. To allow a controlled freezing rate the cryogenic vials were placed inside a cell freezing container (Biocision) for 24 h at -80 °C, and then transferred to a liquid nitrogen storage tank.

### 5.3.5 BioID Pull-down experiments

The cell were induced and treated as explained in section 5.3.1. For each experiment 4e7 cells were collected and spun. The supernatant was removed and the cell pellets snap frozen in liquid nitrogen and stored at -20 °C till further use.

The cell pellets were thawed on ice and resuspended in 9.5 ml lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 1mM EGTA, 1% Triton, 1 mg/ml Aprotinin, 0.5 mg/ml Leupeptin, 250 U Turbonuclease (Accelagen), 1 mM PMFS, 0.1 % SDS). The cells were lysed by incubation for 1 h at 4 °C rotating at 15 rpm. The lysis was equally splitted into 5 2 ml Eppendorf tubes and sonicated 5 times 30 sec at 4 °C. The lysis was spun 30 min at 4 °C at 17000g in a tabletop centrifuge and the supernatants merged into a 15 ml falcon tube. All the following steps involving streptavidin sepharose beads were done using low-affinity binding tips (Biozym). 80 µl of slurry Streptavidin Sepharose High Performance beads (GE Healthcare Life Sciences) were equilibrated in lysis buffer and then added to the lysis and incubated for 3 h at 4 °C rotating at 15 rpm. 9 ml of the supernatant was removed after spinning at 2000 g for 2 min at 4 °C. The remaining supernatant with the beads was transferred to a Spin Column (Pierce, ThermoFisher). All following washes and elution steps were carried out by using a pipette teat to remove the liquid. The beads were washed twice using 700 µl lysis buffer and 4 times with 700 µl wash buffer (50 mM ammonium bicarbonate (AmBic), pH 8.3). After closing the column with a plug at the bottom the beads were transferred to a 2 ml tube by using a cut pipet tip and 3 times 300 µl wash buffer. The beads were spun down at 2000 g for 5 min at 4 °C and 700 µl of the supernatant was removed. 1 µg of Trypsin (Mass Spectrometry Grade, Promega) was added and incubated at 37 °C for 16 h shaking at 500 rpm. Then 0.5 µg of Trypsin was added and the on-bead digest continued for additional 2 h. The beads were transferred to a new Spin Column and the digested peptides were eluted with two times 150 µl 50 mM AmBic. The on-bead elutions were merged. To remove the biotinylated peptides still bound to the beads, 150 µl of 80 % ACN and 20 % TFA was added, briefly mixed and eluted; this step was done twice. The ACN/TFA elutions were merged. The samples were dried using a speed-vac. All following pipetting steps were done using low-affinity binding tips with a filter (Starlab). The elutions were resuspended

in 200 µl buffer A. The desalting and clean-up of the samples was carried out as described in section 5.2.4, except for the use of Micro SpinColumns.

### 5.3.6 Western blot

For general procedure refer to section 5.2.2.

For Western blot analysis antibodies were used in the following concentrations: anti-Biotin, HRP-linked (1:2000 in 2 % BSA in PBST, Cell Signaling Technology, #7075S), mouse-anti FLAG (1:1000 in 5 % milk in PBST, Sigma-Aldrich, #F3165). Incubation with the primary antibody was done for 2 h and with the secondary antibody (1:5000, in 5 % milk in PBST, Jackson ImmunoResearch, #115-035-044) for 2 h at RT.

### 5.3.7 Protein identification by mass spectrometry

The settings for shotgun MS were the same as previously described in section 5.1.5.

The elutions of biotinylated peptides were measured using the same settings but with a stepwise gradient lasting 90 minutes.

### 5.3.8 Quantitative label-free data analysis

For the quantitative label-free analysis, raw files from the Orbitrap were analyzed using MaxQuant (version 1.5.3.28) (Cox and Mann, 2008). MS/MS spectra were searched against the *Human* Swiss-Prot entries of the Uniprot KB (database release 2016_09, 19594 entries) using the Andromeda search engine (Cox et al., 2011). The protein sequences of BirA and Streptavidin were added to the database. The search criteria were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues, unless followed by proline); 3 missed cleavages were allowed; oxidation (M), acetylation (protein N-term) and biotinylation (K) were applied as variable modifications, if applicable; mass tolerance of 20 ppm (precursor) and 0.5 Da (fragments). The retention times were matched between runs, using a time window of 3 minutes. The reversed sequences of the target database were used as decoy database. Peptide and protein hits were filtered at a false discovery rate of 1% using a target-decoy strategy (Elias and Gygi, 2007). Additionally, only proteins identified by at least 2 unique peptides were retained. The Intensity per protein of the proteinGroups.txt output of MaxQuant was used for further analysis. All comparative analyses were performed using R version 3.2.2. (R Core Team, 2012). The R packages MSnbase (Gatto and Lilley, 2012) for processing

proteomics data and the included package imputeLCMD for imputing missing values based on the definitions for values MAR and MNAR in Table 3. Values MNAR were imputed based on the method "MinDet" by replacing values with minimal values observed in the sample. Values MAR were imputed based on the method "knn" (k-nearest neighbors). The data was quantile normalized to reduce technical variations. Protein differential expression was evaluated using the Limma package (Smyth et al., 2005). Differences in protein abundances were statistically determined using the Student's t-test (one-sided) moderated by the empirical Bayes method.

NIP and background proteome

The pairwise comparisons of the AP of NTRs to the control data set were used to define the NIP and background proteome. The comparison was done one-sided and separately for the NIP and the background proteome. The Sime's adjusted p-vales was calculated for each protein using the R cherry package (Goeman and Solari, 2011). The minimum p-value per protein, which defines if this protein is significant in the NIP or background proteome, was adjusted using the method of Benjamini and Hochberg.

ES calculations for specificity of NTRs

The enrichment score (ES) per protein in each experiment was calculated by multiplying all enriched p-values of the selected protein obtained in the pairwise comparison. The new average fold change (FC) was calculated by using the FC of all these proteins calculated in the pairwise comparison. The Fisher transformation was used to define a FDR per experiment. The fisher p-values were adjusted using the method of Benjamini and Hochberg.

Network analysis for specificity of NTRs

In order to identify clusters of proteins displaying specific interaction with NTRs, we applied a network smoothing approach. First, we mapped all the proteins quantified in our experiments to the human STRING protein-protein interaction network (v10, combined score > 0.7, 15478 nodes). The network was then converted into an adjacency matrix and normalized using laplacian transformation. Specificity scores where propagated to adjacent nodes by network smoothing (alpha = 0.5, 30 iterations) followed by topology bias correction. For each NTR sample, proteins were ranked according to their specificity score following network smoothing, and the top 2% proteins for each sample were used for the identification of highly interconnected sub-networks using the Cytoscape (Cline et al., 2007) App MCODE (Bader and Hogue, 2003).

### 5.3.9 Staining of NTR-BirA cell lines

Cells are grown directly on glass slides, previously coated with poly-Lysine, in PBS for 4 h, to enable better attachment for the cells. 100000 cells were seeded per well of a 6 well plate, each containing two glass slides. The cells were induced and treated with biotin as explained in section 5.3.1. All following steps were performed after transferring the glass slides in 24 well plates and with 2 washing steps with PBS containing magnesium and calcium in between each incubation step at RT. First the cells were fixed with 2 % PFA in PBS for 15 min and then permeabilized with 0.4 % Triton in PBS for additional 15 min. Blocking was performed using 2% BSA and 2% FBS in PBS for 1 h. To visualize the nuclear envelope, cells were incubated with mAb414 (1:2500, Covance, #MMS-120P) in blocking buffer for 1 h. As a secondary antibody, an anti-mouse conjugated to Alexa Fluor 488 (1:1000, Life Technologies, #A21204) was used also for 1 h. Starting from here all following step were done with minimum light exposure. DNA was visualized using Hoechst stain (1:10000, Sigma-Aldrich) in PBS for 10 minutes. The last staining step had the goal to visualize all biotinylated proteins. Streptavidin covalently bound to Alexa Fluor 647 (1:1000, Thermo Fisher Scientific, #S21374) in 0.1 % BSA in PBS was used to incubate the cells for 10 minutes. To preserve the stained cells all glass slides were mounted upside down on a microscope slide using one drop of mounting medium (Thermo Fisher Scientific), dried over night at RT and afterwards stored at -20 °C.

