NUCLEOLIN: A NUCLEOLAR RNA-BINDING PROTEIN

INVOLVED IN RIBOSOME BIOGENESIS

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Inaugural-Dissertation

zur Erlangung des Doktorgrades der

Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Julia Fremerey

aus Hamburg

Düsseldorf, April 2016

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen

Fakultät der Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. A. Borkhardt

Korreferent: Prof. Dr. H. Schwender

Tag der mündlichen Prüfung: 20.07.2016

Die vorgelegte Arbeit wurde von Juli 2012 bis März 2016 in der Klinik für Kinder-Onkologie, -Hämatologie und Klinische Immunologie des Universitätsklinikums Düsseldorf unter Anleitung von Prof. Dr. A. Borkhardt und in Kooperation mit dem ‚Laboratory of RNA Molecular Biology' an der Rockefeller Universität unter Anleitung von Prof. Dr. T. Tuschl angefertigt.

Dedicated to my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATION

%	percent
ΔC	delta C terminus
ΔH	enthalpy
ΔN	delta N terminus
ΔRRM	delta RRM
°C	celsius
µ	micro
4SU	4-thiouridine
6SG	6-thioguanosine
A	adenine
A	ampere
aa	amino acid
ad.	add
ADP	adenosine diphosphate
ALL	acute lymphatic leukemia
ALS	amyotrophic lateral sclerosis
AML	acute myeloid leukemia
APS	ammonium persulfate
ARE	AU rich element
ATP	adenosine triphosphate
B23	nucleophosmin
BCA	bicinchoninic acid
BFDR	Bayesian false discovery rate
blast	blasticidin
BMFZ	Biologisch-Medizinisches Forschungszentrum
bp	base pair
BRIX1	ribosome biogenesis protein BRX1 homolog
BSA	bovine serum albumin
C	cytosine
C23	nucleolin
cDNA	complementary DNA
Chip-Seq	chromatin immunoprecipitation-sequencing
CKII	casein kinase II
control counts	spectral counts in the negative control

D	aspartic acid
d0	distance 0 (no mismatch)
d1	distance 1 (one mismatch)
d2	distance 2 (two mismatches)
DAPI	4',6-diamidino-2-phenylindole
DFC	dense fibrillary component
DMEM	dulbecco's modified eagle's medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
Dox	doxycycline
DREME	discriminative regular expression motif elicitation
ds	double stranded
DTT	dithiothreitol
E	glutamic acid
e.g.	exemplī grātiā/for example
ECL	enhanced chemi luminescence
ECM	evolutionary conserved motif
EDTA	ethylenediamine tetraacetic acid
EGTA	ethylene glycol tetraacetic acid
eIF	eukaryotic initiation factor
EMG1	essential for mitotic growth
EMSA	electro mobility shift assay
ETS	external transcribed spacer
F	phenylalanine
FA	formamide
FACT	facilitates chromatin transcription
FBL	fibrillarin
FBS	fetal bovine serum
FC	fibrillary center
FC	fold change, average spectral count in test interaction divided by the average in controls
FDR	false discovery rate
FISH	fluorescence in-situ hybridization
FLAG	affinity tag, sequence: DYKDDDDK
g	gravitation
G	guanine
GC	granular component
GFC	gel filtration chromatography

GFP	green fluorescent protein
GOI	gene of interest
GS	goat serum
GST	glutathione S-transferases
h	hour
H_2N	amino terminus
HA	hemagglutinin
HEK	human embryonic kidney
HeLa	Henrietta Lacks
hg	human genome
HMG	high-mobility group
HRP	horseradish peroxidase
Hygro	hygromycin
IGS	intergenic spacer
IHC	immunohistochemistry
IP	immunoprecipitation
ITC	isothermal-titration calorimetry
ITS	internal transcribed spacer
K	lysine
K_a	affinity constant
Kb	kilo base
K_d	dissociation constant
KD	knockdown
KDa	kilo Dalton
KO	knockout
L	leucine
L	ligand
LC-ESI MS	liquid chromatography-electrospray Ionization-tandem mass spectrometry
LFQ	label-free quantification
LSU	large subunit
M	molar
mAb	monoclonal antibody
mg	milligram
min	minute
miRNA	microRNA
MKI67IP	MKI67 FHA domain-interacting nucleolar phosphoprotein
ml	milliliter

mM	milli molar
mRNA	messenger RNA
MS	mass-spectrometry/mass-spec
MSKCC	Memorial-Sloan-Kettering cancer center
MW	molecular weight
MYBBP1A	Myb-binding protein 1A
N	stoichiometry
NCL	nucleolin
NGS	next generation sequencing
NLS	nuclear localization signal
nm	nanometer
NMR	nuclear magnetic resonance
NoD	nucleolar localization sequence detector
NoLS	nucleolar localization signal
NOP	nucleolar protein
NP40	nonidet P-40
NPM1	nucleophosmin
NRE	nucleolin recognition element
NSCLC	Non-small cell lung cancer
nt	nucleotide
OH	3' hydroxyl terminus
P	protein
pAb	polyclonal antibody
PAGE	polyacrylamide gel electrophoresis
PAR-CLIP	photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pH	potentia hydro genii
PIC	protease inhibitor cocktail
PL	protein-ligand complex
Pol	polymerase
ppm	parts per million
pre	precursor
R	arginine
RBD	RNA-binding domain
RBP	RNA-binding protein
rDNA	ribosomal DNA

RIP-seq	RNA immunoprecipitation sequencing
RNA	ribonucleic acid
RNAi	RNA interference
Rnl2	RNA ligase 2
RP	ribosomal protein
RPKM	reads per kilobase of transcript per million reads mapped
rpm	rounds per minute
RRM	RNA recognition motif
rRNA	ribosomal RNA
RT	reverse transcriptase/reverse transcription
RT	room temperature
S	Svedberg, sedimentation rate
S	serine
SAINT	significant analysis of INTeractome
SAM	significance analysis of microarrays
SBDS	Shwachman-bodian-diamond syndrome gene
SDS	sodium dodecyl sulfate
SEC	size-exclusion chromatography
sec	second
SELEX	systemic evolution of ligands by exponential enrichment
siRNA	small interfering RNA
SL-1	selectivity factor-1
snoRD	C/D box snoRNA
snoRNA	small nucleolar RNA
snoRNP	small nucleolar ribonucleoprotein
spec	spectral counts for the bait-prey pair
ss	single stranded
SSU	small subunit
SUMO	small ubiquitin-related modifier
T	threonine
T	thymine
T0	promoter-proximal terminator
TAE	tris-acetate-EDTA
TBB	tetrabromobenzotriazole
TBE	tris-boric acid-EDTA
TBP	TATA-binding protein
TBS-T	tris buffered saline-tween

TEMED	tetramethylethylenediamine
TFA	trifluoroacetic acid
tRNA	transfer RNA
TTF-1	transcription termination factor-1
U	uracil
UBTF	upstream binding transcription factor
UCE	upstream control element
UTR	untranslated region
UV	ultraviolet
V	Volt
W	Watt
W	tryptophan
XL	crosslink
XRN2	5' – 3' exoribonuclease 2
Y	tyrosine
YBX1	Y box binding protein 1
α	alpha
β	beta
γ ^{32}P-ATP	adenosine triphosphate, labeled on the gamma phosphate group with ^{32}P
γ	gamma
Ψ	psi

ABSTRACT

In eukaryotes, ribosome biogenesis is a highly coordinated process that comprises multiple steps, including polymerase I transcription of ribosomal DNA, processing and modification of the precursor ribosomal RNA transcript, formation of mature ribosomal RNA (18S, 5.8S, 28S) and assembly of the small (40S) and large (60S) ribosomal subunits to form the mature ribosome (80S) within the cytoplasm. This important pathway requires several ribosomal proteins and ribosome biogenesis factors including RNA-binding proteins (RBPs), RNA-helicases, and endo- and exonucleases. RBPs represent 7.5% of all protein-coding genes in humans and coordinate RNA processing and post-transcriptional gene regulation. Altered expression or mutations in genes encoding ribosomal proteins, ribosome biogenesis factors or RBPs can lead to cancer and other genetic diseases. To understand the link between ribosome biogenesis and the molecular mechanisms of cancer, it is crucial to elucidate the function of RBPs that are associated with ribosome biogenesis. Nucleolin (NCL) is one of these factors, it is a multifunctional, highly abundant, nucleolar RBP, which contains four RNA recognition motifs and is implicated in ribosome biogenesis, gene expression and regulation of polymerase I transcription. High expression levels of NCL were previously observed in cancer cells. However, until today, there are no described mutations or splicing variants of NCL which are related to cancer.

This thesis focuses on the characterization of the RNA-binding properties and protein-protein interactions of NCL in the context of ribosome biogenesis. In order to identify the cellular RNA targets of the RNA-binding protein NCL, photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) as well as RNA immunoprecipitation followed by deep sequencing (RIP-Seq) were performed. Small nucleolar RNAs and precursor ribosomal RNA were identified as main targets of NCL, both of them being directly associated with ribosome biogenesis. Biochemical

and structural studies revealed that NCL preferentially interacts with G/C- rich repeat sequences within the precursor ribosomal RNA. This is in concert with the nucleotide composition of the primary ribosomal RNA transcript, which contains up to 72% G and C residues. In a proteomic based approach using NCL-specific immunoprecipitation followed by mass-spectrometry analysis, 119 proteins including several ribosome biogenesis factors, ribosomal proteins of the large and the small subunit, histones, RBPs and RNA-helicases were identified. Furthermore, RNA interference experiments were conducted to further elucidate the impact of NCL on ribosome biogenesis. SiRNA mediated gene knockdown of NCL showed stable expression of mature ribosomal RNA, while precursor ribosomal RNA was slightly decreased.

In summary, the transcriptome and proteome wide targets of NCL that were identified within this thesis, directly associate NCL with the highly coordinated process of ribosome biogenesis.

ZUSAMMENFASSUNG

Die ribosomale Biogenese ist ein fundamentaler und sehr komplexer Prozess, der in Eukaryoten mit der Polymerase-I-vermittelten Transkription ribosomaler DNA im Nucleolus beginnt. Das primäre ribosomale RNA-Transkript durchläuft anschließend mehrere Prozessierungs- und Modifikationsschritte und generiert die reifen ribosomalen RNAs (18S, 5.8S, 28S). Diese bilden zusammen mit weiteren Proteinkomplexen das reife Ribosom (80S), welches aus einer kleinen (40S) und großen (60) ribosomalen Untereinheiten besteht und in das Cytoplasma transportiert wird. Die Ribosomen spielen bei der Proteinbiosynthese, während der Translation der *messenger* RNA eine zentrale Rolle. An der Synthese von Ribosomen sind zahlreiche ribosomale Biogenese Faktoren, ribosomale Proteine, RNA-Helikasen, Endo- und Exonukleasen, sowie zahlreiche Ribonukleoproteinkomplexe und RNA-bindende Proteine beteiligt. Diese umfassen 7.5% aller Protein kodierenden Gene und spielen eine wichtige Rolle bei post-transkriptionellen Mechanismen. Mutationen in Genen, die ribosomale Proteine oder ribosomale Biogenese Faktoren sowie RNA-bindende Proteine kodieren, können Krebs und andere genetisch bedingte Krankheiten, wie z.B. Ribosomopathien, auslösen. Um den Zusammenhang zwischen den molekularen Mechanismen bei der Entstehung von Krebs und dem Prozess der ribosomalen Biogenese genauer zu verstehen, muss die Funktion von ribosomalen Biogenese Faktoren näher untersucht werden. Nucleolin ist ein nukleoläres, multifunktionales, RNA-bindendes Protein, das an der ribosomalen Biogenese, sowie an der Genexpression und Regulation der Polymerase-I Transkription beteiligt ist.

Im Fokus dieser Arbeit steht die Charakterisierung der RNA-Bindestellen und die Protein-Protein Interaktionen von Nucleolin. In *photoactivatable-ribonucleoside-enhanced crosslinking* und *immunoprecipitation* (PAR-CLIP) sowie RNA *immunoprecipitation-sequencing* (RIP-Seq) Analysen wurden *small nucleolar* RNAs

und ribosomale Vorläufer-RNA als Targets von Nucleolin identifiziert. Um die Bindungsaffinität der identifizierten Nucleolin-RNA Komplexe näher zu bestimmen, wurden biochemische und strukturelle Untersuchungen durchgeführt. Diese weisen darauf hin, dass Nucleolin bevorzugt mit G/C-reichen Sequenzen interagiert. Dies stimmt mit der Nukleotidsequenz des primären ribosomalen RNA Transkripts überein, welches ein Target von Nucleolin ist und zu 72% aus G/C- reichen Regionen besteht. Mit Hilfe von massenspektrometrischer Analysen wurden insgesamt 119 Interaktionspartner von Nucleolin identifiziert, darunter frühe und späte ribosomale Biogenese Faktoren, ribosomale Proteine der kleinen und großen ribosomalen Untereinheit, sowie RNA-Helikasen, RNA-bindende Proteine und Histone. Um den Einfluss von Nucleolin auf die ribosomale Biogenese und auf die Expression des primären ribosomalen RNA-Transkripts näher zu untersuchen, wurden RNA-Interferenz Experimente durchgeführt. Diese zeigten eine komplette Abnahme von Nucleolin auf RNA-Ebene, eine stabile Expression reifer ribosomaler RNA und eine geringe Reduktion ribosomaler Vorläufer-RNA.

Zusammenfassend kann gesagt werden, dass Nucleolin mit Faktoren auf Transkriptom- und Proteom- Ebene interagiert, die direkt mit der ribosomalen Biogenese assoziiert sind. Demzufolge spielt Nucleolin eine große Rolle während der rDNA Transkription, sowie in frühen und späten Prozessierungsschritten der ribosomalen Biogenese im Nukleolus.

1 INTRODUCTION

1.1 The nucleolus, the site of ribosome biogenesis

The nucleolus is a dynamic membrane-less suborganelle of the nucleus and was identified more than 200 years ago. Studies in the early 1960s defined the nucleolus as the organelle of ribosomal RNA (rRNA) synthesis and ribosome assembly. Later, in the 1970s and 1980s ribosome biogenesis (section 1.2) and the subcompartments of the nucleolus were described in more detail. Electron microscopy analyses revealed that the nucleolus comprises a 'tripartite architecture' (Lam and Trinkle-Mulcahy, 2015) containing a fibrillar center (FC), a dense fibrillar component (DFC) and a granular component (GC).

Recent bioinformatic analyses of the human nucleolar proteome showed that approximately 30% of the identified proteins are associated with ribosome biogenesis. This is in agreement with the main function of the nucleolus, including transcription and processing of rRNAs and subsequent assembly of the ribosomal subunits (Olson, 2011). Whereas various functions of other nucleolar proteins including pre messenger RNAs (mRNA) processing factors and proteins associated with cell cycle control, DNA replication and DNA repair give rise to additional processes. Nucleoli are also involved in the process and export of several mRNAs and transfer RNAs (tRNA), regulation of the cell cycle, apoptosis, RNA processing and telomerase production (Pederson, 2002; Storck et al., 2007) and maturation and biogenesis of other ribonucleoparticles (RNPs), like the spliceosomal small nuclear RNPs or the signal recognition particle (SRP) (Hein et al., 2013; Quin et al., 2014). However, the most surprising result of the nucleolar proteome analysis was the high level of novel and previously not characterized proteins. Andersen et al. identified 271 nucleolar proteins and grouped them into eight different classes, comprising 22% nucleic acid binding proteins, 14% ribosomal proteins (RPs), 9% RNA modifying enzymes and related proteins, 5% DEAD box

proteins implicating RNA dependent ATPases, 5% chaperones, 3.5% other translation factors, 11% others and 30.5% novel and uncharacterized nucleolar proteins. Over the last 10 years, the nucleolar proteome has increased from less than 100 proteins to more than 4,500 proteins by proteomic analyses using mass-spectrometry (Ahmad et al., 2009; Andersen et al., 2005; Andersen et al., 2002; Leung et al., 2006; Scherl et al., 2002; Scott et al., 2010). A simple target motif common in all of these nucleolar proteins was not identified so far. Therefore, a variety of mechanisms might play a role in nucleolar localization of proteins (Andersen et al., 2002).

Nucleolin (NCL) is one of these nucleolar proteins, it is a non-ribosomal, highly abundant, RNA-binding protein that is involved in ribosome biogenesis (Medina et al., 2010). The structure and function of NCL will be elucidated in more detail in Section 1.4. A comparison of the human and yeast proteome revealed that ~90% of the nucleolus related to yeast proteins have a distinct human homolog. Consequently, the nucleolar proteome is highly conserved throughout evolution (Ahmad et al., 2009; Andersen et al., 2005; Andersen et al., 2002; Leung et al., 2006; Scherl et al., 2002; Scott et al., 2010) (Olson, 2011).