# 6 Appendices

## 6.1 Supplementary Figures

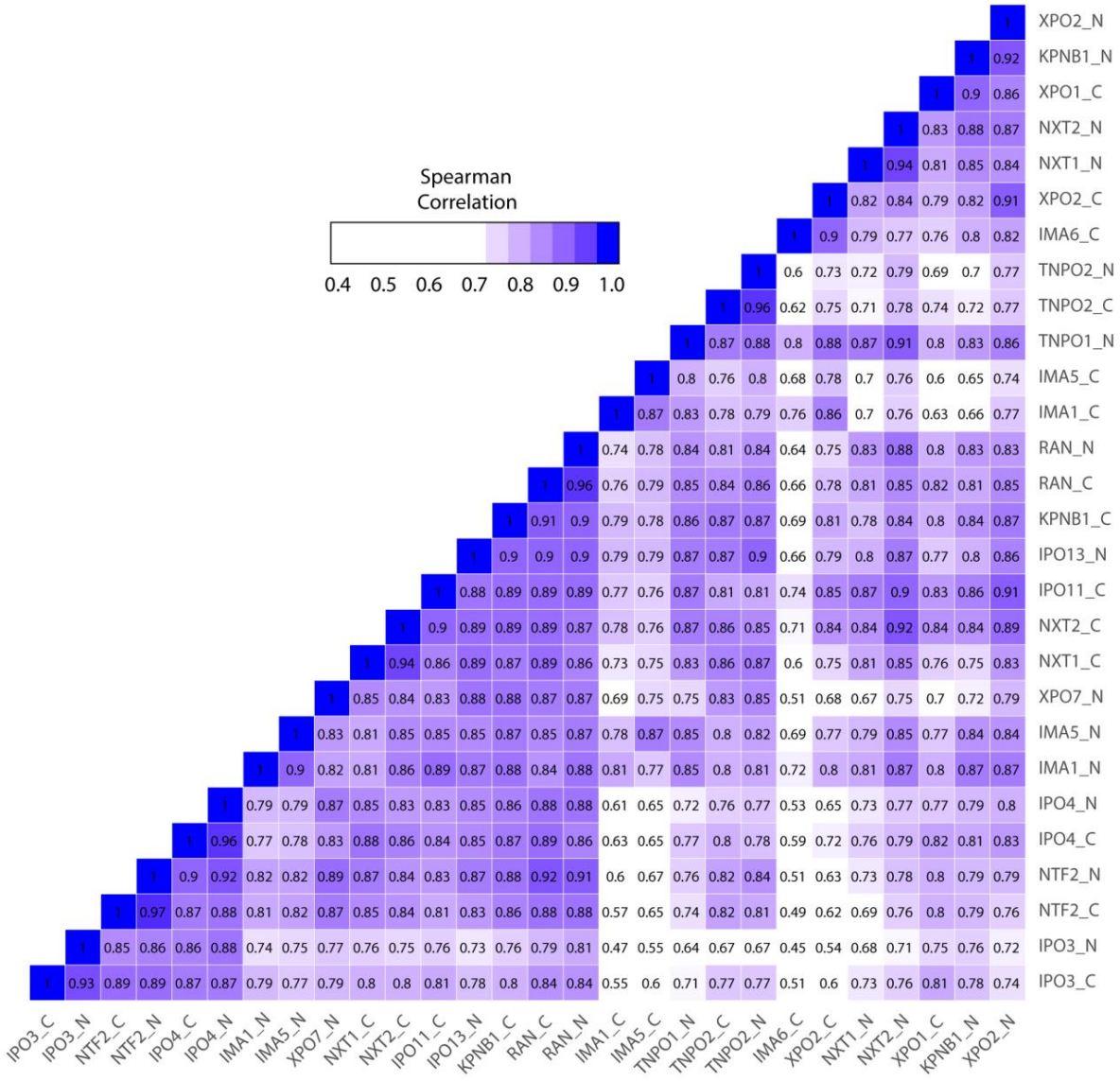


Figure 39: Spearman Correlation of median quantile normalizes iBAQ scores of all NTR-BirA APs.

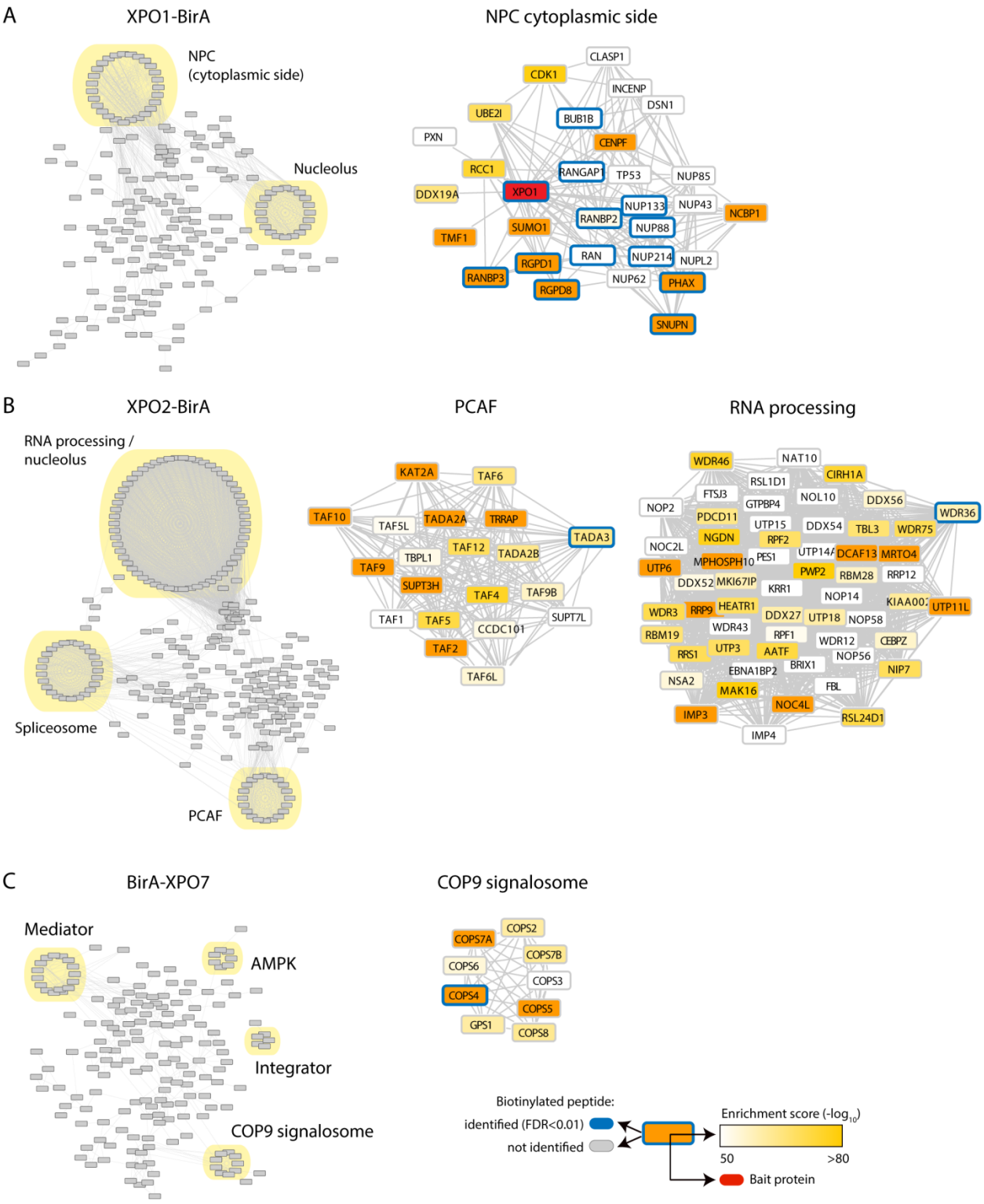


**Figure 40: Network analysis of the top 2 % enriched proteins for XPO1-BirA (A), XPO2-BirA (B) and BirA-XPO7 (C). Selected sub networks with the enrichment score and detected biotinylated peptides are shown. NPC (Nuclear pore complex), PCAF (p300/CBP-associated factor), AMPK (AMP-activated protein kinase), COP9 (Constitutive photomorphogenesis 9).**

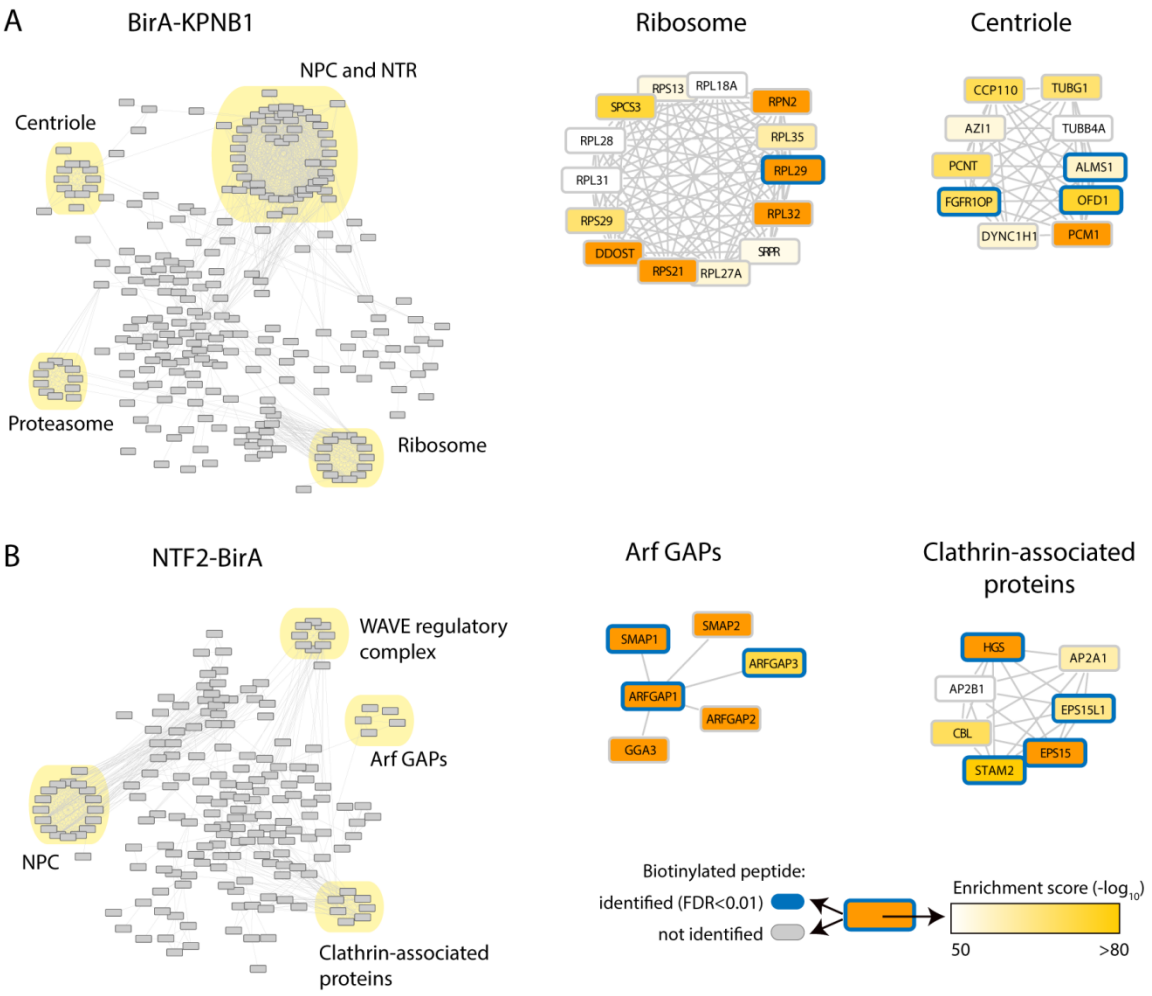


**Figure 41: Network analysis of the top 2 % enriched proteins for BirA-KPNB1 (A) and NTF2-BirA (B). Selected sub networks with the enrichment score and detected biotinylated peptides are shown. NPC (Nuclear pore complex), NTR (Nuclear transport receptors), Arf GAP (ADP-ribosylation factor GTPase-activating proteins), WAVE (WASP-family verprolin homologous protein).**

## 6.2 Supplementary Tables

**Table 7: Overview of NTRs used in this study of the nucleocytoplasmic transport system.**

| | Name used in this study | Protein name | Other names | Directionality of transport | NTR Class | Uniprot accession number |
|---|---|---|---|---|---|---|
| **1** | IMA1 | Importin subunit alpha-1 | KPNA2, RCH1, SRP1 | Import | Importin alpha I | P52292 |
| **2** | IMA5 | Importin subunit alpha-2 | KPNA1, RCH2 | Import | Importin alpha III | P52294 |
| **3** | IMA6 | Importin subunit alpha-6 | KPNA5 | Import | Importin alpha III | O15131 |
| **4** | KPNB1 | Importin subunit beta-1 | IMB1, NTF97 | Import | Importin beta | Q14974 |
| **5** | IPO4 | Importin-4 | IMP4B, RANBP4 | Import | Importin beta | Q8TEX9 |
| **6** | IPO5 | Importin-5 | KPNB3, RANBP5 | Import | Importin beta | O00419 |
| **7** | IPO11 | Importin-11 | RANBP11 | Import | Importin beta | Q9UI26 |
| **8** | IPO13 | Importin-13 | RANBP13, KIAA0724 | Import/Export | Importin beta | O94829 |
| **9** | TNPO1 | Transportin-1 | KPNB2, MIP1, TRN | Import | Importin beta | Q92973 |
| **10** | TNPO2 | Transportin-2 | KPNB2B, TRN-2A | Import | Importin beta | O14787 |
| **11** | XPO1 | Exportin-1 | CRM1 | Export | Importin beta | O14980 |
| **12** | XPO2 | Exportin-2 | CSE1L, CAS | Export | Importin beta | P55060 |
| **13** | XPO7 | Exportin-7 | KIAA0745, RANBP16 | Export | Importin beta | Q9UIA9 |
| **14** | RAN | GTP-binding nuclear protein RAN | - | - | - | P62826 |
| **15** | NTF2 | Nuclear transport factor 2 | NUTF2 | Import | - | P61970 |
| **16** | NXT1 | NTF2-related export protein 1 | p15 | Export | - | Q9UKK6 |
| **17** | NXT2 | NTF2-related export protein 2 | p15-2 | Export | - | Q9NPJ8 |

## 6.3 List of figure

## 6.4 List of tables



## 6.5 Abbreviations

| | | | |
|---|---|---|---|
| 3D | Three-dimensional | Da | Dalton |
| Å | Angstrom | DNA | Deoxyribonucleic acid |
| ACN | Acetonitrile | dsNLS | Dimer specific NLS |
| AL | Annulate lamellae | dpf | Days post fertilization |
| AP | Affinity purification | DMEM | Dulbecco's Modified Eagle's Medium |
| ARM | Armadilla | DMSO | Dimethyl sulfoxide |
| BirA | Biotin ligase | DTT | Dithiothreitol |
| BSA | Bovine serum albumin | EM | Electron microscopy |
| ChIP | Chromatin immunoprecipitation | ER | Endoplasmic reticulum |
| cNLS | Classical nuclear localization signal | ES | Enrichment score |
| CRISPR | Clustered regularly interspaced short palindromic repeats | FA | Formic acid |

| | |
|---|---|
| FACS | Fluorescence-activated cell sorting |
| FBS | Fetal bovine serum |
| FC | Fold change |
| FG | Phenylalanine-Glycine |
| GDP | Guanosine diphosphate |
| GFP | Green fluorescent protein |
| GO | Gene ontology |
| gRNA | Guide ribonucleic acid |
| GTP | Guanosine triphosphate |
| HBV | Hepatits B virus |
| HDACi | Histone deacetylase inhibitors |
| HEK | Human embryonic kidney |
| HeLa | Henrietta Lacks |
| hpf | Hours post fertilization |
| HPLC | High performance liquid chromatography |
| IAA | Iodoacetamide |
| IBB | Importin β binding |
| kDa | Kilodalton |
| LB | Lysogeny broth |
| LC-MS | Liquid chromatography mass spectrometry |
| MAT | Missing at random |
| MDa | Megadalton |
| mESC | Mouse embryonic stem cells |
| min | minute |
| ml | Milliliter |
| mM | Millimolar |
| MNAT | Missing not at random |
| mRNA | Messenger RNA |
| MS | Mass Spectrometry |
| MW | Molecular weight |
| NE | Nuclear envelope |
| NES | Nuclear export sequence |
| NHEJ | Non-homologous end joining |
| NIP | NTR interacting proteome |
| NLS | Nuclear localization sequence |
| nM | Nanomolar |
| NPC | Nuclear pore complex |
| NTR | Nuclear transport receptor |
| Nup | Nucleoporin |
| PAM | Protospacer adjacent motif |
| PBS | Phosphate-buffered saline |
| PCR | Polymerase chain reaction |
| PFA | Paraformaldehyde |
| PTM | Posttranslational modifications |
| RanGAP | Ran GTPase-activating protein |
| RanGEF | Ran guanine nucleotide exchange factor |
| RNA | Ribonucleic acid |
| RNAi | Ribonucleic acid interference |
| RNC | Relative nuclear concentration |

| | |
|---|---|
| RT | Room temperature |
| SDS-PAGE | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| sgRNA | Single guide ribonucleic acid |
| SILAC | Stable isotope labeling by amino acids in cell culture |
| TFA | Trifluoroacetic acid |
| µl | Microliter |
| µm | Micrometer |
| WB | Western Blot |