1.2 Ribosome biogenesis in eukaryotes

Ribosome biogenesis is a highly coordinated and efficient process, which starts with the rDNA transcription in the nucleolus to form the mature ribosome (80S) in the cytoplasm (Mayer and Grummt, 2006). The process comprises five main steps (Figure 1), including (1) polymerase I (Pol I) transcription of ribosomal DNA (rDNA), which occurs in the FC (Carmo-Fonseca et al., 2000), (2) processing of the precursor rRNA in the DFC and (3) modification in the GC, followed by (4) transport and (5) assembly of the small (40S) and the large (60S) ribosomal subunit in the cytoplasm (Cooper, 2000; (Boisvert et al., 2007). Early processing factors like UBTF, NOP56, NOP58 and NCL

(Tafforeau et al., 2013) are mostly located in the DFC, whereas the GC comprises late rRNA processing factors like NPM1, ribosomal subunit assembly factors and several ribosomal proteins (Bouvet et al., 1998; Hein et al., 2013; Medina et al., 2010). The ribosome biogenesis factors are explained in more detail in section 1.2.3.
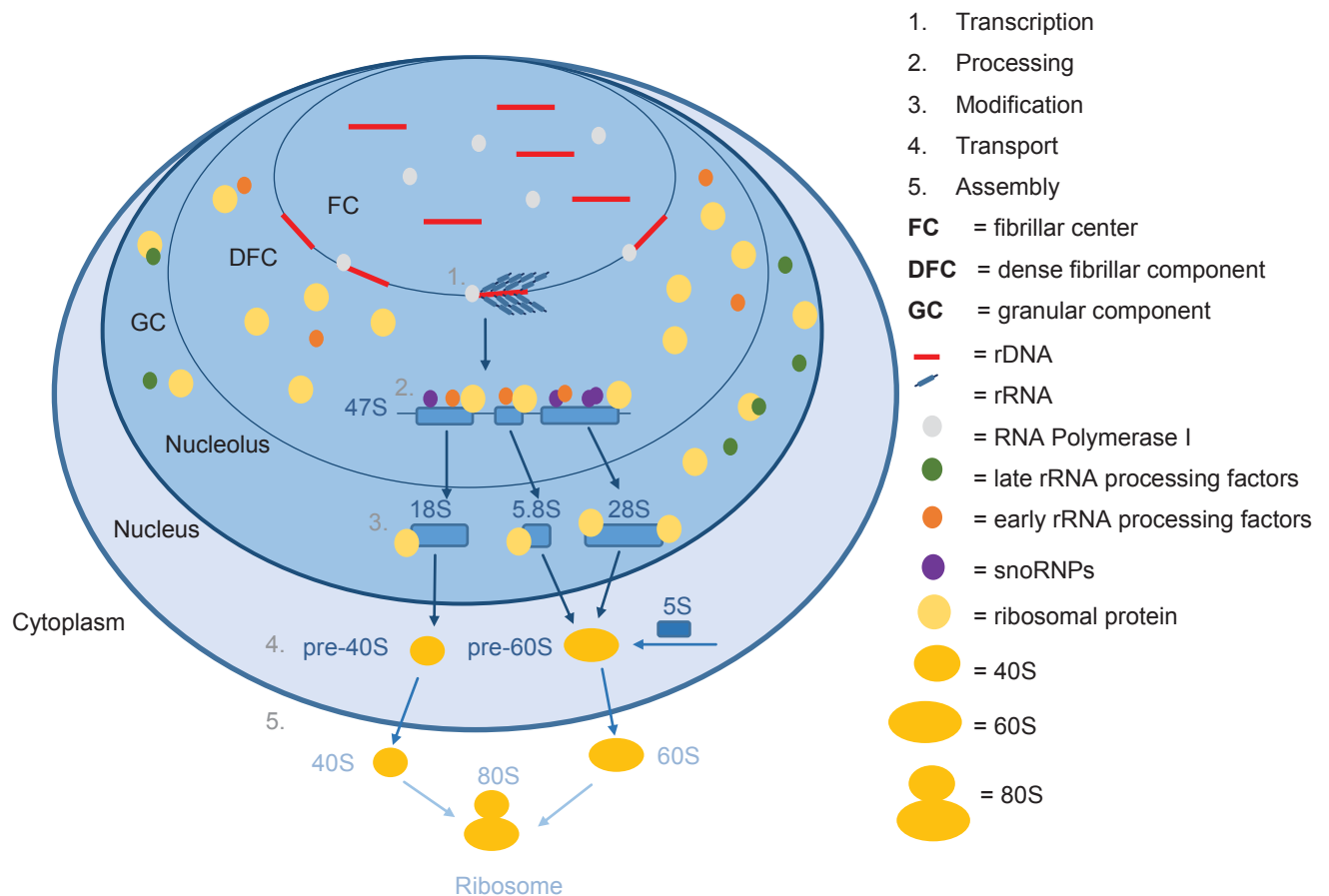
Figure 1: The Nucleolus, the site of ribosome biogenesis in eukaryotes

The three sub-compartments of the nucleolus containing the fibrillar center (FC), dense fibrillar center (DFC) and the granular center (GC). The GC mainly comprises late ribosomal RNA processing factors (green), the DFC contains mostly early processing factors (orange) and the FC region contains the rRNA transcription machinery including Polymerase I, which transcribes rDNA arranged in a head to tail tandem repeat.

Ribosome biogenesis requires four rRNA molecules (5S, 5.8S, 18S, 28S), up to 80 ribosomal proteins, over 150 non-ribosomal proteins and ribosome biogenesis factors, several assembly factors, 70 small nucleolar RNAs (snoRNAs) and small nucleolar RNPs, RNA-helicases, endo- and exoribonucleases and the three RNA polymerases (Pol I, II, III) (Mayer and Grummt, 2006; Robledo et al., 2008; Scherl et al., 2002; Turowski and Tollervey, 2015). RNA Pol I transcribes a long, polycistronic 47S precursor rRNA transcript from tandem rDNA repeats (Turowski and Tollervey, 2015). Pol II transcribes mRNA, snoRNAs, small nuclear RNAs and microRNAs. Pol III transcribes tRNAs, the 5S rRNA, which is part of the large ribosomal subunit and other small RNAs. Among the three RNA polymerases, Pol I is the most active enzyme and accounts up to 60% of transcriptional activity within eukaryotic cells. The RNA Pol I transcription machinery comprises a multiprotein complex and contains among several other factors, the upstream binding transcription factor (UBTF) and the selectivity factor (SL-1). SL-1 is a complex including TATA-binding proteins (TBP)　like TAF_I100, TAF_I63, TAF_I48, TAF_I41 and TAF_I12 (Goodfellow and Zomerdijk, 2013). UBTF, also known as UBF, is a member of the sequence non-specific high mobility group (HMG) that is able to bend DNA. UBTF was found along the entire transcribed region of the rDNA and contributes to the active chromatin state of the rDNA (Cong et al., 2012; Mayer and Grummt, 2006). It binds to the promoter region and recruits SL-1 and Pol I, which leads to the assembly of the preinitiation complex and activates rDNA transcription (Drygin et al., 2010; Quin et al., 2014; Ruggero and Pandolfi, 2003). Therefore, UBTF might play a role in promoter escape as well as in elongation (Drygin et al., 2010; Panov et al., 2006; Stefanovsky et al., 2006). Downregulation of UBTF results in a decrease of transcribed rDNA repeats (Drygin et al., 2010; Sanij et al., 2008).

1.2.1 The human ribosome and ribosomal proteins

The human ribosome is a 4.3 MDa ribonucleoprotein particle (RNP) complex that consists of two-third RNAs and one-third proteins and contains two diverse subunits, a small and a large subunit (Moss et al., 2007). Eukaryotic ribosomes contain approximately 80 proteins that are required for ribosome biogenesis (Korobeinikova et al., 2012; Nikolay et al., 2001), 47 proteins occurring in the large subunit and 33 proteins in the small subunit (Anger et al., 2013; Khatter et al., 2015; Thomson et al., 2013). In eukaryotes, the small subunit (SSU) has a sedimentation coefficient of 40S and consists of the 18S rRNA, whereas the large subunit (LSU) contains the 5.8S, the 28S and the 5S rRNA and sediments at 60S. Both subunits assemble to the 80S ribosome (Meister, 2011) (Anger et al., 2013; Khatter et al., 2015). The large subunit has a peptidyltransferase activity and catalyzes the peptide bond formation, while the small ribosomal subunit is involved in the process, assembly and maturation of the 40S (Phipps et al., 2011a) and has a decoding function of the ribosome (Lafontaine and Tollervey, 2001).

RPs are numbered according to their localization in a two-dimensional polyacrylamide gel. Consequently, small basic proteins have large numbers, while large acidic proteins have small numbers. In general, RPs interact with the negative charges of the phosphates residues of the rRNA and are usually very basic (Nikolay et al., 2001). RPs translocate from the cytoplasm to the nucleolus and translocate back to the cytoplasm as pre ribosomal particles (Leung and Lamond, 2003). RPs are involved in the accurate folding of rRNA and are responsible for correct processing, cleavage and assembly of the ribosome (Robledo et al., 2008). Most of the RPs contain conserved domains that interact with ribosomal RNA (Ruggero and Pandolfi, 2003), some of the them have a RNA-binding motif or a RNA recognition element (Nikolay et al., 2001). RNA-binding proteins are described in more detail in section 1.3.

1.2.2 Ribosomal RNA and the precursor ribosomal RNA transcript

Ribosomal RNAs are the major components of the ribosomal subunits within the ribosome and represent up to 80% of the total RNA in a cell (Turowski and Tollervey, 2015). The coding sequence for three of the four required mature ribosomal RNAs (18S, 5.8S and 28S) is present in a rDNA repeating unit arranged in a head to tail tandem repeat, in the so called nucleolar organizer regions (NORs). The NORs are located on the five acrocentric chromosomes 13, 14, 15, 21 and 22 within the human genome (hg) (Hernandez-Verdun et al., 2002). The fourth rRNA molecule, the 5S rRNA, which is part of the large ribosomal subunit is encoded by genes outside of the nucleolus and is transcribed by Pol III (Carmo-Fonseca et al., 2000). The hg has approximately 400 copies of rDNA, whereas the yeast genome contains around 150 to 200 rDNA repeats (Cong et al., 2012; Goodfellow and Zomerdijk, 2013). In the hg, each rDNA repeating unit is approximately 43 kb. Every cluster is composed of a ~30 kb intergenic spacer (IGS) including regulatory elements like promoter, enhancer and terminator. The transcribed region of the 47S precursor rRNA (Figure 3) spans about 13 kb including the mature rRNA, a 5' and 3' external transcribed spacer (ETS) and two internal transcribed spacer domains (ITS1, ITS2) that separate the mature rRNAs (Goodfellow and Zomerdijk, 2013). The nascent pre rRNA transcript branches off from the actively transcribed rRNA genes and forms a characteristic Christmas tree structure that can be visualized by electron microscopy (Figure 2). The RNP complexes, also named terminal knobs or balls, are responsible for rRNA processing (Kass and Sollner-Webb, 1990; Mougey et al., 1993).

The accessibility of the rDNA is a crucial step in the regulation of ribosome biogenesis and depends on the chromatin state. Only half of the rRNA genes are transcriptionally active and capable for Pol I transcription (Durut and Saez-Vasquez, 2015). These are in an euchromatin state, whereas rDNA that is packaged in a heterochromatin is

inaccessible for the transcription machinery and consequently inactive (Carmo-Fonseca et al., 2000; Turowski and Tollervey, 2015).
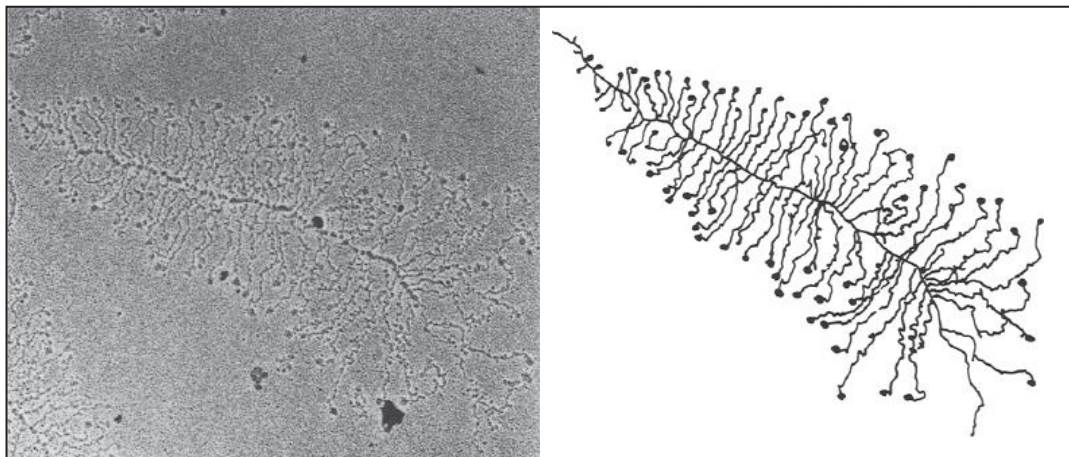
Figure 2: An electronic microscopy image of the transcription unit of rRNA genes

Nuclei of *X. laevis* oocytes were isolated and the chromatin spreading was visualized by electron microscopy using the technique of Miller and Beatty (Miller and Beatty, 1969). When chromatin spreads, active rRNA genes show a characteristic Christmas tree structure comprised of a DNA 'trunk' packed with 'branches' of nascent rRNA transcripts and terminal 'balls' at the end of the ribosomal transcripts. Figure: (Mougey et al., 1993)

The primary 47S rRNA transcript (pre rRNA) undergoes several processing steps including chemical modifications and a series of endo- and exonucleolytic cleavage steps. Two alternative pathways - a minor and a major pathway - were identified and facilitate the cleavage processes. Figure 3 represents the major pathway of pre ribosomal RNA processing. The first cleavage steps of the primary transcript occur in the 5'ETS (01), ~600 nt downstream of the transcription start site (TSS), and in the 3'ETS (02) at position C 12,969. These two cleavage steps result in the 45S pre rRNA. Two cleavage sites for 01 were identified, one at position C414/C416 and a second one six nt downstream, at position G420/U422 (Mullineux and Lafontaine, 2012). The 45S rRNA is directly cleaved in ITS1 at position C6,469/G6,479 (2), which generates the 30S and 32S transcript. The 30S is cleaved at position G1,643 (A0) to create the 26S, which is processed at cleavage at residue U3,653 (1) to generate the 21S transcript. The 18S-E is cleaved at residue A5,527 (3) to form the mature 18S rRNA.

The 32S transcript is cleaved at position C7,849 (3') to produce the mature 28S and the 12S. The 12S is cleaved to maintain the mature 5.8S rRNA (Mullineux and Lafontaine, 2012).
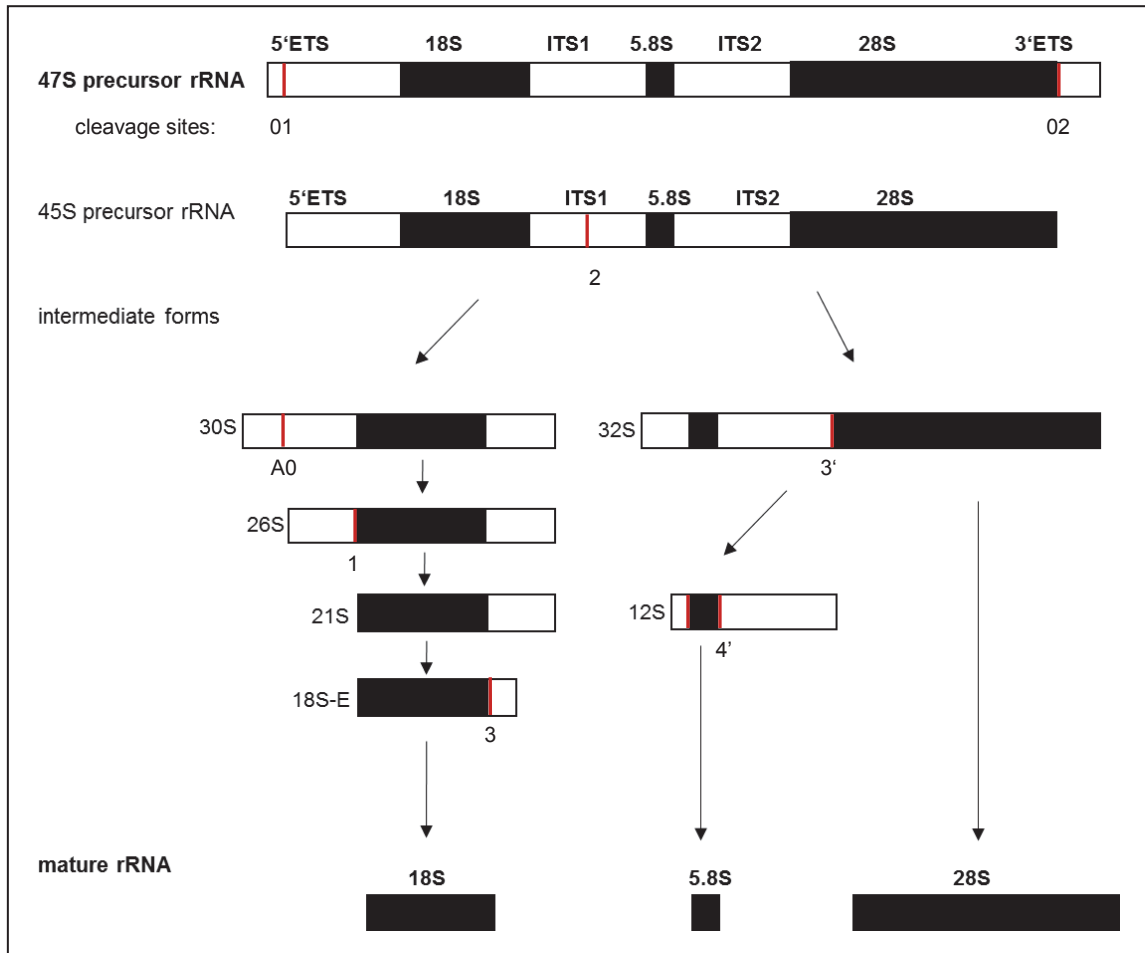
Figure 3: The major pathway of pre ribosomal RNA processing to form the mature rRNAs

Three of the four mature rRNAs (18S, 5.8S, 28S) are encoded in a long 47S rRNA polymerase I transcript, which is cleaved at site 01 and 02 in the 5'ETS and 3'ETS to generate the 45S rRNA. The 45S rRNA is then processed by a minor and a major pathway. This figure illustrates only the major pathway. The cleavage sites are marked in red. The 45S rRNA is directly cleaved in ITS1 at site 2 and generates the 30S and the 32S transcript. The 30S is cleaved at A0 to form the 26S, which is processed at cleavage site 1 to produce the 21S transcript. The 18S-E is cleaved at site 3 to generate the mature 18S rRNA. The 32S transcript is cleaved at site 3' to create the mature 28S and the 12S. The 12S is cleaved to generate the mature 5.8S rRNA. Figure modified after: (Mullineux and Lafontaine, 2012)

In human cells, the first cleavage step of the primary transcript lasts five to eight minutes. The half-life of pre rRNA is very short and takes only some minutes, whereas the half-life of mature rRNA in the cytoplasm are several days (Mullineux and Lafontaine, 2012; Popov et al., 2013; Slomovic et al., 2006; Tafforeau et al., 2013).