## 6.6 Publications


1. Hampoelz B, **Mackmull MT**, Machado P, Ronchi P, Bui KH, Schieber N, Santarella-Mellwig R, Necakov A, Andres-Pons A, Philippe JM, Lecuit T, Yannick Schwab, and Beck M. (2016) **Pre-assembled Nuclear Pores Insert into the Nuclear Envelope during Early Development.** *Cell* 166, 1–15, July 28
2. von Appen A, Kosinski J, Sparks L, Ori A, DiGuilio AL, Vollmer B, **Mackmull MT**, Banterle N, Parca L, Kastritis P, Buczak K, Mosalaganti S, Hagen W, Andres-Pons A, Lemke EA, Bork P, Antonin W, Glavy JS, Bui KH, Beck M. (2015) **In situ structural analysis of the human nuclear pore complex.** *Nature* 526(7571):140-143
3. **Mackmull MT**, Iskar M, Parca L, Singer S, Bork P, Ori A, Beck M. (2015) **Histone Deacetylase Inhibitors (HDACi) Cause the Selective Depletion of Bromodomain Containing Proteins (BCPs).** *Mol. Cell Proteomics* 2015, 14(5):1350-1360
4. Bui KH, von Appen A, DiGuilio AL, Ori A, Sparks L, **Mackmull MT**, Bock T, Hagen W, Andrés-Pons A, Glavy JS, Beck M. (2013) **Integrated structural analysis of the human nuclear pore complex scaffold.** *Cell* 155(6):1233-1243

# 7 Acknowledgements

First of all I would like to express my sincere gratitude to Dr. Martin Beck for the opportunity to work on an exciting and challenging project and write my doctoral thesis in his great laboratory. Thank you for the confidence in me and the constant motivation, patience, freedom and support throughout the time.

I would like to thank the members of my thesis advisory committee Dr. Darren Gilmour, Dr. Anne-Claude Gavin and Prof. Dr. Matthias Mayer for helpful advice, suggestions and intensive discussions. Thank you also for the support outside of the TAC meetings.

Thanks Alessandro for now almost five years of working together. I have learned a lot from you and your knowledge seems infinite. I am grateful for your patience to explain me things over and over again and for your continuous motivation and support.

Thanks Amparo for sharing your excellent knowledge in cell and molecular biology with me. Thank you for having everything in the laboratory under control and defending me against the constant, lurking disorder from all sides.

Thanks Bernhard for showing me how to apply the 3 seconds rule in the laboratory and knowing every story behind these crazy protein names of Drodophila. Show me your best mountain tour as soon as you are officially authorized to do so. Good luck!

Thanks Kasia that you took up a position with me on the better side of the lab and we stood together against the fuzzy picture fraction.

Thanks Shyamal and one day we can perhaps make Aladin fly high. Otherwise, we eat Aladin's Wunderteller until we are sick.

Thanks Jan for the in the truest sense of the word sweaty hours with many victories and for the help of understanding the mysterious floating Nups.

I would like to thank all former lab members starting with Lenore (thanks for your unconventional way), Alex (thanks for the many very special jokes at a questionable level), Huy, Thomas, Amanda (thanks for a lot of fun and that I could get to know New York from another side), Jakob, Borja (thanks for many test experiments for the BirA project), Scott and Stephan, and all the current lab members Panos, Matteo, Andre, Christian, Shotaro and Jacob. Thank you very much. You are responsible for the fact that I almost

always like to go to the lab. Thank you not only for the support and many great advices on scientific problems, but also about the insane important questions about the number of lives of a cat and many other things.

Special thanks goes to Joanna and all former and current members of the Proteomics Core Facility for their support in mass spectrometry and maintenance of all the machines.

I would like to thank Darren Gilmour for the opportunity to work in his fish laboratory. Thanks Andreas, Sarfi, Sevi, Sabine and all other lab members for helping me out in the unknown field of the zebrafish.

Apart from all people who enrich my academic life, I would like to thank the people who have taken me there. First of all, I would like to thank my parents Andrea and André for the support in my whole life and the free decision of my future at any time. Then I would like to thank my grandparents Anton, Maria and Erna. Finally, I would like to thank Daniel for his patience and support and the attempt to understand a s scientist.

# 8 References

Allgrove, J., G. S. Clayden, D. B. Grant, and J. C. Macaulay, 1978, Familial glucocorticoid deficiency with achalasia of the cardia and deficient tear production: Lancet, v. 1, p. 1284-6.

Andrade, M. A., and P. Bork, 1995, HEAT repeats in the Huntington's disease protein: Nat Genet, v. 11, p. 115-6.

Andrade, M. A., C. Petosa, S. I. O'Donoghue, C. W. Müller, and P. Bork, 2001, Comparison of ARM and HEAT protein repeats: J Mol Biol, v. 309, p. 1-18.

Baake, M., M. Bäuerle, D. Doenecke, and W. Albig, 2001, Core histones and linker histones are imported into the nucleus by different pathways: Eur J Cell Biol, v. 80, p. 669-77.

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

Bannister, A. J., E. A. Miska, D. Görlich, and T. Kouzarides, 2000, Acetylation of importin-alpha nuclear import factors by CBP/p300: Curr Biol, v. 10, p. 467-70.

Boersema, P. J., R. Raijmakers, S. Lemeer, S. Mohammed, and A. J. Heck, 2009, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics: Nat Protoc, v. 4, p. 484-94.

Brown, C. R., C. J. Kennedy, V. A. Delmar, D. J. Forbes, and P. A. Silver, 2008, Global histone acetylation induces functional genomic reorganization at mammalian nuclear pore complexes: Genes Dev, v. 22, p. 627-39.

Capelson, M., and M. W. Hetzer, 2009, The role of nuclear pores in gene regulation, development and disease: EMBO Rep, v. 10, p. 697-705.

Carvalhal, S., S. A. Ribeiro, M. Arocena, T. Kasciukovic, A. Temme, K. Koehler, A. Huebner, and E. R. Griffis, 2015, The nucleoporin ALADIN regulates Aurora A localization to ensure robust mitotic spindle formation: Mol Biol Cell, v. 26, p. 3424-38.

Cañas, B., D. López-Ferrer, A. Ramos-Fernández, E. Camafeita, and E. Calvo, 2006, Mass spectrometry technologies for proteomics: Brief Funct Genomic Proteomic, v. 4, p. 295-320.

Chapman-Smith, A., and J. E. Cronan, 1999, The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity: Trends Biochem Sci, v. 24, p. 359-63.

Chen, B. B., and R. K. Mallampalli, 2009, Masking of a nuclear signal motif by monoubiquitination leads to mislocalization and degradation of the regulatory enzyme cytidylyltransferase: Mol Cell Biol, v. 29, p. 3062-75.

Choi-Rhee, E., H. Schulman, and J. E. Cronan, 2004, Promiscuous protein biotinylation by Escherichia coli biotin protein ligase: Protein Sci, v. 13, p. 3043-50.

Chook, Y. M., A. Jung, M. K. Rosen, and G. Blobel, 2002, Uncoupling Kapbeta2 substrate dissociation and ran binding: Biochemistry, v. 41, p. 6955-66.

Chook, Y. M., and K. E. Süel, 2011, Nuclear import by karyopherin-βs: recognition and inhibition: Biochim Biophys Acta, v. 1813, p. 1593-606.

Christie, M., C. W. Chang, G. Róna, K. M. Smith, A. G. Stewart, A. A. Takeda, M. R. Fontes, M. Stewart, B. G. Vértessy, J. K. Forwood, and B. Kobe, 2016, Structural Biology and Regulation of Protein Import into the Nucleus: J Mol Biol, v. 428, p. 2060-90.

Clark, A. J., and A. Weber, 1998, Adrenocorticotropin insensitivity syndromes: Endocr Rev, v. 19, p. 828-43.

Cline, M. S., M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P. L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, and G. D. Bader, 2007,

Integration of biological networks and gene expression data using Cytoscape: Nat Protoc, v. 2, p. 2366-82.
Clouse, K. N., M. J. Luo, Z. Zhou, and R. Reed, 2001, A Ran-independent pathway for export of spliced mRNA: Nat Cell Biol, v. 3, p. 97-9.
Cook, A., F. Bono, M. Jinek, and E. Conti, 2007, Structural biology of nucleocytoplasmic transport: Annu Rev Biochem, v. 76, p. 647-71.
Cordes, V. C., S. Reidenbach, and W. W. Franke, 1996, Cytoplasmic annulate lamellae in cultured cells: composition, distribution, and mitotic behavior: Cell Tissue Res, v. 284, p. 177-91.
Couzens, A. L., J. D. Knight, M. J. Kean, G. Teo, A. Weiss, W. H. Dunham, Z. Y. Lin, R. D. Bagshaw, F. Sicheri, T. Pawson, J. L. Wrana, H. Choi, and A. C. Gingras, 2013, Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions: Sci Signal, v. 6, p. rs15.
Cox, J., and M. Mann, 2008, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification: Nat Biotechnol, v. 26, p. 1367-72.
Cox, J., N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann, 2011, Andromeda: a peptide search engine integrated into the MaxQuant environment: J Proteome Res, v. 10, p. 1794-805.
Coyaud, E., M. Mis, E. M. Laurent, W. H. Dunham, A. L. Couzens, M. Robitaille, A. C. Gingras, S. Angers, and B. Raught, 2015, BioID-based Identification of Skp Cullin F-box (SCF)β-TrCP1/2 E3 Ligase Substrates: Mol Cell Proteomics, v. 14, p. 1781-95.
Cronshaw, J. M., A. N. Krutchinsky, W. Zhang, B. T. Chait, and M. J. Matunis, 2002, Proteomic analysis of the mammalian nuclear pore complex: J Cell Biol, v. 158, p. 915-27.
Cronshaw, J. M., and M. J. Matunis, 2003, The nuclear pore complex protein ALADIN is mislocalized in triple A syndrome: Proc Natl Acad Sci U S A, v. 100, p. 5823-7.
D'Angelo, M. A., D. J. Anderson, E. Richard, and M. W. Hetzer, 2006, Nuclear pores form de novo from both sides of the nuclear envelope: Science, v. 312, p. 440-3.
D'Angelo, M. A., J. S. Gomez-Cavazos, A. Mei, D. H. Lackner, and M. W. Hetzer, 2012, A change in nuclear pore complex composition regulates cell differentiation: Dev Cell, v. 22, p. 446-58.
D'Angelo, M. A., M. Raices, S. H. Panowski, and M. W. Hetzer, 2009, Age-dependent deterioration of nuclear pore complexes causes a loss of nuclear integrity in postmitotic cells: Cell, v. 136, p. 284-95.
Dickmanns, A., R. H. Kehlenbach, and B. Fahrenkrog, 2015, Nuclear Pore Complexes and Nucleocytoplasmic Transport: From Structure to Function to Disease: Int Rev Cell Mol Biol, v. 320, p. 171-233.
Dingwall, C., J. Robbins, S. M. Dilworth, B. Roberts, and W. D. Richardson, 1988, The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen: J Cell Biol, v. 107, p. 841-9.
Domon, B., and R. Aebersold, 2010, Options and considerations when selecting a quantitative proteomics strategy: Nat Biotechnol, v. 28, p. 710-21.
Doucet, C. M., and M. W. Hetzer, 2010, Nuclear pore biogenesis into an intact nuclear envelope: Chromosoma, v. 119, p. 469-77.
Eden, E., D. Lipson, S. Yogev, and Z. Yakhini, 2007, Discovering motifs in ranked lists of DNA sequences: PLoS Comput Biol, v. 3, p. e39.
Eden, E., R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, 2009, GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists: BMC Bioinformatics, v. 10, p. 48.
Elias, J. E., and S. P. Gygi, 2007, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry: Nature methods, v. 4, p. 207-14.

Fagerlund, R., L. Kinnunen, M. Köhler, I. Julkunen, and K. Melén, 2005, NF-{kappa}B is transported into the nucleus by importin {alpha}3 and importin {alpha}4: J Biol Chem, v. 280, p. 15942-51.

Fan, S., and B. Margolis, 2011, The Ran importin system in cilia trafficking: Organogenesis, v. 7, p. 147-53.

Fischer, U., J. Huber, W. C. Boelens, I. W. Mattaj, and R. Lührmann, 1995, The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs: Cell, v. 82, p. 475-83.

Fornerod, M., M. Ohno, M. Yoshida, and I. W. Mattaj, 1997, CRM1 is an export receptor for leucine-rich nuclear export signals: Cell, v. 90, p. 1051-60.

Fredriksson, K., C. M. Van Itallie, A. Aponte, M. Gucek, A. J. Tietgens, and J. M. Anderson, 2015, Proteomic analysis of proteins surrounding occludin and claudin-4 reveals their proximity to signaling and trafficking networks: PLoS One, v. 10, p. e0117074.

Gagnon, J. A., E. Valen, S. B. Thyme, P. Huang, L. Akhmetova, L. Ahkmetova, A. Pauli, T. G. Montague, S. Zimmerman, C. Richter, and A. F. Schier, 2014, Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs: PLoS One, v. 9, p. e98186.

Gatto, L., and K. S. Lilley, 2012, MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation: Bioinformatics, v. 28, p. 288-9.

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

Goeman, J., and A. Solari, 2011, Multiple testing for exploratory research, Statistical Science, p. 584-597.

Goldfarb, D. S., A. H. Corbett, D. A. Mason, M. T. Harreman, and S. A. Adam, 2004, Importin alpha: a multipurpose nuclear-transport receptor: Trends Cell Biol, v. 14, p. 505-14.

Gontan, C., T. Güttler, E. Engelen, J. Demmers, M. Fornerod, F. G. Grosveld, D. Tibboel, D. Görlich, R. A. Poot, and R. J. Rottier, 2009, Exportin 4 mediates a novel nuclear import pathway for Sox family transcription factors: J Cell Biol, v. 185, p. 27-34.

Green, N. M., 1975, Avidin: Adv Protein Chem, v. 29, p. 85-133.

Grünwald, D., R. H. Singer, and M. Rout, 2011, Nuclear export dynamics of RNA-protein complexes: Nature, v. 475, p. 333-41.

Grüter, P., C. Tabernero, C. von Kobbe, C. Schmitt, C. Saavedra, A. Bachi, M. Wilm, B. K. Felber, and E. Izaurralde, 1998, TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus: Mol Cell, v. 1, p. 649-59.

Gstaiger, M., and R. Aebersold, 2009, Applying mass spectrometry-based proteomics to genetics, genomics and network biology: Nat Rev Genet, v. 10, p. 617-27.

Gupta, G. D., É. Coyaud, J. Gonçalves, B. A. Mojarad, Y. Liu, Q. Wu, L. Gheiratmand, D. Comartin, J. M. Tkach, S. W. Cheung, M. Bashkurov, M. Hasegan, J. D. Knight, Z. Y. Lin, M. Schueler, F. Hildebrandt, J. Moffat, A. C. Gingras, B. Raught, and L. Pelletier, 2015, A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface: Cell, v. 163, p. 1484-99.

Görlich, D., S. Kostka, R. Kraft, C. Dingwall, R. A. Laskey, E. Hartmann, and S. Prehn, 1995, Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope: Curr Biol, v. 5, p. 383-92.

Güttinger, S., E. Laurell, and U. Kutay, 2009, Orchestrating nuclear envelope disassembly and reassembly during mitosis: Nat Rev Mol Cell Biol, v. 10, p. 178-91.

Güttler, T., and D. Görlich, 2011, Ran-dependent nuclear export mediators: a structural perspective: EMBO J, v. 30, p. 3457-74.