1.2.3 Ribosome biogenesis factors

Formation of eukaryotic ribosomes require a large inventory of biogenesis factors such as transcription factors of Pol I that guide the polymerase to the rDNA promoter and facilitate several rounds of transcription, including initiation, promoter escape, elongation and termination. Besides transcription factors and ribosomal proteins, non-ribosomal proteins comprising ribosome biogenesis factors are essential for ribosome synthesis (Turowski and Tollervey, 2015). To mention only a few important factors; Nucleophosmin (NPM1), also known as B23, is a nucleolar phosphoprotein with a histone chaperone activity. It is involved in late steps of pre rRNA processing and plays a role in the cleavage event of the 32S rRNA and in the production of the 5.8S and the 28S (Yanagida et al., 2001). Upon stress, NPM1 translocates to the nucleoplasm (Burger et al., 2013). A variant of NPM1, missing the histone chaperone activity decreases rRNA transcription (Drygin et al., 2010; Murano et al., 2008). Histone chaperones interact with histones, regulate nucleosome assembly and are involved in multiple steps during nucleosome formation (Burgess and Zhang, 2013). NCL, another nucleolar phosphoprotein, also known as C23, is a RNA-binding protein (RBP) that plays a dual role in rDNA chromatin transcription (Durut and Saez-Vasquez, 2015). The nucleolar protein Fibrillarin (FBL) is a core protein of the U3 snoRNP complex and is involved in early as well as late rRNA processing steps (Durut and Saez-Vasquez, 2015; Yanagida et al., 2001). DDX51, a RNA-helicase, regulates early processing events and is involved in snoRNA U8 hybridization. Depletion of DDX51 results in inhibition of the primary 3'ETS processing event. DDX21, another RNA-helicase is involved in early and late pre ribosomal intermediates. It crosslinks to pre rRNA and recruits snoRD56 and snoRD68 (Calo et al., 2015; Sloan et al., 2015).

RNA-helicases are further explained in section 1.2.3.2 and snoRNAs are described in more detail in section 1.2.4. MYC, a nuclear phosphoprotein functions as a

transcription factor and is involved in the chromatin structure regulation, DNA replication and also in ribosome biogenesis (van Riggelen et al., 2010). MYC belongs to the family of mammalian oncogenes (Greasley et al., 2000) and can directly increase synthesis of rRNA and RPs. It binds to the coding region of rDNA and regulates Pol I transcription as well as the transcription of genes that encode proteins that are involved in ribosome biogenesis, like the nucleolar protein 56 (NOP56), block of proliferation 1 (BOP1), FBL, dyskerin (DKC1) and also activates transcription of NCL (Greasley et al., 2000; van Riggelen et al., 2010). DKC1 is the pseudouridine-synthase of the H/ACA-box snoRNPs that are involved in the modification of rRNA. The eukaryotic initiation factor 6 (eIF6) is a component of the pre ribosomal particles and is involved in the formation of the 60S subunit (Brina et al., 2015; Miluzio et al., 2009). Table 1 summarizes the function of several important ribosome biogenesis factors.

Table 1: Ribosome biogenesis factors

Name	Comments and Function	Reference
5'-3' exoribonuclease 2	nuclear 5'-3' single stranded RNA exonuclease, involved in the first cleavage steps	(Woolford and Baserga, 2013)
60S ribosome subunit biogenesis protein NIP7 homolog	involved in the 3' cleavage step, PUA domain	(Woolford and Baserga, 2013)
Fibrillarin	component of snRNP, associated with U3, U8, U13 snoRNAs, 2'O-methyltransferase	(Durut and Saez-Vasquez, 2015; Schlosser et al., 2003; Woolford and Baserga, 2013; Yanagida et al., 2001)
MKI67 FHA domain-interacting nucleolar phosphoprotein	involved in the first cleavage steps, RRM domain	(Woolford and Baserga, 2013)
Myb-binding protein 1A	required for rRNA transcription	(Phipps et al., 2011b; Woolford and Baserga, 2013)
Nucleolar protein 56	component of the C/D-box snoRNPs	(Dupuis-Sandoval et al., 2015; Schlosser et al., 2003)
Nucleolar protein 58	component of the C/D-box snoRNPs	(Dupuis-Sandoval et al., 2015; Schlosser et al., 2003)
Nucleolin	ribosome biogenesis, RNA-binding protein, RRM domain	(Cong et al., 2012; Durut and Saez-Vasquez, 2015; Ghisolfi-Nieto et al., 1996; Roger et al., 2003; Schlosser et al., 2003)
Nucleophosmin	nucleolar phosphoprotein, pre rRNA processing of the 5.8S and 28S, ribosomal protein assembly	(Murano et al., 2008; Schlosser et al., 2003; Yanagida et al., 2001)
pre-rRNA processing protein FTSJ3	methyltransferase, involved in the processing of the 34S pre-rRNA to 18S rRNA and in 40S ribosomal subunit formation	(Morello et al., 2011b)
Ribosome biogenesis protein BOP1	ribosomal RNA processing, involved in the first cleavage steps	(Schlosser et al., 2003)
RNA binding protein 28	involved in the first cleavage steps, RRM domain	(Woolford and Baserga, 2013)
RNA binding protein 34	involved in the first cleavage steps, RRM domain	(Woolford and Baserga, 2013)
Suppressor of SWI4 1 homolog	involved in the first cleavage steps, Brix domain	(Woolford and Baserga, 2013)

Table adapted from: (Woolford and Baserga, 2013)

1.2.3.1 Phosphorylation of ribosome biogenesis factors by protein kinases

Growth factor signaling, nutrient availability and protein kinases are able to control ribosomal RNA transcription. For example, casein kinase II (CKII) and the mammalian target of rapamycin (mTOR) are implicated in rDNA transcription and play a major role in ribosome biogenesis (Drygin et al., 2010; James et al., 2014).

CKII is a serine/threonine protein kinase that is ubiquitously expressed in eukaryotes. It consists of two regulatory β-subunits (CKII 2β), which have a regulatory role in stimulating the catalytic activity of the two α-subunits (α and/or α') that facilitate recognition of the substrates and the catalytic activity. CKII usually exists in a heterotetrameric structure, in a 2α2β, αα'2β or 2α'2β configuration. CKII is located in the nucleolus as well as in the cytoplasm (Duncan and Litchfield, 2008; Li et al., 1996a; Pfaff and Anderer, 1988). Unlike other kinases that are only active in response to stimuli, the catalytic α-subunits of CKII are constitutively active (Pagano et al., 2007).

Overexpression of CKII is highly correlated with aggressive tumorigenesis and upregulated levels of CKII were found in various types of cancer, including leukemia and solid tumors (Ruggero and Pandolfi, 2003). CKII phosphorylates several proteins involved in Pol I transcription, including UBTF, SL-1, NPM1 and NCL (Drygin et al., 2010; Panova et al., 2006; Schneider and Issinger, 1988). UBTF is phosphorylated in the acidic region, in the C-terminal domain, which stabilizes binding of SL-1 to rDNA. NCL is phosphorylated in the acidic region of the N-terminal domain. CKII regulates Pol I transcription at multiple steps, including the assembly of the initiation complex, promoter escape, elongation and reinitiation of the Pol I complex. Consequently, inhibition of CKII decreases rDNA transcription, which results in cell cycle arrest, apoptosis and also in nucleolar disruption (Drygin et al., 2010). CKII is one of the only kinases that is able to use GTP and ATP as a substrate for the phosphotransferase kinase reaction (Fritz et al., 2009; Jakobi and Traugh, 1995; Leiva et al., 1987; Medina et al., 2010; Pfaff and Anderer, 1988). Several small molecules like

tetrabromobenzotriazole (TBB) and tetrabromocinnamic (TBCA) as well as heparin are able to inhibit the phosphorylation reaction of CKII (Fritz et al., 2009; Hathaway et al., 1980; Pagano et al., 2007; Sarno et al., 2001; Zien et al., 2005).

The mammalian target of rapamycin is also a serine/threonine protein kinase, which is involved in cell proliferation, cell growth, protein synthesis and also regulates Pol I transcription. The mTOR pathway is characterized as a contributing factor in several types of cancer. Treatment with rapamycin inhibits the mTOR pathway and results in a reduction of the nuclear size, inhibits rDNA transcription and leads to an increase of pseudouridylation of the 28S, performed by H/ACA-box snoRNAs. This indicates that the mTOR pathway is also involved in the regulation of rRNA modification (Drygin et al., 2010; Dupuis-Sandoval et al., 2015; James et al., 2014).

1.2.3.2 RNA-helicases involved in ribosome biogenesis

RNA-helicases use the energy of ATP hydrolysis to unwind RNA duplex structures and play a key role in all major pathways of RNA metabolism, such as ribosome biogenesis (Leary and Huang, 2001) (Meister, 2011). Table 2 summarizes important DEAD RNA-helicases that are involved in ribosome biogenesis. Proteomic analysis of the nucleolus identified over 30 putative RNA-helicases (Martin et al., 2013), several of these were identified to be involved in ribosome biogenesis. For example, DDX10, DDX18, DHX15, DHX37 are involved in the biogenesis of the SSU, whereas DDX25 and DDX56 are associated with the LSU. DHX33 regulates rRNA transcription and recruits Pol I by interaction with UBTF. DDX21 which is also known as RNA-helicase II is involved in early and late pre ribosomal intermediates, crosslinks to pre rRNA and recruits snoRNAs (Calo et al., 2015; Sloan et al., 2015). Since knockdown of DDX21 results in decreased levels of 18S and 28S rRNA, it is assumed that the RNA-helicase II is required for ribosome biogenesis (Martin et al., 2013). It was also proposed that

RNA-helicases are required for the release of snoRNAs from pre rRNA and specific snoRNPs from the pre ribosomes (Martin et al., 2013).

Table 2: RNA-helicases involved in ribosome biogenesis
The table lists DEAD RNA-helicases with the corresponding yeast homolog, their localization and function.

Name	Yeast homolog	Localization	Function	Reference
DDX3X	Dbp1	Nucleus, Cytoplasm	rRNA transcription	(Boisvert et al., 2012; Yedavalli et al., 2004; Zhang et al., 2011)
DDX5	Dbp2	Nucleolus	rRNA transcription, 5.8S maturation, U8 snoRNA release during pre 32S cleavage	(Boisvert et al., 2012; Iggo et al., 1991; Jalal et al., 2007; Saporita et al., 2011)
DDX10	Dbp4	Nucleolus	rRNA transcription, component of the SSU processome, SSU biogenesis	(Boisvert et al., 2012; Turner et al., 2009; Wild et al., 2010; Zhang et al., 2011)
DDX17	Dbp2	Nucleolus	rRNA transcription, 5.8S maturation, U8 snoRNA release during pre 32S cleavage	(Boisvert et al., 2012; Zhang et al., 2011)
DDX18	Has1	Nucleolus	rRNA transcription, SSU biogenesis	(Boisvert et al., 2012; Turner et al., 2009; Wild et al., 2010; Zhang et al., 2011)
DDX21	-	Nucleolus	28S rRNA stabilization, assembly of ribosomal proteins, 18S, 28S production	(Boisvert et al., 2012)
DDX23	Prp28	Nucleus	rRNA transcription	(Boisvert et al., 2012; Zhang et al., 2011)
DDX24	Mak5	Nucleolus	LSU biogenesis	(Boisvert et al., 2012; Wild et al., 2010)
DDX46	Prp5	Nucleus	rRNA transcription	(Boisvert et al., 2012; Zhang et al., 2011)
DDX47	Rrp3	Nucleolus	rRNA transcription, 18S and 28S maturation	(Boisvert et al., 2012; Zhang et al., 2011)
DDX50	-	Nucleolus	rRNA transcription, 18S and 28S maturation (antagonist of DDX21)	(Boisvert et al., 2012; Zhang et al., 2011)
DDX51	Dbp6	Nucleus, Nucleolus	rRNA transcription, 3' end maturation of 28S rRNA, release of U8 snoRNA from pre rRNA	(Boisvert et al., 2012; Zhang et al., 2011)
DDX56	Dbp9	Nucleus, Nucleolus	rRNA transcription, LSU biogenesis	(Wild et al., 2010; Zhang et al., 2011)
DHX33	-	Nucleolus	rRNA transcription; facilitating conformational change of rDNA	(Boisvert et al., 2012; Zhang et al., 2011)

Table adapted from: (Martin et al., 2013).

1.2.4 Structure and function of small nucleolar RNAs

Small nucleolar RNAs (snoRNAs) belong to a large class of small, regulatory, non-coding RNAs (ncRNAs) localized within the nucleolus. SnoRNAs are components of the small nucleolar RNP complex that are required for modification and processing of the pre rRNA during ribosome biogenesis (Dupuis-Sandoval et al., 2015; Esteller, 2011; Lafontaine, 2015). Every pre rRNA is transiently packed with 75 different snoRNAs like snoRNA U3, U14 and U17 that are involved in the early processing steps and are required for the first cleavage sites (Turowski and Tollervey, 2015). Ribosomal RNAs undergo several co- and posttranscriptional modifications that are essential in eukaryotes and responsible for cell growth, RNA stability, secondary structure formation and activity of the ribosomes (Granneman and Baserga, 2004; Henras et al., 2008; Sharma and Lafontaine, 2015). Modifications of rRNA during ribosome biogenesis are therefore very important and include base methylation, acetylation, amino-carboxypropylation, methylation of the sugar backbone and pseudouridylation by formation of the uridine isomers (Lafontaine, 2015). SnoRNAs guide sequence-specific 2'-O-methylation (-CH3) or pseudouridylation of rRNA. Pseudouridylation is a modification process of a site-specific conversion of a uridine to a pseudouridine (Ψ). A single uridine is exposed to the enzymatic activity of the pseudouridine synthase dyskerin that facilitates the modification of the specific residue (Dupuis-Sandoval et al., 2015; Esteller, 2011; Ruggero and Pandolfi, 2003).