Hamamoto, T., S. Gunji, H. Tsuji, and T. Beppu, 1983, Leptomycins A and B, new antifungal antibiotics. I. Taxonomy of the producing strain and their fermentation, purification and characterization: J Antibiot (Tokyo), v. 36, p. 639-45.

Hampoelz, B., M. T. Mackmull, P. Machado, P. Ronchi, K. H. Bui, N. Schieber, R. Santarella-Mellwig, A. Necakov, A. Andrés-Pons, J. M. Philippe, T. Lecuit, Y. Schwab, and M. Beck, 2016, Pre-assembled Nuclear Pores Insert into the Nuclear Envelope during Early Development: Cell, v. 166, p. 664-78.

Handschug, K., S. Sperling, S. J. Yoon, S. Hennig, A. J. Clark, and A. Huebner, 2001, Triple A syndrome is caused by mutations in AAAS, a new WD-repeat protein gene: Hum Mol Genet, v. 10, p. 283-90.

Hinshaw, J. E., and R. A. Milligan, 2003, Nuclear pore complexes exceeding eightfold rotational symmetry: J Struct Biol, v. 141, p. 259-68.

Hirano, M., Y. Furiya, H. Asai, A. Yasui, and S. Ueno, 2006, ALADINI482S causes selective failure of nuclear protein import and hypersensitivity to oxidative stress in triple A syndrome: Proc Natl Acad Sci U S A, v. 103, p. 2298-303.

Hoelz, A., E. W. Debler, and G. Blobel, 2011, The structure of the nuclear pore complex: Annu Rev Biochem, v. 80, p. 613-43.

Howe, K., M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E. Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, I. Sealy, M. Caccamo, C. Churcher, C. Scott, J. C. Barrett, R. Koch, G. J. Rauch, S. White, W. Chow, B. Kilian, L. T. Quintais, J. A. Guerra-Assunção, Y. Zhou, Y. Gu, J. Yen, J. H. Vogel, T. Eyre, S. Redmond, R. Banerjee, J. Chi, B. Fu, E. Langley, S. F. Maguire, G. K. Laird, D. Lloyd, E. Kenyon, S. Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M. Jones, M. Quail, D. Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis, C. Clee, K. Oliver, R. Clark, C. Riddle, D. Elliot, D. Eliott, G. Threadgold, G. Harden, D. Ware, S. Begum, B. Mortimore, B. Mortimer, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M. Dunn, C. Johnson, J. Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P. Howden, N. Barker, C. Lloyd, C. Stevens, J. Harley, K. Holt, G. Panagiotidis, J. Lovell, H. Beasley, C. Henderson, D. Gordon, K. Auger, D. Wright, J. Collins, C. Raisen, L. Dyer, K. Leung, L. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C. Griffiths, D. Manthravadi, et al., 2013, The zebrafish reference genome sequence and its relationship to the human genome: Nature, v. 496, p. 498-503.

Huebner, A., A. M. Kaindl, R. Braun, and K. Handschug, 2002, New insights into the molecular basis of the triple A syndrome: Endocr Res, v. 28, p. 733-9.

Huebner, A., P. Mann, E. Rohde, A. M. Kaindl, M. Witt, P. Verkade, S. Jakubiczka, M. Menschikowski, G. Stoltenburg-Didinger, and K. Koehler, 2006, Mice lacking the nuclear pore complex protein ALADIN show female infertility but fail to develop a phenotype resembling human triple A syndrome: Mol Cell Biol, v. 26, p. 1879-87.

Hutten, S., and R. H. Kehlenbach, 2006, Nup214 is required for CRM1-dependent nuclear protein export in vivo: Mol Cell Biol, v. 26, p. 6772-85.

Izaurralde, E., U. Kutay, C. von Kobbe, I. W. Mattaj, and D. Görlich, 1997, The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus: EMBO J, v. 16, p. 6535-47.

Jäkel, S., and D. Görlich, 1998, Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells: EMBO J, v. 17, p. 4491-502.

Kalderon, D., W. D. Richardson, A. F. Markham, and A. E. Smith, 1984, Sequence requirements for nuclear location of simian virus 40 large-T antigen: Nature, v. 311, p. 33-8.

Kalverda, B., and M. Fornerod, 2010, Characterization of genome-nucleoporin interactions in Drosophila links chromatin insulators to the nuclear pore complex: Cell Cycle, v. 9, p. 4812-7.

Kalverda, B., H. Pickersgill, V. V. Shloma, and M. Fornerod, 2010, Nucleoporins directly stimulate expression of developmental and cell-cycle genes inside the nucleoplasm: Cell, v. 140, p. 360-71.

Kann, M., B. Sodeik, A. Vlachou, W. H. Gerlich, and A. Helenius, 1999, Phosphorylation-dependent binding of hepatitis B virus core particles to the nuclear pore complex: J Cell Biol, v. 145, p. 45-55.

Katahira, J., K. Strässer, A. Podtelejnikov, M. Mann, J. U. Jung, and E. Hurt, 1999, The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human: EMBO J, v. 18, p. 2593-609.

Kelley, J. B., A. M. Talley, A. Spencer, D. Gioeli, and B. M. Paschal, 2010, Karyopherin alpha7 (KPNA7), a divergent member of the importin alpha family of nuclear import receptors: BMC Cell Biol, v. 11, p. 63.

Kessel, R. G., 1992, Annulate lamellae: a last frontier in cellular organelles: Int Rev Cytol, v. 133, p. 43-120.

Kim, D. I., K. C. Birendra, W. Zhu, K. Motamedchaboki, V. Doye, and K. J. Roux, 2014, Probing nuclear pore complex architecture with proximity-dependent biotinylation: Proc Natl Acad Sci U S A, v. 111, p. E2453-61.

Kimmel, C. B., W. W. Ballard, S. R. Kimmel, B. Ullmann, and T. F. Schilling, 1995, Stages of embryonic development of the zebrafish: Dev Dyn, v. 203, p. 253-310.

Kimura, M., and N. Imamoto, 2014, Biological significance of the importin-β family-dependent nucleocytoplasmic transport pathways: Traffic, v. 15, p. 727-48.

Kimura, M., S. Kose, N. Okumura, K. Imai, M. Furuta, N. Sakiyama, K. Tomii, P. Horton, T. Takao, and N. Imamoto, 2013a, Identification of cargo proteins specific for the nucleocytoplasmic transport carrier transportin by combination of an in vitro transport system and stable isotope labeling by amino acids in cell culture (SILAC)-based quantitative proteomics: Mol Cell Proteomics, v. 12, p. 145-57.

Kimura, M., Y. Morinaka, K. Imai, S. Kose, P. Horton, and N. Imamoto, 2017, Extensive cargo identification reveals distinct biological roles of the 12 importin pathways: Elife, v. 6.

Kimura, M., N. Okumura, S. Kose, T. Takao, and N. Imamoto, 2013b, Identification of cargo proteins specific for importin-β with importin-α applying a stable isotope labeling by amino acids in cell culture (SILAC)-based in vitro transport system: J Biol Chem, v. 288, p. 24540-9.

Kind, B., K. Koehler, M. Lorenz, and A. Huebner, 2009, The nuclear pore complex protein ALADIN is anchored via NDC1 but not via POM121 and GP210 in the nuclear envelope: Biochem Biophys Res Commun, v. 390, p. 205-10.

Kırlı, K., S. Karaca, H. J. Dehne, M. Samwer, K. T. Pan, C. Lenz, H. Urlaub, and D. Görlich, 2015, A deep proteomics perspective on CRM1-mediated nuclear export and nucleocytoplasmic partitioning: Elife, v. 4.

Kitagawa, Y., M. Kameoka, S. Shoji-Kawata, Y. Iwabu, H. Mizuta, K. Tokunaga, M. Fujino, Y. Natori, Y. Yura, and K. Ikuta, 2008, Inhibitory function of adapter-related protein complex 2 alpha 1 subunit in the process of nuclear translocation of human immunodeficiency virus type 1 genome: Virology, v. 373, p. 171-80.

Kobiler, O., N. Drayman, V. Butin-Israeli, and A. Oppenheim, 2012, Virus strategies for passing the nuclear envelope barrier: Nucleus, v. 3, p. 526-39.

Komeili, A., and E. K. O'Shea, 1999, Roles of phosphorylation sites in regulating activity of the transcription factor Pho4: Science, v. 284, p. 977-80.

Kosugi, S., M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita, and H. Yanagawa, 2009, Six classes of nuclear localization signals specific to different binding grooves of importin alpha: J Biol Chem, v. 284, p. 478-85.