According to their motifs, structure, and chemical modification reactions, snoRNAs are divided in two classes; the C/D box (Figure 4) and the H/ACA box (Figure 5) snoRNAs (Dupuis-Sandoval et al., 2015; Esteller, 2011; Lafontaine, 2015). The C/D box snoRNAs are typically 60 nt to 90 nt in length and have conserved C and D boxes with consensus sequences of UGAUGA and CUGA, respectively. Additional C' and D' boxes are less conserved, but contain the same consensus sequence as the C and D

box. The guide sequence is complementary to a target sequence that is downstream of the C/C' boxes. The guide region binds with the fifth residue upstream of the D/D' box motif (Dupuis-Sandoval et al., 2015). The snoRNP contains four core proteins, including NOP56, NOP58, NHP2-like and FBL, which catalyze the 2'-O-ribose methylation (Dupuis-Sandoval et al., 2015; Esteller, 2011). SnoRNA U3 and U8 belong to the C/D box snoRNAs. They function in the endonucleolytic cleavage steps and are required for folding the pre rRNA during ribosome biogenesis. Both snoRNAs are transcribed by Pol II and are encoded in their individual independent transcription unit (Dupuis-Sandoval et al., 2015). U3, one of the most abundant snoRNA, is associated with the 5'ETS and with early processing steps in ribosome biogenesis (Ginisty et al., 1998). First cleavage steps in the 5'ETS and 3'ETS of the pre rRNA, require snoRNAs U3, U14, U17 and E3. Depletion of these snoRNAs decreases 5'ETS processing *in vitro* (Ginisty et al., 1998).
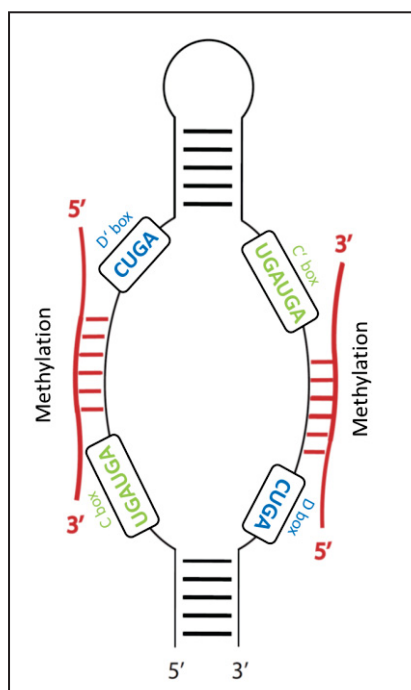
Figure 4: Structure of the C/D box snoRNA

C/D box snoRNAs contain C/C' boxes with a consensus sequence of UGAUGA (green) and D/D' boxes with a consensus sequence of CUGA (blue). The guide sequence is complimentary to the target sequence (red), which is upstream of the D/D' boxes. The snoRNP contains four core proteins, including FBL, NOP56, NOP58, and NHP2-like, which catalyze the 2'-O-ribose methylation. Figure modified after: (Gardner et al., 2010).

The small nucleolar H/ACA box RNAs are 120 nt to 140 nt in length (Figure 5). Their secondary structure implicates two hairpins containing a H and ACA box that are both located in the hinge region of the hairpin. The H box with a common ANANNA motif is located in the first hinge, the ACA box with a conserved ACANNN motif (N represents any nucleotide) is located in the second hinge of the hairpin, three nt upstream of the 3' end of the RNA (Dupuis-Sandoval et al., 2015). The two guide sequences with complementary sequences to the target site are located in the middle part of the hairpin structure and catalyze the pseudouridylation reaction. In general, the modified uridine is found 14 nt to 15 nt upstream of the H or ACA box. The snoRNPs of H/ACA snoRNAs consist of dyskerin, GAR1, NHP2 and NOP10 (Dupuis-Sandoval et al., 2015; Esteller, 2011; Mannoor et al., 2012).
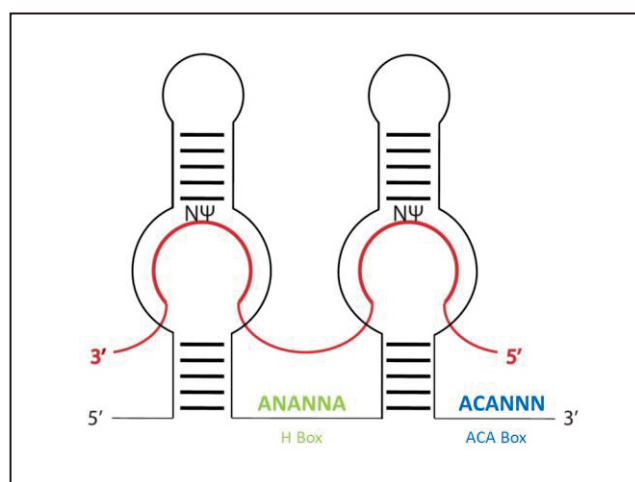
Figure 5: Structure of the H/ACA snoRNA

H/ACA snoRNAs consist of two hairpins and contain a H box (green) and an ACA box (blue) located in the hinge region of the hairpin. N stands for any nucleotide. The guide sequence marked in red is located in the middle of the hairpin structure. The snoRNPs of H/ACA snoRNAs consist of dyskerin, GAR1, NHP2 and NOP10. Figure modified after: (Gardner et al., 2010).

SnoRNAs are involved in posttranslational gene silencing, e.g. the HBII-52, a human snoRD (C/D box snoRNA) gene plays a role in mRNA splicing of the secretion receptor 2C (Kishore and Stamm, 2006). In humans, up to 98% of snoRNAs are encoded in introns of a host gene, which can carry several snoRNAs in different introns, like NCL. Consequently, the expression of snoRNAs strongly depends on their host gene.

Intronic snoRNAs are formed by splicing, debranching of the intron lariat and exonucleolytic degradation of the 5' and 3' sequences adjacent to the snoRNA (Dupuis-Sandoval et al., 2015; Esteller, 2011). SnoRNA U82, U20 (also known as snoRD20) and U23 are respectively encoded in intron 5, 11 and 12 of the NCL gene (Ginisty et al., 1999; Nicoloso et al., 1994; Rebane and Metspalu, 1999). SnoRNA U82 and U20 contain respectively a 11 nt and 21 nt complementarity to the 18S rRNA. These snoRNAs function in the methylation of the human 18S rRNA and are therefore involved in the processing and modification of the small ribosomal subunit (SSU). Usually, host genes of snoRNAs have related functions and are mainly involved in ribosome biogenesis (Rebane and Metspalu, 1999).

1.2.5 The role of ribosome biogenesis in cancer and genetic diseases

In general, loss-of function mutations in ribosome biogenesis factors or ribosomal proteins are lethal in model organisms and presumably embryonic lethal in humans when they occur homozygous (de la Cruz et al., 2015; Ruggero and Pandolfi, 2003). However, altered expression of RPs or mutation in genes that encode proteins involved in ribosome biogenesis can cause defects in rRNA transcription and are found in various types of cancer and genetic diseases, like ribosomopathies. Ribosomopathies like the Diamond-Blackfan anaemia or the Shwachman-Diamond syndrome show severe effects in the bone marrow and skin (De Keersmaecker et al., 2015; Gerstberger et al., 2014b; Ruggero and Pandolfi, 2003).

Over a century ago, pathologists revealed a strong correlation between enlarged and abnormal nucleoli in cancer cells. Changes in size of the nucleolus can contribute to the prediction of the clinical outcome. Enlarged nucleoli are associated with abnormal high expression of ribosome biogenesis and with poor clinical prognosis (Quin et al., 2014; Su et al., 2014). Recent studies showed that alteration in morphology is related

to hyperactivation of rDNA transcription and overexpression of ribosome biogenesis linked to altered levels of rDNA transcription is a common feature in cancer cells. Mutation in genes encoding Pol I components and ribosomal transcription factors are found in genetic diseases causing mental retardation and growth defects (Hannan et al., 2013). Williamson et al. proposed that changes in pre rRNA synthesis are one of the most important alterations in tumor cells. Overexpression of pre rRNA is a general phenomenon in cancer and correlates with poor prognosis e.g. in a type of alveolar rhabdomyosarcoma (Drygin et al., 2010; Williamson et al., 2006).

High expression of certain ribosomal proteins is associated with cancer and increased cell proliferation and cell growth. RPL19 is overexpressed in prostate cancer, RPS6 is upregulated in Non-Hodgkin lymphoma, whereas depletion of RPL16 leads to a decrease of the large ribosomal subunit. This results in downregulation of the polyribosomes and consequently leads to cellular growth defects (Ruggero and Pandolfi, 2003; Shenoy et al., 2012). Table 3 summarizes several RPs, which show altered expression in cancer (van Riggelen et al., 2010).

Table 3: Altered expression of ribosomal proteins in cancer

Ribosomal protein	Cancer
RPL15	Oesophageal
RPL13	Gastrointestinal
RPL13, RPL36a, RPS8, RPL12, RPL23a, RPL27, RPL30	Hepatocellular
RPS3, RPS6, RPS8, RPS12, RPL5, RPL22, RPL26, RPL35	Colorectal
RPL37, RPL7a, RPL19	Prostate
RPS12	Cervical
RPS6	Non-Hodgkin lymphoma

Table adapted and modified after: (van Riggelen et al., 2010).

The Diamond-Blackfan anemia is directly related to alterations in ribosome biogenesis and is linked to mutations in RPS19 and RPS24. These mutations disrupt the

maturation of the 18S rRNA and result in decreased expression of the small ribosomal subunit. Diamond-Blackfan anemia shows increased susceptibility to hematological malignancies and affects also the production of red blood cells (Gerstberger et al., 2014a; Ruggero and Pandolfi, 2003; Shenoy et al., 2012). Whereas, the Shwachman-Diamond and the Bowen-Conradi syndrome affect the production of white blood cells and are caused by mutations in ribosome biogenesis factors like SBDS (Shwachman-Bodian-Diamond Syndrome gene) and EMG1 (Essential for Mitotic Growth 1) (Armistead et al., 2014), respectively.

It was also proposed that certain snoRNAs show altered expression in breast and prostate cancer (Berger et al., 2015; Su et al., 2014). Knockdown of the core protein components of the C/D box snoRNP complex, like NOP56, NOP58 and FBL result in decreased levels of snoRNAs (Su et al., 2014). Depletion of snoRNA U50 is associated with the development of prostate and breast cancer (Dong et al., 2009; Lafontaine, 2015; Liao et al., 2010; Mannoor et al., 2012). H/ACA snoRNA42 is highly expressed in Non-small cell lung cancer (NSCLC) and implicated to play a role in this type of cancer (Mannoor et al., 2012; Su et al., 2014). Elevated expression of snoRNAs were also identified in meningioma. Table 4 summarizes snoRNAs that function as an oncogene and are increased in NSCLC and breast cancer.

Table 4: SnoRNAs involved in various types of cancer

snoRNA	Class	Cancer
snoRD33	C/D box	NSCLC
snoRD66	C/D box	NSCLC
snoRD76	C/D box	NSCLC
snoRA42	H/ACA box	NSCLC
snoRD44	C/D box	Breast cancer

SnoRNAs that function as oncogenes and are increased in cancer are listed according to their class and cancer type. Table adapted from: (Mannoor et al., 2012).

Within the last few years, the nucleolus and the transcription machinery of Pol I became interesting targets in cancer therapy (Drygin et al., 2010; Hein et al., 2013). Until now, there are several anti-cancer drugs that target Pol I transcription including Actinomycin-D (Act-D), Cisplatin, Mitomycin C, CX-3543 and CX-5461. Changes in size of the nucleolus might also be related to the response of chemotherapeutic drugs and small molecule inhibitors. Act-D has an impact on the structure of the nucleus, the integrity of the nucleolus and localization patterns of several nucleolar proteins (Andersen et al., 2002). Low concentrations of Act-D intercalate in G/C rich regions of the rDNA and can inhibit elongation of Pol I. Cisplatin inhibits Pol I transcription and introduces crosslinks to DNA with a high affinity for HMG containing proteins. CX-3543, also known as quarfloxin, is a clinical phase I approved drug that targets Pol I. It binds to G-rich quadruplex region in the rDNA and blocks Pol I elongation by disrupting NCL G-quadruplexes, which leads to translocalization of NCL to the nucleoplasm. Inhibition of elongation and subnucleolar translocation of NCL to the nucleoplasm can lead to apoptotic cell death in cancer cells. The small molecule, CX-5461 inhibits Pol I transcription at a very early stage of initiation by disrupting the binding of SL-1 to the rDNA promoter. CX-5461 activates expression of the tumor suppressor p53 and induces tumor-specific apoptosis in malignant B cells. In normal B cells, p53 is not activated and apoptosis is therefore not induced (Berger et al., 2015; Drygin et al., 2010; Drygin et al., 2009; Hein et al., 2013; Pickard and Bierbach, 2013; Quin et al., 2014). The effects of the small molecules on ribosome biogenesis and on the nucleolus are summarized in Table 5.

Table 5: Small molecule inhibitors of RNA Polymerase I

Small molecule	Mechanism	Effect on ribosome biogenesis	Effect on the nucleolus
Actinomycin D	Intercalates into GC-rich duplex DNA in the rDNA repeats	Low concentration Inhibit Pol I transcription	Nucleolar disintegration
Cisplatin	DNA crosslinking via alkylating DNA bases	Inhibition of Pol I transcription	Nucleolar disintegration
CX-3543 (quarfloxin)	Disruption of the Nucleolin - rDNA G-quadruplex complex	Selective inhibition of Pol I transcription at the stage of elongation	Translocation of non-rDNA bound Nucleolin to the nucleoplasma, causing apoptosis in cancer cells, no effect on fibrillarin
CX-5461	Inhibits SL-1 preinitiation complex formation at the rDNA promoter	Selective inhibition of Pol I transcription at the stage of initiation	Nucleolar disintegration
Mitomycin C	DNA crosslinking via alkylating 5-CpG-3 guansoine	Inhibition of Pol I transcription	Nucleolar disintegration

The table summarizes small molecules, which show an inhibitory effect on ribosome biogenesis.
Table: (Hein et al., 2013).

1.3 RNA-binding proteins

The human genome encodes 1,542 RNA-binding proteins (RBPs), which implicates 7.5% of all protein-coding genes (Gerstberger et al., 2014b). RBPs interact directly with all classes of RNAs including precursor and mature RNA, coding as well as non-coding RNAs and are able to form RNP complexes. These complexes are involved in posttranscriptional gene regulation (PTGR) including RNA maturation, transport, stability and translation.

At first, RNA was supposed to serve either as a template, in form of mRNAs or as a structural component during protein synthesis in form of tRNAs or rRNAs. Subsequently, with the discovery of catalytic RNAs and several ncRNAs, it was demonstrated that RNAs are involved in many regulatory functions within a cell. RNA can functions as a guide RNA, which recognizes RNA sequence motifs or as a scaffold that is able to recruit proteins (Cech and Steitz, 2014). RNA can be divided by their function in coding and ncRNAs. The most abundant RNAs, comprising 90% of cellular RNAs according to their copy number include mRNA as coding RNAs, and rRNAs and

tRNAs, which belong to ncRNAs. Messenger RNAs are essential for protein synthesis and encode the information for protein coding genes that are translated by the ribosomes (Dreyfuss et al., 2002), whereas tRNAs function as a RNA adaptor and recognize specific triplet codons on the mRNA (Simos and Hurt, 1999). The remaining 10% represents ncRNAs, which either serve as a guide or a molecular scaffold in processes like RNA modification, splicing and silencing. This class comprises lncRNA (long non-coding RNA), snRNAs (small nuclear RNA), including U1, U2, U,4, U5, U6, which are the major components of the spliceosome (Kiss, 2004), snoRNAs that guide chemical modifications including methylation and pseudouridylation of rRNAs (Dupuis-Sandoval et al., 2015), miRNA (microRNA), which are associated with AGO proteins and induce degradation and translational silencing by target sequences mainly in the 3'UTRs of mRNAs (Bartel, 2009), siRNA (small interfering RNAs), piRNA (piwi associated RNAs) and scaRNA (small cajal body specific RNAs) (Figure 6).
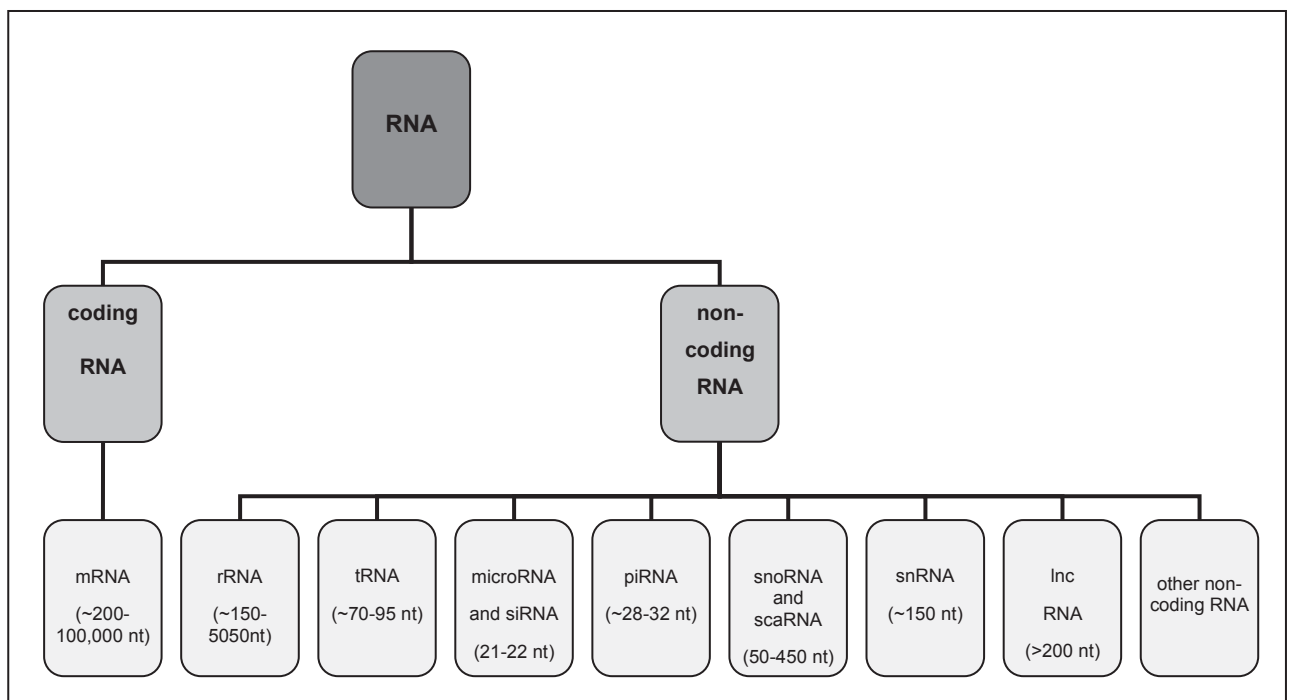
Figure 6: Coding and non-coding RNAs

RNAs are divided into two groups; coding RNA representing mRNAs and non-conding RNAs, which include rRNA, tRNA, microRNA, siRNA, piRNA snoRNA, scaRNA, snRNA, lncRNA and other ncRNA.

Figure modified after: (Ronnau et al., 2014).

The majority of RBPs is involved in protein synthesis, whereas specific biological functions of one-third of the identified RBPs remain unknown (Gerstberger et al., 2014b). Out of the 1,542 RBPs, 692 RBPs are involved in mRNA-binding, 169 are ribosomal proteins, 153 are implicated in tRNA biogenesis, 163 are associated with ribosome biogenesis comprising 122 pre rRNA and 41 snoRNAs targets, 122 are involved in other ncRNA-binding, 90 in small nuclear RNA-binding and 153 contain unknown and diverse RNA targets (Figure 7).
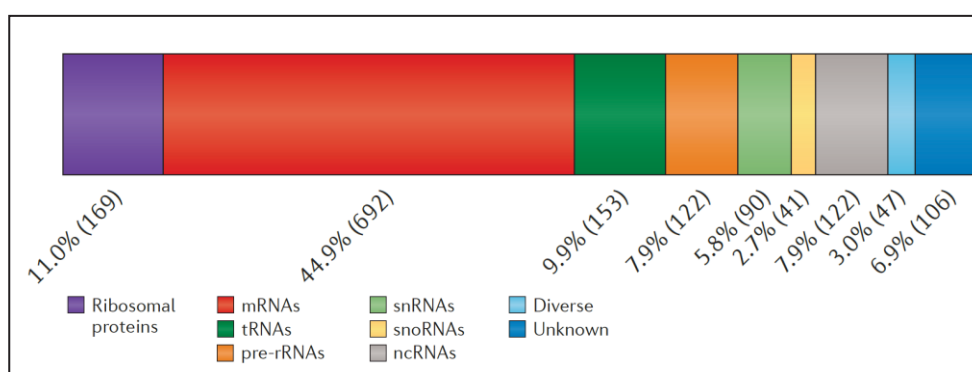
Figure 7: RNA targets of RNA-binding proteins

The 1,542 RNA-binding proteins are grouped according to their RNA targets clustered in ribosomal proteins, mRNAs, tRNAs, pre rRNAs, snRNAs, snoRNAs, ncRNAs, diverse and unknown. Figure: (Gerstberger et al., 2014b).

1.3.1 RNA-binding domains

RBPs are generally characterized and grouped by their specific RNA-binding domains that are responsible for RNA recognition (Table 6); the RRM (RNA recognition motif), KH (K-homology), ZnF (Zink finger), PAZ, PIWI, Pumilio and S1 (Ascano et al., 2012; Gerstberger et al., 2014b; Lunde et al., 2007).

Table 6: RNA-binding domains

Domain	RNA-protein interaction and domain description	Reference
CSD	cold-shock domain, ssRNA/ssDNA binding, RBD functions as a RNA chaperone in bacteria, involved in regulation of translation in eukaryotes	(Ascano et al., 2012)
DEAD/DEAH	DEAD/DEAH box helicase motif, unwinds RNA and DNA	(Glisovic et al., 2008; Rocak and Linder, 2004)
Dsrm	double-stranded RNA binding motif, found in dsRNA-dependent protein kinase, protein kinase R (PKR), RNA-helicases, Drosophila staufen protein, *E.coli* RNase III, RNase H1, dsRNA dependent adenosine deaminases	(Ascano et al., 2012; Lunde et al., 2007)
KH	two different K homology domains, share a single KH motif, KH recognizes four nucleotides of ssRNA through hydrophobic interactions	(Lunde et al., 2007)
LSM	the like SM (LSM) domain contains sm proteins, found in snRNP complexes, binds A/U rich regions	(Ascano et al., 2012)
PAZ	Piwi Argonaut and Zwille (PAZ) domain, posttranscriptional silencing domain, recognizes two-base 3'overhangs of siRNA	(Ascano et al., 2012; Lunde et al., 2007; Meister, 2013)
Piwi	Piwi domain (P-element induced wimpy testis), found in the protein piwi, dsRNA guided hydrolysis of ssRNA recognizes the defining 5' phosphate group in the siRNA guide, posttranscriptional silencing domain	(Ascano et al., 2012; Lunde et al., 2007; Meister, 2013)
PUF	tandem repeat of eight domains, ssRNA-binding domain	(Zamore et al., 1997)
RRM	found in a variety of RBPs, in proteins implicated in regulation of alternative splicing and in components of snRNPs, interacts with four nucleotides of ssRNA through stacking, electrostatics and hydrogen bonding	(Ascano et al., 2012; Gerstberger et al., 2014b; Lunde et al., 2007)
S1, S1-like	S1 ssRNA-binding domain, found in ribosomal proteins, similar to cold-shock domain	(Lunde et al., 2007)
SAM	shape-dependent recognition of RNA stem-loop, interactions with the sugar-phosphate backbone and single base in the loop	(Lunde et al., 2007)
SAP	SAP (SAF-A/B, Acinus and PIAS) motif, putative DNA/RNA-binding domain, found in nuclear and cytoplasmic proteins	(Ascano et al., 2012)
zf-C2H2_jaz	JAZ dsRNA-binding protein zinc-fingers, dsRNA-binding	(Krishna et al., 2003; Lunde et al., 2007)
zf-CCCH	Zinc finger are found in many RBP, Zinc finger motif type C-x8-C-x5-C-x3-H, stacking interactions between aromatic residues and bases creates a kink in the RNA that enables direct recognition of Watson-Crick edges of the bases by protein backbone	(Ascano et al., 2012; Lunde et al., 2007)
zf-CCHC	Zinc knuckle, C-x2-C-x4-H-x4-C, ssRNA-binding	(Krishna et al., 2003; Lunde et al., 2007)

Table adapted and modified from: (Ascano et al., 2012; Lunde et al., 2007)

Proteins contain either multiple copies of the same domain or copies of different domains to expand their functional binding repertoire. Multiple RNA-binding domains within the protein recognize a longer stretch of nucleic acids. A single RRM or KH domain can interact with four to eight nucleotides of a single stranded (ss) RNA (Lunde et al., 2007). The RRM is the most abundant and common RNA-binding domain followed by the KH domain and the zinc-finger domain within RBPs. The RRM is comprised of 80 to 90 amino acids (aa) that form a βαββαβ topology with a four-stranded anti-parallel β-sheets with two helices. Interaction of the RNA takes place on the surface of the β-sheet and is mediated by three conserved arginine (R) or lysine (K) residues. Approximately 1% of all human genes contain a RRM domain arranged in multiple copies in the same polypeptide (Gerstberger et al., 2014b; Lunde et al., 2007). RBPs like NCL, TIAR (TIA-1 related protein), several hnRNPs (heterogeneous nuclear ribonucleoproteins), PABP (poly (A) binding protein), and HuR, also known as ELAVL1 contain multiple copies of RRM domains. Common RNA targets of HuR are mRNAs and snoRNAs (Eberhardt et al., 2012; Soeno et al., 2010). Several other RBPs like PACP, Vigilin, Pumilio, Dicer and Staufen are illustrated with their RBDs in Figure 8.
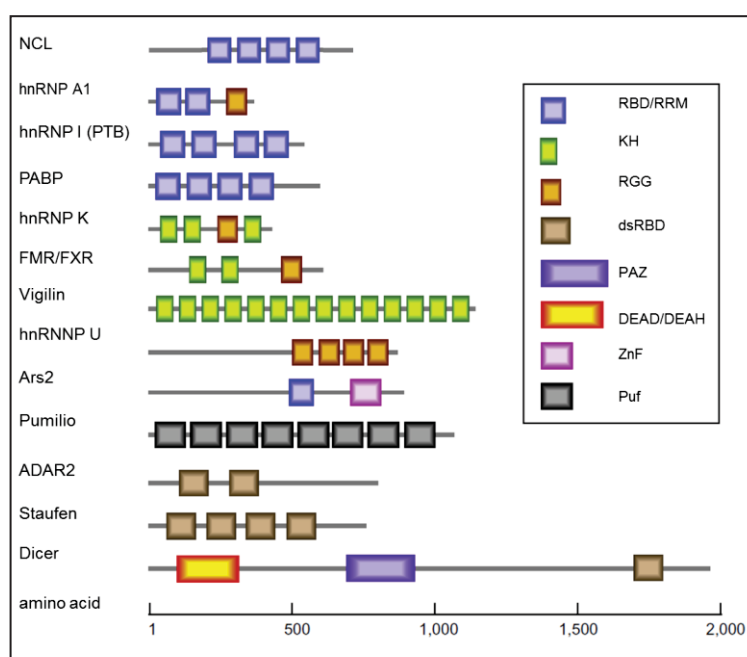
Figure 8: RNA-binding domains of RNA-binding proteins

RNA-binding proteins contain different RNA-binding domains like RNA-binding domain (RBD) or RNA-recognition motif (RRM), K-homology (KH), RGG (Arg-Gly-Gly) domain, double stranded RNA-binding domain (dsRBD), Piwi/Argonaut/Zwille (PAZ) domain, RNA-helicases with DEAD/DEAH box, RNA-binding zinc finger (ZnF) and Puf RNA-binding repeats (Puf). NCL, hnRNP I and PABP contain 4 RBD/RRM. Figure modified after: (Glisovic et al., 2008)

1.3.2 RNA-binding proteins in human diseases

RNA-binding proteins are key players in RNA metabolism and are highly conserved between bacteria, archaea and eukaryotes. Immortalized cell lines and tumor tissues contain higher expression levels of mRNA-binding proteins and ribosomal proteins than normal tissues. This implicates high requirement of protein production and illustrates one major hallmark of cancer. RBPs or RNP complexes function as posttranscriptional gene regulators and altered expression levels, including loss- or gain- of function are associated with human diseases ranging from neurodegenerative disorders to cancer (Lukong et al., 2008). Table 7 summarizes several RBPs that are associated with human diseases.

Table 7: RNA-binding proteins implicated in human diseases

RNA-binding protein	Disease	Reference
FXS, FXTAS, PEM/SN, FMRP1, POMA, hnRNP-P2, ELAV/HuB/C/D, NOVA1/2, NOP56, QKI	Neurological disorders and neurodegenerative disease	(Keene, 2007; Kobayashi et al., 2011; Lukong et al., 2008; Zhou et al., 2014)
FMR1	Mental retardation	(Keene, 2007; Wang et al., 2012)
FMR1, RPL10	Autism	(Khandjian, 1999; Klauck et al., 2006; Wang et al., 2012)
Ribosomal proteins	Turner Syndrome	(Keene, 2007)
Poly-A binding protein 2	Muscular dystrophy	(Keene, 2007)
SMN1/2, OPMD, SMA, DM, PABPN1	Muscular atrophy	(Keene, 2007; Lukong et al., 2008)
CUG-BP/EDEN, CELF3/4/5	Myotonic dystrophy	(Keene, 2007)
ELAV/Hu, hnRNP-L1, αCP1/2, ETR3	Cardiovascular disease	(Keene, 2007)
SF2/ASF, eIF4E, Sam68, QKI, TET, EWS, TLS/FUS, DICER1	Cancer	(Foulkes et al., 2014; Lukong et al., 2008)
TPP, TIA, TIAR, HuR	Immunoregulatory disorders	(Keene, 2007)
HuR, ELAV/HuB/C/D, NOVA-1/2	Autoimmunity disorders	(Lukong et al., 2008)
Ribosomal proteins, SBDS	Diamond-Blackfan anaemia, Shwachman-Diamond syndrome	(De Keersmaecker et al., 2015; Gerstberger et al., 2014b; Ruggero and Pandolfi, 2003)
EIF4G1/G2/G3, PARK7	Parkinson disease	(Chartier-Harlin et al., 2011; Gasser, 2004)
FUS, EWSR1, TAF15	Amyotrophic Lateral Sclerosis	(Coady and Manley, 2015)

Table modified and adapted from: (Keene, 2007; Lukong et al., 2008)

Defects in RPs and ribosome biogenesis factors can cause ribosomopathies (section 1.2.5), whereas mutations in mRBPs are mostly implicated in neuropathological diseases by dysregulation of splicing. Mutations in mRBPs like FUS and hnRNP A2B1/A1 can cause amyotrophic lateral sclerosis (ALS) (Ascano et al., 2013; Gerstberger et al., 2014b), loss-of function of FMR1 (fragile mental retardation 1) leads to mental retardation and autism (Keene, 2007; Khandjian, 1999; Wang et al., 2012) (Table 7).

1.3.3 Identification of RNA-protein interaction

Over the last few years, several different methods have been developed to identify targets of RBPs. Methods like systematic evolution of ligands by exponential enrichment (SELEX) were developed in the 1990s by Ellington and Szostak (Ellington and Szostak, 1990). SELEX is a low-throughput method for the identification and characterization of sequence-specific binding preferences of RBPs *in vitro*. High-affinity binding targets are selected from a randomized RNA oligonucleotide pool by several sequential rounds of interaction to the purified RBP, each followed by PCR amplification. The identified targets are then cloned and subsequently sequenced in to order to identify a set of short sequences that are recognized by the RBP. These short sequences are analyzed to define sequence and structural preferences of the RBP (Li et al., 2014). RNA immunoprecipitation (RIP) experiments like RIP-Chip (RNA IP followed by microarray analysis) or RIP-Seq (RNA IP followed by sequencing) (Keene et al., 2006; Morris et al., 2010; Townley-Tilson et al., 2006; Zhao et al., 2010), methods using either short-wavelength (254 nm) or long-wavelength (365 nm) UV-crosslinking and immunoprecipitation (CLIP) linked to high-throughput deep sequencing, like HiTS-CLIP (high throughput sequencing CLIP), iCLIP (individual-nucleotide resolution CLIP) and PAR-CLIP (photoactivatable-ribonucleoside enhanced

CLIP), respectively, were developed to identify transcriptome-wide binding sites of RBPs (Darnell, 2010; Hafner et al., 2010a; Konig et al., 2011; Sutandy et al., 2016; Ule et al., 2005). The different CLIP methods are illustrated in Figure 9.
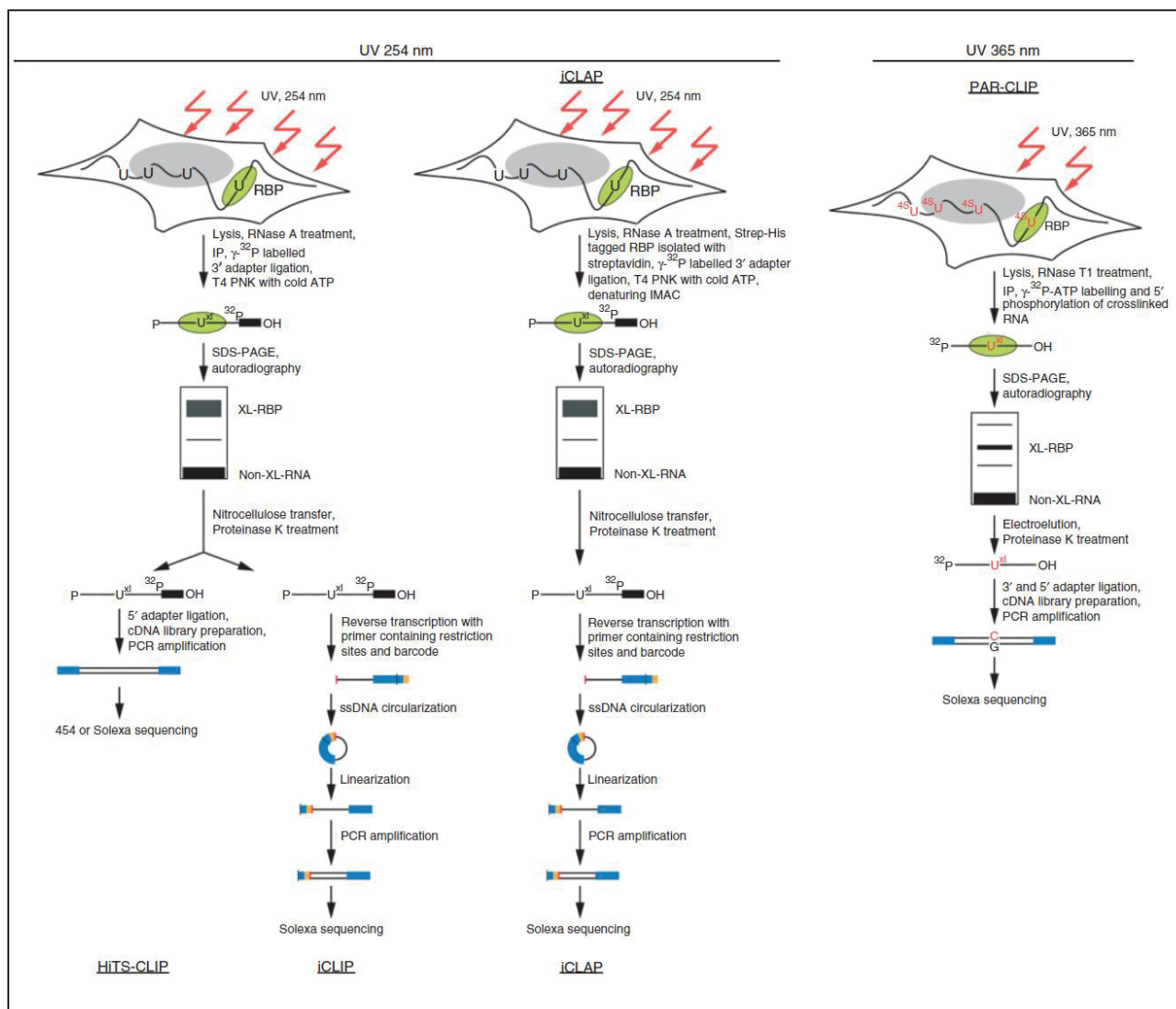
Figure 9: HiTS-CLIP, iCLIP, iCLAP and PAR-CLIP

Overview of crosslinking and immunoprecipitation methods in order to identify RNA-protein interactions. High throughput sequencing CLIP, individual nucleotide resolution CLIP and individual nucleotide resolution crosslinking affinity purification use short-wavelength ultraviolet light of 254 nm (left panel), whereas photoactivatable ribonucleoside-enhanced CLIP utilizes long-wavelength UV of 365 nm (right panel). Figure: (Ascano et al., 2012).