Kotera, I., T. Sekimoto, Y. Miyamoto, T. Saiwaki, E. Nagoshi, H. Sakagami, H. Kondo, and Y. Yoneda, 2005, Importin alpha transports CaMKIV to the nucleus without utilizing importin beta: EMBO J, v. 24, p. 942-51.

Krull, S., J. Dörries, B. Boysen, S. Reidenbach, L. Magnius, H. Norder, J. Thyberg, and V. C. Cordes, 2010, Protein Tpr is required for establishing nuclear pore-associated zones of heterochromatin exclusion: EMBO J, v. 29, p. 1659-73.

Kudo, N., B. Wolff, T. Sekimoto, E. P. Schreiner, Y. Yoneda, M. Yanagida, S. Horinouchi, and M. Yoshida, 1998, Leptomycin B inhibition of signal-mediated nuclear export by direct binding to CRM1: Exp Cell Res, v. 242, p. 540-7.

Kutay, U., F. R. Bischoff, S. Kostka, R. Kraft, and D. Görlich, 1997, Export of importin alpha from the nucleus is mediated by a specific nuclear transport factor: Cell, v. 90, p. 1061-71.

Kwon, K., and D. Beckett, 2000, Function of a conserved sequence motif in biotin holoenzyme synthetases: Protein Sci, v. 9, p. 1530-9.

Köhler, M., C. Speck, M. Christiansen, F. R. Bischoff, S. Prehn, H. Haller, D. Görlich, and E. Hartmann, 1999, Evidence for distinct substrate specificities of importin alpha family members in nuclear protein import: Mol Cell Biol, v. 19, p. 7782-91.

Lee, B. J., A. E. Cansizoglu, K. E. Süel, T. H. Louis, Z. Zhang, and Y. M. Chook, 2006, Rules for nuclear localization sequence recognition by karyopherin beta 2: Cell, v. 126, p. 543-58.

Lee, S. J., Y. Matsuura, S. M. Liu, and M. Stewart, 2005, Structural basis for nuclear import complex dissociation by RanGTP: Nature, v. 435, p. 693-6.

Lemke, E. A., 2016, The Multiple Faces of Disordered Nucleoporins: J Mol Biol, v. 428, p. 2011-24.

Leopold, P. L., and K. K. Pfister, 2006, Viral strategies for intracellular trafficking: motors and microtubules: Traffic, v. 7, p. 516-23.

Lim, Y. S., and B. L. Tang, 2015, A role for Rab23 in the trafficking of Kif17 to the primary cilium: J Cell Sci, v. 128, p. 2996-3008.

Lu, C., C. Shao, E. Cobos, K. P. Singh, and W. Gao, 2012, Chemotherapeutic sensitization of leptomycin B resistant lung cancer cells by pretreatment with doxorubicin: PLoS One, v. 7, p. e32895.

Lun, A. T., and G. K. Smyth, 2014, De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly: Nucleic Acids Res, v. 42, p. e95.

Lupu, F., A. Alves, K. Anderson, V. Doye, and E. Lacy, 2008, Nuclear pore composition regulates neural stem/progenitor cell differentiation in the mouse embryo: Dev Cell, v. 14, p. 831-42.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, and G. M. Church, 2013, RNA-guided human genome engineering via Cas9: Science, v. 339, p. 823-6.

Matsuura, Y., A. Lange, M. T. Harreman, A. H. Corbett, and M. Stewart, 2003, Structural basis for Nup2p function in cargo release and karyopherin recycling in nuclear import: EMBO J, v. 22, p. 5358-69.

Matsuura, Y., and M. Stewart, 2005, Nup50/Npap60 function in nuclear protein import complex disassembly and importin recycling: EMBO J, v. 24, p. 3681-9.

Meyer, T., A. Begitt, I. Lödige, M. van Rossum, and U. Vinkemeier, 2002, Constitutive and IFN-gamma-induced nuclear import of STAT1 proceed through independent pathways: EMBO J, v. 21, p. 344-54.

Mingot, J. M., S. Kostka, R. Kraft, E. Hartmann, and D. Görlich, 2001, Importin 13: a novel mediator of nuclear import and export: EMBO J, v. 20, p. 3685-94.

Miyamoto, Y., M. Hieda, M. T. Harreman, M. Fukumoto, T. Saiwaki, A. E. Hodel, A. H. Corbett, and Y. Yoneda, 2002, Importin alpha can migrate into the nucleus in an importin beta- and Ran-independent manner: EMBO J, v. 21, p. 5833-42.

Nardozzi, J. D., K. Lott, and G. Cingolani, 2010, Phosphorylation meets nuclear import: a review: Cell Commun Signal, v. 8, p. 32.

Nishi, K., M. Yoshida, D. Fujiwara, M. Nishikawa, S. Horinouchi, and T. Beppu, 1994, Leptomycin B targets a regulatory cascade of crm1, a fission yeast nuclear protein, involved in control of higher order chromosome structure and gene expression: J Biol Chem, v. 269, p. 6320-4.

Oeckinghaus, A., M. S. Hayden, and S. Ghosh, 2011, Crosstalk in NF-κB signaling pathways: Nat Immunol, v. 12, p. 695-708.

Ogawa, Y., Y. Miyamoto, M. Asally, M. Oka, Y. Yasuda, and Y. Yoneda, 2010, Two isoforms of Npap60 (Nup50) differentially regulate nuclear protein import: Mol Biol Cell, v. 21, p. 630-8.

Okada, N., Y. Ishigami, T. Suzuki, A. Kaneko, K. Yasui, R. Fukutomi, and M. Isemura, 2008, Importins and exportins in cellular differentiation: J Cell Mol Med, v. 12, p. 1863-71.

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

Ori, A., P. Free, J. Courty, M. C. Wilkinson, and D. G. Fernig, 2009, Identification of heparin-binding sites in proteins by selective labeling: Mol Cell Proteomics, v. 8, p. 2256-65.

Oudhoff, M. J., S. A. Freeman, A. L. Couzens, F. Antignano, E. Kuznetsova, P. H. Min, J. P. Northrop, B. Lehnertz, D. Barsyte-Lovejoy, M. Vedadi, C. H. Arrowsmith, H. Nishina, M. R. Gold, F. M. Rossi, A. C. Gingras, and C. Zaph, 2013, Control of the hippo pathway by Set7-dependent methylation of Yap: Dev Cell, v. 26, p. 188-94.

Panté, N., and M. Kann, 2002, Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm: Mol Biol Cell, v. 13, p. 425-34.

Patel, S. S., and M. F. Rexach, 2008, Discovering novel interactions at the nuclear pore complex using bead halo: a rapid method for detecting molecular interactions of high and low affinity at equilibrium: Mol Cell Proteomics, v. 7, p. 121-31.

Peri, F., and C. Nüsslein-Volhard, 2008, Live imaging of neuronal degradation by microglia reveals a role for v0-ATPase a1 in phagosomal fusion in vivo: Cell, v. 133, p. 916-27.

Prasad, R., L. A. Metherell, A. J. Clark, and H. L. Storr, 2013, Deficiency of ALADIN impairs redox homeostasis in human adrenal cells and inhibits steroidogenesis: Endocrinology, v. 154, p. 3209-18.

Quan, Y., Z. L. Ji, X. Wang, A. M. Tartakoff, and T. Tao, 2008, Evolutionary and transcriptional analysis of karyopherin beta superfamily proteins: Mol Cell Proteomics, v. 7, p. 1254-69.

R Core Team, 2012, R: A Language and Environment for Statistical Computing: Vienna, Austria, R Foundation for Statistical Computing.

Rabe, B., A. Vlachou, N. Panté, A. Helenius, and M. Kann, 2003, Nuclear import of hepatitis B virus capsids and release of the viral genome: Proc Natl Acad Sci U S A, v. 100, p. 9849-54.

Rabut, G., V. Doye, and J. Ellenberg, 2004, Mapping the dynamic organization of the nuclear pore complex inside single living cells: Nat Cell Biol, v. 6, p. 1114-21.

Reichelt, R., A. Holzenburg, E. L. Buhle, M. Jarnik, A. Engel, and U. Aebi, 1990, Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components: J Cell Biol, v. 110, p. 883-94.

Ribbeck, K., and D. Görlich, 2001, Kinetic analysis of translocation through nuclear pore complexes: EMBO J, v. 20, p. 1320-30.

Riggleman, B., E. Wieschaus, and P. Schedl, 1989, Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene: Genes Dev, v. 3, p. 96-113.