Photo-crosslinking and non-crosslinking methods that are used to discover RNA-binding sites are summarized and described in Table 8.

Table 8: Methods for the identification of RNA-protein interaction and cellular RNA targets

Method	Aim of the method	Description
RIP-Seq	target gene identification	no UV-crosslinking, IP of epitope tagged RBP or endogenous RBP to isolate associated RNA, RNA is quantified by RNA-Seq, targets are defined by calculating enrichment scores of the IP samples versus a control
HiTS-CLIP	identification of target site, definition of RNA recognition element	254 nm UV-crosslinking of RNA to the RBP, crosslinked RNA is digested by RNAse and RNA-protein complexes are fractionated by SDS-PAGE, isolation of RNA, ligation of a 3'- and 5'-adapter for RT and PCR, cDNA is deep sequenced, overlapping sequence reads are clustered and targets are defined based on enrichment according to a negative control
iCLIP	identification of target site, definition of RNA recognition element	254 nm UV-crosslinking of RNA to the RBP, crosslinked RNA is digested by RNAses and RNA-protein complexes are fractionated by SDS-PAGE, crosslinked RNA is isolated, RT with primer containing restriction sites and barcodes, cDNA that is generated from RT is circularized and deep sequenced
iCLAP	identification of target site, definition of RNA recognition element	254 nm UV-crosslinking of RNA, the RBP is tagged with a strep- and polyhistidine epitope, streptavidine beads are used in the first purification step, immobilized metal-ion affinity chromatography is performed in a second purification step, crosslinked RNA is isolated and converted into cDNA and subsequently deep sequenced
PAR-CLIP	identification of target site, definition of RNA recognition element	incubation of cells with a photoactivatable ribonucleoside (4SU, 6SG), which is incorporated into newly transcribed RNA, 365 nm UV-crosslinking of RNA, IP of epitope tagged or endogenous RBP, isolation of crosslinked RNA, 3'- and 5'-adapter ligation for RT and PCR, cDNA library preparation, deep sequencing, putative RBP target sites are scored by frequency of crosslinking evidence, upon 4SU or 6SG characteristic T to C or G to A conversion, respectively, are introduced

Table adapted from: (Ascano et al., 2012)

The method of choice to analyze RNA-protein interactions within this thesis are PAR-CLIP and RIP-Seq. The methods are described in more detail in section 2.2.9 and 2.2.10. The identified RNA-binding sites are composed of a short RNA sequence and/or a structural element that interacts with amino acid residues in the RNA-binding domain of the RBP. These domains were also known as RNA recognition elements (RRE) (Ascano et al., 2013). Furthermore, biochemical and structural studies like electrophoretic mobility shift assays (EMSA) (Hellman and Fried, 2007) or isothermal-

titration calorimetry (ITC) (Feig, 2007) can be performed to test RNA-protein interactions and to confirm binding of the complex, these methods are further described in section 2.2.12.

1.4 Nucleolin, a nucleolar RNA-binding protein

Nucleolin (NCL) was first described by L.R. Orrick in 1973 and was isolated from nucleoli of rat liver cells (Orrick et al., 1973). The human NCL gene is composed of 14 exons and 13 introns and is located on chromosome 2q12-qter. The protein contains 710 aa and has a predicted molecular weight of 77 kDa. According to the acidic stretches within the N-terminal domain, the protein shows aberrant mobility in a SDS-polyacrylamide gel and runs at ~110 kDa. Regarding its mobility in a two-dimensional gel, NCL was previously known as C23 (Creancier et al., 1993; Ginsty et al., 1999; Orrick et al., 1973).

NCL consists of three structural domains, a N-terminal, a central and a C-terminal domain (Figure 10). The N terminus spans one-third of the protein and is rich in aspartic and glutamic acids (D/E) interspersed with basic sequences. NCL contains a bipartite nuclear localization signal (NLS) localized in exon 5 and a central domain, which contains four RRMs. The GAR (glycine and arginine rich) domain, also known as the RGG (Arg-Gly-Gly) domain, is located in the C terminus.
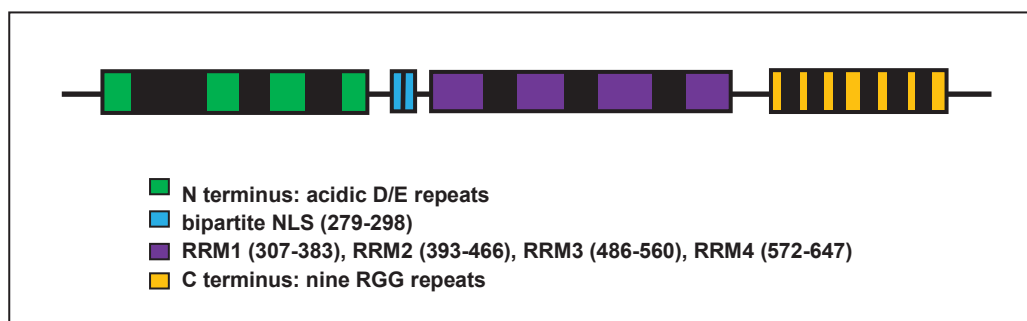
Figure 10: The structure of human Nucleolin

Nucleolin has three distinct domains. The acidic stretches (D/E) of the N terminus are marked in green, the bipartite nuclear localization signal (NLS) localized in exon 5 (279-298) is marked in blue. The four RNA-binding domains (RRM1-4) are responsible for specific RNA interactions and are illustrated in purple. The C terminus, which is rich in arginine and glycine residues (RGG) is depicted in yellow.

Each structural domain of NCL contributes to the multifunctional properties of the protein. The N terminus comprises several phosphorylation sites for the cyclin-dependent kinase I (CDK1) and CKII, which recognizes a consensus sequence that includes serine (S) and threonine (T) residues. It is proposed that the N-terminal domain interacts with several factors of the pre rRNA processing complex and is involved in rRNA transcription (Abdelmohsen and Gorospe, 2012). NCL also functions as a histone chaperone in chromatin remodeling and has similar properties like the FACT complex (Durut and Saez-Vasquez, 2015; Ginisty et al., 1999; Mongelard and Bouvet, 2007). FACT is a histone chaperone that is involved in several aspects of nucleosome formation and interacts with histones H3-H4 and H1A-H2B (Burgess and Zhang, 2013). It is supposed that the acidic stretch of NCL, which is associated with histone H1, is required for histone chaperone activity that facilitates RNA transcription through the nucleosomes (Angelov et al., 2006; Mongelard and Bouvet, 2007; Storck et al., 2007). In general, acidic stretches, also known as acid blobs or negative noodles, stimulate transcription and are common in transcriptional activating proteins (Sigler, 1988). Therefore, the N terminus of NCL might function as an acid blob involved in transcription activation and regulation of Pol I (Hanakahi et al., 1997). The central domain of NCL, containing four RRMs (RRM1-4), shows high binding affinity to RNA. Structural and biochemical studies revealed specific RNA interaction and two binding motifs of NCL were identified. A NCL recognition element (NRE) characterized by a small stem-loop structure and an evolutionary conserved motif (ECM) with a short ssRNA sequence (Allain et al., 2000a; Ghisolfi-Nieto et al., 1996; Ginisty et al., 2001; Serin et al., 1997). NCL also binds with high affinity to ssDNA and interacts with G-rich sequences that were preferentially found in non-transcribed spacer (NTS) region of the rDNA (Egyhazi et al., 1988). The RGG domains of the C terminus are responsible for interaction with RNA and RPs (Yanagida et al., 2001) and are implicated in nucleolar localization. RG or RGG domains are common in nucleolar proteins and RNA-binding

proteins containing RRMs. The RGG motif facilitates protein-mediated interactions of RNA and DNA *in vitro.* Therefore, the C-terminal domain is associated with non-specific and low RNA-binding affinity (Berger et al., 2015; Durut and Saez-Vasquez, 2015; Ginisty et al., 1999; Hanakahi et al., 1997; Medina et al., 2010).

The structure of NCL is highly conserved across species and can be found in animals, plants and yeast, whereas the amount of the acidic stretches and the RRMs differ slightly between species. Human NCL contains four RRMs, while plants and yeast have only two RRMs in their central domain (Ginisty et al., 1999). NCL counterpart in *Saccaromyces cerevisiae* is NSR1 (nuclear signal recognition 1) and GAR2 (glycine-arginine-rich 2) in *Schizosaccaromyces pombe* (Durut and Saez-Vasquez, 2015).

1.4.1 Localization of Nucleolin

NCL is mainly localized within the nucleolus in the DFC, weaker expressed in the GC and rarely expressed in the FC of the nucleolus (Escande et al., 1985). NCL contains a bipartite NLS in exon 5 (Figure 10). However, the precise mechanisms how NCL is localized within the nucleolus remain unknown. Several groups, including M.S. Schmidt-Zachmann and Nigg, and L. Creansier determined the structural domains of NCL, which are responsible for nucleolar localization. Full length NCL and NCL lacking the N terminus (ΔN) are mainly localized in the nucleolus, whereas NCL either missing the C-terminal domain (ΔC) or the central domain (ΔRRM) is primarily localized in the nucleoplasm (Creancier et al., 1993; Schmidt-Zachmann and Nigg, 1993). Consequently, it is assumed that the C terminus and the central domain with the RRMs are implicated in nucleolar localization (Hanakahi et al., 1997). Table 9 summarizes the localization of NCL according to the domains tested by Schmidt-Zachmann and Nigg (Schmidt-Zachmann and Nigg, 1993).

Table 9: Localization of Nucleolin

Nucleolin	Nucleoplasm	Nucleolus
full length		+
ΔN NCL		+
ΔC NCL	+	
ΔRRM1-4 NCL	+	

Full length Nucleolin and ΔN NCL are primarily localized in the nucleolus, whereas NCL either missing the C terminus (ΔC NCL) or the central domain (ΔRRM1-4 NCL) are mainly localized in the nucleoplasm. Therefore, it is proposed that the C terminal domain and the four RRMs are responsible for nucleolar localization. + indicates localization in the compartment (Schmidt-Zachmann and Nigg, 1993).

NCL represents ~5% of all nucleolar proteins within the nucleolus in exponentially growing cells (Tuteja and Tuteja, 1998). The major cellular fraction of NCL is localized within the nucleolus and represents more than 90% (Berger et al., 2015), whereas NCL is also localized in the nucleoplasm, cytoplasm and on the cell surface. It was also reported that NCL is a shuttling protein that relocalizes from the nucleolus to the cytoplasm and might therefore be implicated in the transport of pre ribosomal subunits (Ginisty et al., 1999). In response to stress, NCL relocalizes from the nucleolus to the nucleoplasm by formation with p53 (Tajrishi et al., 2011).

1.4.2 The role of Nucleolin in ribosome biogenesis

NCL is one of the most multifunctional RBPs involved in several aspects of gene expression and regulation, including chromatin remodeling, RNA transcription by RNA Pol I and II, nucleocytoplasmic transport, mRNA stabilization, apoptosis and ribosome biogenesis including rRNA processing, maturation and ribosome assembly (Ginisty et al., 1998; Ginisty et al., 1999; Storck et al., 2007).

Chip-Seq experiments revealed distribution of NCL along the entire rDNA coding region and promoter region. NCL showed similar binding patterns compared to the transcription factor UBTF and RPA116, the second largest subunit of Pol I (Cong et al., 2012). Previous studies suggest that NCL plays a role in the regulation of Pol I elongation. The N terminus of NCL could interact with the Pol I transcription machinery,

while the four RRMs are responsible for specific interaction with nascent rRNA and therefore, NCL is involved in the first processing and cleavage steps within the nucleolus and is associated with the highly coordinated process of ribosome biogenesis. NCL recruits ribosome biogenesis factors and the U3 snoRNA complex, which is linked to early processing. The complex binds to rDNA and is responsible for the primary cleavage within the 5'ETS. It was also assumed that NCL is implicated in the assembly of ribosomal subunits (Abdelmohsen and Gorospe, 2012; Berger et al., 2015; Ghisolfi-Nieto et al., 1996), although, NCL is absent from the mature cytoplasmic ribosome (Cong et al., 2012). Due to the histone chaperone activity *in vitro*, NCL is involved in the early remodeling complex of the chromatin and assists transcription of the Pol I complex through the nucleosome. It enables proper folding of pre rRNA and modulates chromatin structure by interaction with histone H1 through the N-terminal domain (Angelov et al., 2006; Mongelard and Bouvet, 2007). NCL binds to H2A-H2B dimers and facilitates the assembly of the nucleosomes (Cong et al., 2012; Roger et al., 2002). It was also suggested that interaction with RNA during rRNA transcription leads to a chaperone activity in NCL that permits correct folding of pre rRNA (Allain et al., 2000a; Allain et al., 2000b). Structural nuclear magnetic resonance (NMR) studies of NCL revealed that the first two RRMs (RRM1-2) bind with high affinity to an 18 nt consensus sequence found in the 5'ETS of the pre rRNA. The NRE motif, known as the Nucleolin recognition element, was identified *in vitro* using SELEX (Ghisolfi-Nieto et al., 1996). Another RNA-binding motif of NCL, known as the ECM has a consensus sequence of UCGA and is located five nt downstream of the first cleavage site in the 5'ETS of the primary rRNA transcript (Ginisty et al., 2001).

It was proposed that NCL plays a dual role in rDNA transcription (Durut and Saez-Vasquez, 2015). Small interfering RNA mediated depletion of NCL leads to cell cycle arrest, increase of apoptosis, disruption of the nucleolus (Berger et al., 2015; Tajrishi et al., 2011) and inhibition of Pol I transcription along with a decrease in pre rRNA.

Conditional knockout (KO) of NCL in DT40 cells and partial knockdown (KD) of NCL showed decreased levels of gene expression, inhibition of RNA Pol I transcription and decreased accumulation of pre rRNA (Storck et al., 2009). Depletion of NCL leads to an increase of Pol I at the 5' end of the coding region and a reduction of UBTF along the entire rDNA (Cong et al., 2012). In HeLa cells, a decrease of Pol I transcription correlates with NCL depletion and overexpression of NCL with increased levels of pre rRNA and Pol I transcription (Berger et al., 2015; Cong et al., 2012; Durut and Saez-Vasquez, 2015; Rickards et al., 2007; Storck et al., 2009). NCL was therefore identified as a transcription activator of Pol I (Durut and Saez-Vasquez, 2015). Contradictory results demonstrated that interaction of NCL with nascent pre rRNA results in inhibition of Pol I transcription, whereas Pol II and III transcription were not affected. Injection of a 2- to 3-fold excess of *Xenopus* or hamster NCL in *X. laevis* stage IV oocytes showed a reduced effect in the accumulation of pre rRNA. Injection of an anti-NCL antibody into the nucleus of *Chironomus tentans* salivary glands revealed a 2- to 3.5-fold increase in the transcription rate of rDNA (Durut and Saez-Vasquez, 2015; Egyhazi et al., 1988; Roger et al., 2002).

1.4.3 The role of Nucleolin in cancer

Until now, there are no described mutations or splicing variants of NCL involved in diseases, whereas altered protein expression is associated with different types of cancer. This can be linked to increased synthesis of rRNA (Durut and Saez-Vasquez, 2015). In actively dividing and proliferating cells, including stem cells and cancer cells, NCL is highly expressed (Abdelmohsen and Gorospe, 2012) and regulates apoptosis, cell survival and proliferation (Berger et al., 2015; Storck et al., 2007). In colorectal tumor tissues a two-fold overexpression of NCL was identified. Overexpression of NCL is also associated with several types of cancer including breast, lung, gastric and

leukemia (Hammoudi et al., 2013; Otake et al., 2007; Pichiorri et al., 2013; Qiu et al., 2013; Shen et al., 2014; Soundararajan et al., 2009; Zhao et al., 2013). In tumor cells, NCL is mostly localized in the cytoplasm and is highly expressed on the cell membrane, where it serves as an interaction partner for proteins that are involved in tumorigenesis and angiogenesis (Berger et al., 2015; Borer et al., 1989; Hovanessian et al., 2010; Semenkovich et al., 1990).

Small molecules and chemotherapeutic drugs that target NCL on the cell surface of cancer cells are used for effective cancer therapy. Although, specific targeting and depletion of NCL needs to be carefully coordinated and cannot be underestimated, since NCL is indicated in many essential aspects of gene regulation (Berger et al., 2015; Hovanessian et al., 2010). So far, AS-1411 also known as AGRO100 is a 26 nt G-rich aptamer with the following sequence GGTGGTGGTGGTTGTGGTGGTGGTGG that targets NCL in various pathways. A model proposed by Reyes-Reyes et al. indicates that cell surface NCL serves as a receptor for the aptamer and results in a selective uptake in cancer cells (Reyes-Reyes et al., 2010). AS-1411 shows high binding affinity to NCL and directly reduces the RNA-binding capacity of NCL. The aptamer was already tested in clinical trials e.g. in AML patients and in solid tumors and showed promising effects in cancer therapy (Abdelmohsen and Gorospe, 2012; Mongelard and Bouvet, 2010; Storck et al., 2007; Teng et al., 2007). CX-3543, another small molecule inhibits Pol I transcription by disrupting the interaction of NCL and rDNA and consequently induces apoptosis in cancer cells (Berger et al., 2015; Drygin et al., 2009). CX-3543, also known as quarfloxin is also summarized in Table 5.

1.5 Aim of the thesis

Nucleolin (NCL) is a highly abundant, nucleolar RNA-binding protein that is associated with the well-coordinated process of ribosome biogenesis within the nucleolus. This thesis focuses on the characterization of the RNA-binding properties of NCL, identification of cellular RNA targets and protein-protein interactions of NCL in the context of ribosome biogenesis.

First, to characterize the domains of Nucleolin, which contribute to the nucleolar accumulation, immunohistochemistry and RNA-FISH experiments of full length NCL versus NCL lacking the N terminus (ΔN) were performed. Immunohistochemistry was also conducted to examine the localization of NCL upon 4-thiouridine (4SU) treatment and in the presence of the Polymerase I inhibitor, Actinomycin-D. In order to identify transcriptome-wide RNA targets of NCL, methods like photoactivatable-ribonucleoside enhanced crosslinking and immunoprecipitation (PAR-CLIP) and RNA immunoprecipitation-sequencing (RIP-Seq) were performed, followed by biochemical and structural experiments like electrophoretic mobility shift assays, isothermal-titration calorimetry, size-exclusion chromatography and crystallography. These experiments were maintained to test the binding affinity of the identified NCL-RNA complex. Additionally, mass-spectrometry, a proteomic based approach was carried out to further elucidate the protein-protein interaction partners of NCL and to gain insight into the function of NCL in ribosome biogenesis. Furthermore, RNAi experiments in HEK293 cells were performed to verify if NCL has an impact on the expression of precursor ribosomal RNA and other factors that are associated with ribosome biogenesis.

2 MATERIAL AND METHODS

This chapter gives an overview of the utilized materials and performed methods. The chapter is divided in two parts; the first part provides the material including devices, consumables, buffers, chemicals, kits, peptides, antibodies, RNA and DNA oligos, plasmids and cell lines, while the second part describes the methods including a brief procedure of the experiment.

2.1 Material

Consumables	Company
1.5 ml siliconized tubes	Bio Plas, Inc 4165SL
1.5 ml tubes	Eppendorf
15 ml and 50 ml Falcon tubes	Greiner, bio-one
15 cm culture dishes	Greiner, bio-one
1kb Plus DNA ladder 250 µg (1 µg/µl)	Invitrogen, Carlsbad, CA, USA
365 nm, XL-1500 UV Transilluminator	Spectromics Corporation
NuPAGE Novex 4-12% BT	Invitrogen
Acclaim PepMap100 trap Column	Dionex/Thermo Scientific
Amersham Hybond-Transfer membrane	GE Healthcare, Freiburg, Germany
Centrifuge 5403	Eppendorf, Hamburg, Germany
Centrifuge 5417R	Eppendorf, Hamburg, Germany
CO_2 incubator for cell culture	Thermo Scientific
Cryo freezing container, 500 ml	Nalgene, NY, USA
Culture dish, 150 x 25 mm	Corning, Amsterdam, Netherlands
Dynabeads Protein G, 30 mg/ml	Invitrogen, Carlsbad, CA, USA

extra thick blotting paper	Bio-Rad, Munich Germany
FLA-7000 Phosphor Imager	Fujifilm, Düsseldorf, Germany
Gene Amp PCR System 2700	ABI, Carlsbad, CA, USA
HiSeq	Illumina, San Diego, CA,USA
Illustra MicroSpin, G-25 Columns	GE Healthcare, Freiburg, Germany
LAS-3000 mini 2UV Transilluminator	Fujifilm, Düsseldorf, Germany
Milli-Q Integral 15	Millipore, Billerica, MA, USA
NanoDrop Spectrophotometer ND-1000	PeqLab, Erlangen, Germany
PCR thermocycler	AB, Applied Biosystems
pH-meter 7310	inolab
Phosphorimager	FLA-7000, Fujifilm
Radioisotope laboratory	KMT-lab, Düsseldorf/Tuschl lab, NY
Scalpels	Feather
Thermomixer Comfort	Eppendorf, Hamburg, Germany
Transfer Blot Mini Trans Blot Cell	Bio-rad, Munich, Germany
Transfer Blot SD Semi-Dry transfer Cell	Bio-Rad, Munich, Germany
Ultimate 3000 Rapid Separation liquid	Dionex/ Thermo Scientific, Idstein, Germany
Vortex2 Genie	Scientific Industries, NY, USA
Orbitrap Elite high resolution Instrument	Thermo Scientific, Bremen, Germany
MicroCal iTC$_{2000}$	Malvern, USA
GeneChip Human Genome U133 A 2.0	Affymetrix, Santa Clara, CA, USA
Magnetic rack for 1.5 ml tubes	Invitrogen

2.1.1 Media, Buffer and Solution

Media, Buffer, Solution	Company/Composition of final concentration
Agarose gel	
Agarose gel	1%-2.5% agarose/low melt agarose in 1x TBE/1x TAE Buffer
Running buffer	1x TBE or 1x TAE
Crystallography	
Dialysis Buffer/ITC buffer	25 mM Tris HCL (pH 7.5), 200 mM NaCl
Lysis buffer	50 mM Tris HCL (pH 7.5), 200 mM NaCl, 40 mM Imidazol (pH 8.0), 10% glycerol, 2 mM PMI
Elution buffer	25 mM Tris, 300 mM NaCl, 250 mM Imidazol (pH 8.0)
Size-Exclusion Chromatography	
SEC buffer	25 mM Tris pH 7.5, 50 mM NaCl and DTT
Electrophoretic mobility shift assays	
5x EMSA binding buffer	10 mM Tris-HCl pH 7.5, 50 mM NaCl, 4% Glycerol, freshly added to 1x EMSA Buffer: 1 mM DTT, 0.1U/µl RNasin Ribonuclease Inhibitor (Promega)
NCL-GST storage buffer	200 mM Tris pH 8.0, 50 mM NaCl, 5 mM EDTA, 5 mM EGTA
5x TBE	445 mM Tris, 445 mM Boric acid, 1 mM EDTA (pH 8.0) (run the gel in 0.25x TBE buffer)
4% EMSA gel	6.7 ml Acrylamide (30%), 5 ml 50% glycerol, 2.5 mL 5x TBE, 500 µl APS (10%), 50 µl TEMED, add 50 ml H_2O

Isothermal-Titration Calorimetry

ITC buffer   25 mM Tris (pH 7.5), 200 mM NaCl

Mass-Spectrometry

IP-wash buffer (with and without NP40)   50 mM HEPES-KOH pH 7.5, 300 mM KCl, 0.05% (v/v) NP40, H_2O, add fresh: 0.5 mM DTT and EDTA-free protease inhibitor cocktail (PIC), (Roche)

De-staining buffer   15 mM $Na_2S_2O_3$, 50 mM $K_3[Fe(CN)_6]$

Reducing buffer   10 mM DTT, 50 mM $(NH_4) HCO_3$

Alkylation buffer   50 mM C_2H_4INO, 50 mM NH_4HCO_3

Media

LB media   5g Yeast extract, 10g Tryptone, 10g NaCl, H_2O

LB Agar plates   5g Yeast extract, 10g Tryptone, 10g NaCl, 15g Agar, 1 ml NaOH, H_2O

PAR-CLIP

IP-wash buffer   50 mM HEPES-KOH pH 7.5, 300 mM KCl, 0.05% (v/v) NP40, H_2O, add fresh: 0.5 mM DTT and EDTA-free protease inhibitor cocktail (PIC), (Roche)

High-salt wash buffer   50 mM HEPES-KOH pH 7.5, 500 mM KCl, 0.05% (v/v) NP40, H_2O, add fresh: 0.5 mM DTT and EDTA-free protease inhibitor cocktail (PIC), Roche

2x Formamide loading buffer	20 mM EDTA-NaOH pH 8.0, 0.05% (w/v) Bromophenol blue, Fromamide
Polynucleotide Kinase (PNK) buffer with DTT	50 mM Tris-HCl pH7.5, 50 mM NaCl, 10 mM MgCl2*6H$_2$O, H$_2$O, add fresh: 5 mM DTT
NP40 lysis buffer	50 mM HEPES-KOH pH 7.5, 150 mM KCl, 2 mM EDTA-NaOH pH 8.0, 2% NP40, H$_2$O, add fresh: 0.5 mM DTT and EDTA-free protease inhibitor cocktail (PIC), Roche
Phosphatase buffer	50 mM Tris-HCl pH 7.5, 20 mM EGTA-NaOH pH 7.5, 0.5% (v/v) NP40, H$_2$O
10x RNA ligase buffer without ATP	500 mM Tris-HCl pH 7.6, 100 mM MgCl2*6H$_2$O, 100 mM 2-Mercaptoethanol, 1 mg/ml acetylated BSA, H$_2$O
10x RNA Ligase buffer with ATP	500 mM Tris-HCl pH 7.6, 100 mM MgCl2*6H$_2$O, 100 mM 2-Mercaptoethanol, 1 mg/ml acetylated BSA, 2mM ATP, H$_2$O
10x Dephosphorylation buffer (NEB Buffer 3)	50 mM Tris-HCl pH 7.9, 100 mM NaCl, 10 mM MgCl2*6 H$_2$O /NEB
4x SDS-PAGE loading buffer	50 mM Tris-HCl pH 6.8, 2 mM EDTA-NaOH pH 8.0, 10% (v/v) Glycerol, 2% (v/v) SDS, 100 mM DTT, 0.1% (w/v) Bromophenol blue, H$_2$O
Page Ruler prestained Protein ladder	Thermo Scientific
2x Proteinase-K buffer	100 mM Tris-HCL pH 7.5, 12.5 mM EDTA-NaOH pH 8.0, 150 mM NaCl, 2% (v/v) SDS, H$_2$O

12% Sequa Gel	3 ml Buffer, 12.6 ml Diluent, 14.4 ml Concentrate, 12 µl TEMED, 240 µl 10% APS
15% Sequa Gel	3 mL Buffer, 9 ml Diluent, 18 ml Concentrate, 12 µl TEMED, 240 µl 10% APS
10x Tris-Borate-EDTA (TBE) buffer	445 mM Tris, 445mM boric acid, 10 mM EDTA-NaOH pH 8.0, H_2O
Polynucleotide kinase (PNK) buffer without DTT	50 mM Tris-HCl pH7.5, 50 mM NaCl, 10 mM MgCl2*6 H_2O, H_2O
Acidic Phenol/Chloroform (25:24:1)	25 ml acidic phenol, 24 ml chloroform, 1 ml isoamyl alcohol
50% DMSO	Dilute 10 ml DMSO in 10 ml H_2O
10x PCR buffer	100 mM Tris-HCl, 500 mM KCl, 10 mM 2-Mercaptoethanol, 1% (v/v) Triton X-100, 20 mM $MgCl_2$*6 H_2O, H_2O
10x dNTP Solution	2 mM dATP, 2 mM dCTP, 2 mM dGTP, 2 mM dTTP, H_2O
1M DTT	1.54 g in 10 ml H_2O
5x DNA Loading buffer	50 mM EDTA-NaOH pH 8.0, 0.2% (w/v) Bromophenol blue, 20% (w/v) Ficoll type 400, H_2O

RNA Deprotection

2'-Deprotection buffer	100 mM TEMED-Acetate, pH 3.8, Thermo Scientific

RNA-FISH

1M Tris (pH 7.4)	820 ml 1M Tris-HCl, 180 mL 1M Tris base
1M Tris (pH 8.5)	266 ml 1M Tris-HCl, 734 ml 1 M Tris base
1x TBS	0.1 M NaCl, 0.01 M Tris Buffer (pH 7.4)

4% PFA (10% Formalin)	10 ml 16% PFA, 30 mL 1x TBS
Antibody blocking solution	5% goat serum in 1x TBS-T Buffer
Mounting solution (MOWIOL)	25% glycerol, 0.1M Tris (pH 6.8), MOWIOL, H_2O
Hybridization buffer (50% FA)	12.6 M (50% (v/v)) Formamide, 1 M NaCl, 75 mM Tris Buffer (pH 8.5), 1x Denhardt's, 250 µg/ml Baker's yeast tRNA (20 mg/ml), 500 µg/ml Salmon sperm DNA, 2.5 mM Chaps, 0.5% (v/v) Tween 20, H_2O
Wash buffer I	12.6 M (50% (v/v)) Formamide, 0.25 M NaCl, 0.075 M Tris Buffer (pH 8.5), 0.1% (v/v) Tween 20
Wash buffer II	0.05 M NaCl, 0.075 M Tris Buffer (pH 8.5), 0.1% (v/v) Tween 20
1x TBS-T	10 mM Tris (pH 7.4), 100 mM NaCl, 0.1% (v/v) Tween

RNA-Labeling

10x Kinase buffer	500 mM Tris, pH 7.5, 100 mM MgCl2, 50 mM DTT

SDS-PAGE and Western Blot

1x MOPS running buffer	20x Buffer, (Invitrogen), dilute 1:20 in H_2O
10x SDS running buffer	25 mM Tris, 192 mM Glycine, 0.1% SDS
10x Transfer buffer	25 mM Tris, 192 mM Glycine
2x Laemmli buffer	125 mM Tris (pH 6.8), 10% Glycerin, 4%SDS, Bromphenolblue, ad 50 ml H_2O. Add 200 µl β-mercaptoethanol/1.8 ml aliquot freshly prior use

SDS stacking gel buffer	30% Acryl-Bisacrylamide (19:1), 126 mM Tris-HCL pH 6.8, 0.1% SDS, 0.1% TEMED, 0.1% APS
SDS separation gel buffer	30% Acryl-Bisacrylamide (19:1), 378 mM Tris-HCL pH 8.0, 0.1% SDS, 0.06% TEMED, 0.1% APS
Silver Staining	
Fixing solution	10% Acetic Acid, 50% Methanol, H_2O
Staining solution	20% Methanol, 20% Stainer A, 5% Stainer B, H_2O
Sensitizing solution	30% Ethanol, 10% Sensitizer, H_2O
Developing Solution	10% Developer, 1 drop Derveloper Enhancer, H_2O
Stopper	Life Technologies
Wash buffer	30 % ethanol, H_2O

2.1.2 Cells

Cell line/Cells	Company
HEK293	ACC-305
Flp-In T-Rex HEK293	Life Technologies
OneShot, TOP10 *E.coli* cells	Invitrogen

2.1.3 Cell Culture Media and Antibiotics

Cell culture Media and Antibiotics	Company
Dulbecco's Modified Eagle Medium (DMEM)	Gibco, Invitrogen, Carlsbad, USA
Fetal Bovine Serum (FBS) heat inactivated	PAA, Pasching, Austria
L-Glutamine, 200 mM	Gibco, Invitrogen, Carlsbad, USA
Penicillin, 10,000 U/ml	Gibco, Invitrogen, Carlsbad, USA
Streptomycin, 10,000 µg/µl	Gibco, Invitrogen, Carlsbad, USA

Hygromycin, 100 mg/ml	InvivoGen
Zeocin, 100 mg/ml	InvivoGen
Blasticidin, 10 mg/ml	InvioGen
1x Trypsin-EDTA, 0.05%	Gibco, Life Technologies
Ampicillin sodium salt	SIGMA
Kanamycin sulfate	SIGMA
Doxycycline	SIGMA