Roth, D. M., G. W. Moseley, C. W. Pouton, and D. A. Jans, 2011, Mechanism of microtubule-facilitated "fast track" nuclear import: J Biol Chem, v. 286, p. 14335-51.

Roux, K. J., D. I. Kim, M. Raida, and B. Burke, 2012, A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells: J Cell Biol, v. 196, p. 801-10.

Róna, G., H. L. Pálinkás, M. Borsos, A. Horváth, I. Scheer, A. Benedek, G. N. Nagy, I. Zagyva, and B. G. Vértessy, 2014, NLS copy-number variation governs efficiency of nuclear import--case study on dUTPases: FEBS J, v. 281, p. 5463-78.

Sangel, P., M. Oka, and Y. Yoneda, 2014, The role of Importin-βs in the maintenance and lineage commitment of mouse embryonic stem cells: FEBS Open Bio, v. 4, p. 112-20.

Savas, J. N., B. H. Toyama, T. Xu, J. R. Yates, and M. W. Hetzer, 2012, Extremely long-lived nuclear pore proteins in the rat brain: Science, v. 335, p. 942.

Schooley, A., B. Vollmer, and W. Antonin, 2012, Building a nuclear envelope at the end of mitosis: coordinating membrane reorganization, nuclear pore complex assembly, and chromatin de-condensation: Chromosoma, v. 121, p. 539-54.

Sessler, R. J., and N. Noy, 2005, A ligand-activated nuclear localization signal in cellular retinoic acid binding protein-II: Mol Cell, v. 18, p. 343-53.

Smyth, G. K., R. Gentleman, S. Carey, R. Dudoit, R. Irizarry, and W. Huber, 2005, Limma: linear models for microarray data: Bioinformatics and Computational Biology Solutions using R and Bioconductor, Springer.

Soniat, M., and Y. M. Chook, 2015, Nuclear localization signals for four distinct karyopherin-β nuclear import systems: Biochem J, v. 468, p. 353-62.

Spilianakis, C., J. Papamatheakis, and A. Kretsovali, 2000, Acetylation by PCAF enhances CIITA nuclear accumulation and transactivation of major histocompatibility complex class II genes: Mol Cell Biol, v. 20, p. 8489-98.

Spiller, M. P., M. A. Reijns, and J. D. Beggs, 2007, Requirements for nuclear localization of the Lsm2-8p complex and competition between nuclear and cytoplasmic Lsm complexes: J Cell Sci, v. 120, p. 4310-20.

Strimmer, K., 2008, fdrtool: a versatile R package for estimating local and tail area-based false discovery rates: Bioinformatics, v. 24, p. 1461-2.

Sukegawa, J., and G. Blobel, 1993, A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and faces the nucleoplasm: Cell, v. 72, p. 29-38.

Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc, 2011, REVIGO summarizes and visualizes long lists of gene ontology terms: PLoS One, v. 6, p. e21800.

Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, and C. von Mering, 2015, STRING v10: protein-protein interaction networks, integrated over the tree of life: Nucleic Acids Res, v. 43, p. D447-52.

Thakar, K., S. Karaca, S. A. Port, H. Urlaub, and R. H. Kehlenbach, 2013, Identification of CRM1-dependent Nuclear Export Cargos Using Quantitative Mass Spectrometry: Mol Cell Proteomics, v. 12, p. 664-78.

Truong, K., T. D. Lee, B. Li, and Y. Chen, 2012, Sumoylation of SAE2 C terminus regulates SAE nuclear localization: J Biol Chem, v. 287, p. 42611-9.

Tullio-Pelet, A., R. Salomon, S. Hadj-Rabia, C. Mugnier, M. H. de Laet, B. Chaouachi, F. Bakiri, P. Brottier, L. Cattolico, C. Penet, M. Bégeot, D. Naville, M. Nicolino, J. L. Chaussain, J. Weissenbach, A. Munnich, and S. Lyonnet, 2000, Mutant WD-repeat protein in triple-A syndrome: Nat Genet, v. 26, p. 332-5.

Twyffels, L., C. Gueydan, and V. Kruys, 2014, Transportin-1 and Transportin-2: protein nuclear import and beyond: FEBS Lett, v. 588, p. 1857-68.

Varnaitė, R., and S. A. MacNeill, 2016, Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID: Proteomics, v. 16, p. 2503-2518.

von Appen, A., J. Kosinski, L. Sparks, A. Ori, A. L. DiGuilio, B. Vollmer, M. T. Mackmull, N. Banterle, L. Parca, P. Kastritis, K. Buczak, S. Mosalaganti, W. Hagen, A. Andres-Pons, E. A. Lemke, P. Bork, W.

Antonin, J. S. Glavy, K. H. Bui, and M. Beck, 2015, In situ structural analysis of the human nuclear pore complex: Nature, v. 526, p. 140-3.

Wang, W., X. Yang, T. Kawai, I. López de Silanes, K. Mazan-Mamczarz, P. Chen, Y. M. Chook, C. Quensel, M. Köhler, and M. Gorospe, 2004, AMP-activated protein kinase-regulated phosphorylation and acetylation of importin alpha1: involvement in the nuclear import of RNA-binding protein HuR: J Biol Chem, v. 279, p. 48376-88.

Weil, T. T., R. M. Parton, and I. Davis, 2010, Making the message clear: visualizing mRNA localization: Trends Cell Biol, v. 20, p. 380-90.

Wente, S. R., and M. P. Rout, 2010, The nuclear pore complex and nuclear transport: Cold Spring Harb Perspect Biol, v. 2, p. a000562.

Westerfield, M., 2007, The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio): Eugene, University of Oregon Press.

Winkler, J., S. Roessler, C. Sticht, A. L. DiGuilio, E. Drucker, K. Holzer, E. Eiteneuer, E. Herpel, K. Breuhahn, N. Gretz, P. Schirmacher, A. Ori, and S. Singer, 2016, Cellular apoptosis susceptibility (CAS) is linked to integrin β1 and required for tumor cell migration and invasion in hepatocellular carcinoma (HCC): Oncotarget, v. 7, p. 22883-92.

Wälde, S., and R. H. Kehlenbach, 2010, The Part and the Whole: functions of nucleoporins in nucleocytoplasmic transport: Trends Cell Biol, v. 20, p. 461-9.

Wühr, M., T. Güttler, L. Peshkin, G. C. McAlister, M. Sonnett, K. Ishihara, A. C. Groen, M. Presler, B. K. Erickson, T. J. Mitchison, M. W. Kirschner, and S. P. Gygi, 2015, The Nuclear Proteome of a Vertebrate: Curr Biol, v. 25, p. 2663-71.

Yamazumi, Y., A. Kamiya, A. Nishida, A. Nishihara, S. Iemura, T. Natsume, and T. Akiyama, 2009, The transmembrane nucleoporin NDC1 is required for targeting of ALADIN to nuclear pore complexes: Biochem Biophys Res Commun, v. 389, p. 100-4.

Yang, F., Y. Shen, D. G. Camp, and R. D. Smith, 2012, High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis: Expert Rev Proteomics, v. 9, p. 129-34.

Yasuhara, N., M. Oka, and Y. Yoneda, 2009, The role of the nuclear transport system in cell differentiation: Semin Cell Dev Biol, v. 20, p. 590-9.

Yasuhara, N., R. Yamagishi, Y. Arai, R. Mehmood, C. Kimoto, T. Fujita, K. Touma, A. Kaneko, Y. Kamikawa, T. Moriyama, T. Yanagida, H. Kaneko, and Y. Yoneda, 2013, Importin alpha subtypes determine differential transcription factor localization in embryonic stem cells maintenance: Dev Cell, v. 26, p. 123-35.

Yates, J. R., C. I. Ruse, and A. Nakorchevsky, 2009, Proteomics by mass spectrometry: approaches, advances, and applications: Annu Rev Biomed Eng, v. 11, p. 49-79.

Zhu, W., J. W. Smith, and C. M. Huang, 2010, Mass spectrometry-based label-free quantitative proteomics: J Biomed Biotechnol, v. 2010, p. 840518.

Zuccolo, M., A. Alves, V. Galy, S. Bolhy, E. Formstecher, V. Racine, J. B. Sibarita, T. Fukagawa, R. Shiekhattar, T. Yen, and V. Doye, 2007, The human Nup107-160 nuclear pore subcomplex contributes to proper kinetochore functions: EMBO J, v. 26, p. 1853-64.