2.1.4 Chemicals and Enzymes

Chemicals and Enzymes	Company
[γ-^{32}P] Adenosine 5'-triphosphate (ATP)	Hartmann Analytic, Germany
[γ-^{32}P] Adenosine 5'-triphosphate (ATP)	Perkin Elmer, USA
[α-^{32}P] Adenosine 5'-triphosphate (ATP)	Perkin Elmer, USA
[γ-^{32}P] Guanosine 5'-triphosphate (GTP)	Perkin Elmer, USA
[5'-^{32}P] Cytidine 3', 5' bis(phosphate) (pCp)	Perkin Elmer, USA
1,4 dithiothreitol (DTT)	Roth, Karlsruhe, Germany
4-thioruidine (4SU)	Sigma-Aldrich, St. Louis/Tuschl lab
50 x Denhardt Solution	AppliChem, Darmstadt, Germany
50 x TAE (Tris/Acetic Acid/EDTA) Buffer	Bio-Rad, Munich, Germany
AccuPrime SuperMix I	Invitrogen
AccuPrime Pfx DNA Polymerase	Invitrogen
Agarose	Biozym, Hessisch Odendorf, Germany
Ammonium Persulfate (APS)	Sigma-Aldrich, St. Louis, MO, USA
Bis-Acrylamide 30% (19:1)	Bio-Rad, Munich, Germany

Bromphenol blue	Sigma-Aldrich, St. Louis, MO, USA
Chloroform	Merck, Darmstadt, Germany
Complete, EDTA-free Protease Inhibitor	Roche Diagnostics, Mannheim
DMSO (Dimethyl sulfoxide)	Sigma-Aldrich, St. Louis, MO, USA
DAPI (4',6-diamidino-2-phenylindole)	Sigma-Aldrich, St. Louis, MO, USA
EDTA (Ethylenediamine-tetraacetic acid)	Sigma-Aldrich, St. Louis, MO, USA
Ethanol	Merck, Darmstadt, Germany
Ethidium bromide Solution	Sigma-Aldrich, St. Louis, MO, USA
Glycerine	Merck, Darmstadt, Germany
Glycine	Merck, Darmstadt, Germany
GlycoBlue 300 µl (15 mg/ml)	Ambion, Cambridgeshire, UK
Gateway LR Clonase II enzyme mix	Invitrogen
Gateway BP clonase II enzyme mix	Invitrogen
H_3BO_3 (Boric acid)	Merck, Darmstadt, Germany
HCl (Hydrochloric acid)	Merck, Darmstadt, Germany
Isoamylalcohol	Merck, Darmstadt, Germany
KCl	Merck, Darmstadt, Germany
KOH	Merck, Darmstadt, Germany
Magnesium	Merck, Darmstadt, Germany
Methanol	Merck, Darmstadt, Germany
Mg_2Cl	Merck, Darmstadt, Germany
MOWIOL	Polysciences
Na_2EDTA	Sigma-Aldrich, St. Louis, MO, USA
Na_2HPO_4	Merck, Darmstadt, Germany
NaCl	Merck, Darmstadt, Germany
NaF	AppliChem, Darmstadt, Germany

NaH$_2$PO$_4$	Merck, Darmstadt, Germany
NaOH	Merck, Darmstadt, Germany
Phenol	Merck, Darmstadt, Germany
Phenol acid	Sigma-Aldrich, St. Louis, MO, USA
Precision Plus Protein Standard	Bio-Rad, Munich, Germany
RNA Ligase 2 (Rnl2 (1-249) K227Q) (1 µg/µl)	NEB, Frankfurt, Germany
RNAse T1 (1,000 U/µl)	Thermo Scientific
T4 RNA Ligase (10 U/µl)	Thermo Scientific
Alkaline Phosphatase Calf Intestinal	NEB, Frankfurt, Germany
T4 RNA Ligase 1	NEB, Frankfurt, Germany
T4 Polynucleotide Kinase (10,000 U/ml)	NEB, Frankfurt, Germany
Salmon Sperm DNA (10 mg/ml)	Applichem, USA
SDS 20%	Life technologies, Germany
SYBR Green PCR Master Mix	ABI, Carlsbad, CA, USA
T4 Polynucleotide Kinase, 10,000 U/ml	NEB, Frankfurt, Germany
TEMED (tetramethylethylenediamine)	Merck, Darmstadt, Germany
Tris	Roth, Karlsruhe, Germany
TRIzol Reagent	Invitrogen, Carlsbad, CA, USA
Tween20	Merck, Darmstadt, Germany
Super Script III	Invitrogen

2.1.5 Commercial Kits

Kits	Company
Cloning	
Gateway Recombination Cloning Technology	Thermo Scientific
QIAquick PCR Purification Kit	Qiagen, Hilden, Germany
QIAfilter Plasmid Midi kit	Qiagen, Hilden, Germany
QIAprep Spin Miniprep	Qiagen, Hilden, Germany
QIAquick Gel Extraction Kit	Qiagen, Hilden, Germany
GenScript's CloneEZ PCR Clonning Kit	GenScript, NJ, USA
Sequencing	
TruSeq SBS Kit v3 – HS	Illumina
TruSeq Small RNA Sample Preparation Kit	Illumina
TruSeq SR Cluster Kit v3 – cBot – HS	Illumina
TruSeq Stranded mRNA LT kit	Illumina
Crystallography	
JESG Cove	Qiagen, Hilden, Germany
JESG Cove II	Qiagen, Hilden, Germany
Classics	Qiagen, Hilden, Germany
Classics II	Qiagen, Hilden, Germany
PEGs II	Qiagen, Hilden, Germany

2.1.6 Plasmids of the Gateway Cloning System

Plasmid	Company
pENTR4 (entry vector)	Invitrogen, Life Technologies
pOG44 (expression vector for Flp recombinase)	Invitrogen, Life Technologies

pFRT_DESTFLAG/HA (N-terminal double affinity tag, for stable expression)	Invitrogen, Life Technologies, www.addgene.com
pFRT_TO_DESTFLAG/HA (N-terminal double affinity tag, for tetracycline inducible expression)	Invitrogen, Life Technologies, www.addgene.com
pFRT_DESTFLAG/HA_eGFP (stable expression of eGFP	Invitrogen, Life Technologies, www.addgene.com

2.1.7 DNA and RNA oligos

PAR-CLIP

3'adapter_21.930	5'-rAppTCGTATGCCGTCTTCTGCTTGT-3'
5'adapter_26.68	5'-GUUCAGAGUUCUACAGUCCGACGAUC-3'
19 nt size marker	5'-CGUACGCGGGUUUAAACGA-3'
24 nt size marker	5'-CGUACGCGGAAUAGUUUAAACUGU-3'
35 nt size marker	5'UCAUCUUGGUCGUACGCGGAAUAGUUU AAACUGU-3'

RNA oligos

Oligo_1	5'-CGUCCGAGCGUCCCGACUCCCGGUG-3'
Oligo_2	5'-GGUGCGUGUGGGAAGGCGUGGGGUG-3'
Oligo_3	5'-UGUCGACUUGCGGGCGGCCCCCCU-3'
Oligo_4	5'-CUUCGUGAUCGAUGUGGUGACGUCG-3'
Oligo_5	5'-CGUCCGAGCGUCCCGACUCCCGGUGC-3'
Oligo_6	5'-CGUCCGAGCGUCCCCACUCCCCGUG-3'
Oligo_7	5'- CGUCCGAGCGACGU-3'
Oligo_8	5'-AGCGUCCCGACUCCCG-3'
Oligo_9	5'-AGCGUCCCGACUCCCGGUGC-3'
Oligo_10	5'-UCCGAGCGUCCCGACUCCCG-3'

Oligo_11	5'-CGUCCGAGCGUCCCGACGUG-3'
Oligo_12	5'-CGUCCGAGCGACUCCCGGUG-3'
Oligo_13	5'-CGUCCGAGCGUCCCGACUCCCGGUGC-3'
Oligo_14	5'-GGCCGGUUCCGGUUCCUU-3'
Oligo_15	5'-CUGGGG CUGGGG CUGGGG-3'
Oligo_16	5'-CUG CUG CUG CUG CUG CUG-3'
Oligo_17	5'-GGU GGU GGU GGU GGU GGU-3'
Oligo_18	5'-GUU GUU GUU GUU GUU GUU-3'
Oligo_19	5'-GGC GGC GGC GGC GGC GGC-3'
Oligo_20	5'-CGUCCGAGCGUCCCGACUCCCGGUG-3'
Oligo_21	5'- AAA AAA AAA AAA AAA AAA AA-3'
Oligo_22	5'-ACU ACU ACU ACU ACU ACU-3'
Oligo_23	5'-AUU AUU AUU AUU AUU AUU-3'
Oligo_24	5'-GU GU GU GU GU GU GU GU GU-3'
Oligo_25	5'-GUU GUU GUU GUU GUU GUU-3'
Oligo_26	5'-GGC GGC GGC GGC GGC GGC-3'
Oligo_27	5'-GU GU GU GU GU GU GU GU GU-3'
Oligo_28	5'-GGU GGU GGU GGU GGU GGU-3'
Oligo_29	5'-GUU GUU GUU GUU GUU GUU-3'
Oligo_30	5'-CGGGCCGGCGGGGUCCUCUGACG-3'
Oligo_31	5'-AC AC AC AC AC AC AC AC AC-3'
poly (A) (25 nt)	5'- AAA AAA AAA AAA AAA AAA AAA AAA A-3'
RNA-FISH probes	Z: 5'-Amino Modifier C6-TFA,
	W: 3'-Fmoc-Amino-modifier C7 CPG
28 S_1	Z-GCTTAAATTCAGCGG-W

28 S_2	Z-GGTCCTAACACGTGC-W
28 S_3	Z-AGGCACTCGCATTCC-W
28 S_4	Z-CTTCGCGATGCTTTG-W

2.1.7.1 Primer

Primer	Sequence (5' – 3')
NCL Sequencing primer	
NCL_414_fw	5'- GAATGCCAAGAAGGAAGACAG-3'
NCL_903_fw	5'- AGAACCGACTACGGCTTTCA-3'
NCL_1311_fw	5'- GACAGAAGCTGATGCAGAGAAA-3'
NCL_1796_fw	5'- GGATAGTTACTGACCGGGAAA-3'
NCL_rev	5'- CTATTCAAACTTCGTCTTCTTTCCTTGTGG-3'
Cloning primer	
NCL_298_fw	5'- acagattggtGGATCCgtggaaggcacagaaccg -3'
NCL_468_rev	5'- ctttaccagaCTCGAGTCAacctttctctccagtata -3'
NCL_565_rev	5'- ctttaccagaCTCGAGTCAattaggtgatcccctggg -3'
NCL_648_rev	5'- ctttaccagaCTCGAGTCActtaggtttggcccagtc -3'
NCL_710_rev	5'- ctttaccagaCTCGAGTCAttcaaacttcgtcttctt -3'
Gateway Cloning primers	
NCL_attB1_fw	5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTC ATGGTGAAGCTCGCGAAGGCA-3'
NCL_attB1_rev	5'GGGGACCACTTTGTACAAGAAAGCTGGGT CTATTCAAACTTCGTCTTCTTTCC-3'
NCL_ fw	5'-ATGGTGAAGCTCGCGAAGGCA -3'
NCL_rev	5'-CTATTCAAACTTCGTCTTCTTTCCTTGTGG -3'

Sequencing primer

M13/pENTR4_fw	5'- TGTAAAACGACGGCCAGT -3'
M13/pENTR4_rev	5'- CAGGAAACAGCTATGAC -3'
pFRT_DESTFLAG/HA_fw	5'- GGTCTATATAAGCAGAGCTCTC -3'
pFRT_GFP_dest_fw	5'- ACAACCACTACCTGAGCA -3'
pFRT_DESTFLAG/HA_rev	5'- TAGAAGGCACAGTCGAGG-3'
pFRT_to_DESTHA_fw	Same as pFRT_DESTFLAG/HA_fw
pFRT_to_DESTHA_rev	Same as pFRT_DESTFLAG/HA_rev

PAR-CLIP primer

3' PAR-CLIP PCR_21.929	5'- CAAGCAGAAGACGGCATACGA -3'
5' PAR-CLIP PCR_44.32	5'-AATGATACGGCGACCACCGACAGGTTC AGAGTTCTACAGTCCGA-3'

2.1.7.2 siRNAs

siRNA	sense sequence	antisense sequence
NCL I	CUAUAGAGGUGGAAAGAAUuu	AUUCUUUCCACCUCUAUAGuu
NCL II	GAAAGAAGGAAAUGGCCAAuu	UUGGCCAUUUCCUUCUUUCuu
NCL III	GGAAAUGUCAGAAGAUGAAuu	UUCAUCUUCUGACAUUUCCuu

2.1.8 Inhibitors

Inhibitor	Company
Actinomycin-D	Sigma-Aldrich, St. Louis, MO, USA
Casein Kinase II Inhibitor I, TBB	Calbiochem, Millipore, USA
Casein Kinase II Inhibitor III, TBCA	Calbiochem, Millipore, USA

2.1.9 Antibodies and Peptides

Antibody/Peptide	Company
Anti-Nucleolin, rabbit, polyclonal	Abcam, ab70493
Anti-C23 (MS-3), mouse, monoclonal	Santa Cruz, sc-8031
Anti-RPA194, mouse, monoclonal	Santa Cruz, sc-46699
Anti-FLAG M2, F3165	Sigma-Aldrich, St. Louis, MO, USA
3x FLAG-peptide	Tuschl lab, NY, USA
Anti-Casein Kinase II, rabbit, polyclonal	Millipore, 06-873
Goat-anti-Mouse-HRP	Dako
Goat-anti-Rabbit-HRP	Dako
Goat-anti-Mouse, Alexa Fluor 488, polyclonal	Thermo Scientific
Anti β-Tubulin, mouse, monoclonal, T4026	Sigma-Aldrich, St. Louis, MO, USA
Anti β-Actin, mouse, monoclonal, A2228	Sigma-Aldrich, St. Louis, MO, USA

2.1.10 Recombinant Protein

Protein	Company
GST-NCL, (289-709)	Invigate, Jena, Germany
NCL, RRM1-2	M. Teplova, MSKCC, NY, USA
NCL, RRM1-4	M. Teplova, MSKCC, NY, USA

2.2 Methods

2.2.1 Generation of stable and inducible Flp-In T-REx HEK293 cells

The use of the Gateway recombination cloning technology and Flp-In T-REx lines enables rapid creation of stable and inducible mammalian cell lines for regulated protein expression. Stable and inducible human embryonic kidney (HEK) 293 cells flanked with a FLAG/HA or GFP affinity tag were generated using the Gateway Flp-In T-REx expression system from Invitrogen. The constructs were cloned according to the protocol by Spitzer et al. (Spitzer et al., 2013). For subsequent purification and immunoprecipitation (IP) a 17 amino acid FLAG/HA-double affinity tag was linked to the N terminus of the gene of interest (GOI). The use of conventional cloning-based Flp-In recombination vectors facilitates rapid creation of the cell lines.

In the first step, the GOI flanked by *att*B sites is cloned into a donor vector flanked with *att*P sites producing an entry vector using a BP clonase reaction. In a second step, the GOI is subcloned from the entry vector flanked by *att*L sites into a destination vector flanked by *att*R sites. Subsequently, a LR clonase reaction is performed to generate the final expression vector. The BP and the LR reaction are both reversible. Therefore, it is a very flexible approach to shuttle the insert DNA from one expression vector to another. Figure 11 illustrates the Gateway recombination cloning technology with the shuttling procedure of the vectors.
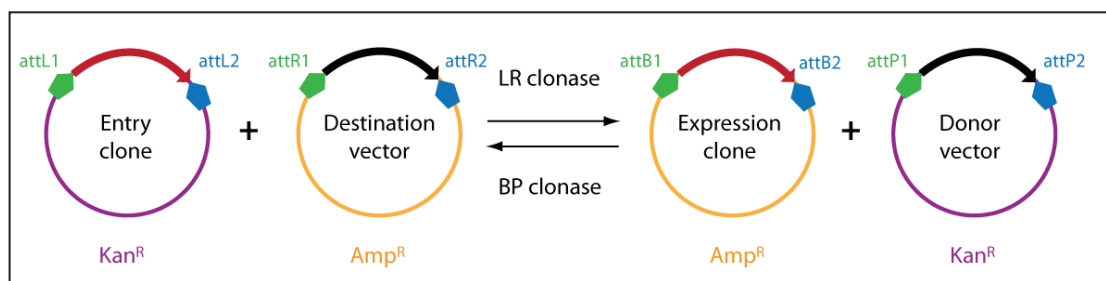
Figure 11: The Gateway recombination cloning technology

Rapid creation of stable and inducible mammalian cell lines for regulated protein expression is maintained using the Gateway recombination cloning technology and Flp-In T-REx lines. The cloning technology from Invitrogen ensures easy and rapid cloning. In the first step, the gene of interest flanked by *att*B sites is cloned into a donor vector flanked with *att*P sites. The entry vector flanked by *att*L sites is produced performing a BP clonase reaction. In a second step, the gene of interest is sublconed from the entry vector into a destination vector flanked by *att*R sites. In a final step, a LR clonase reaction is performed to create the expression vector that is flanked by *att*B sites. Figure: Invitrogen

First, the GOI is amplified from cDNA, in a second PCR the *att*B sites are integrated with specific PCR primers at both ends of the gene. The *att*B sites serve for further recognition of the vector and for subsequent LR and BP reaction. The AccuPrime Kit from Thermo Scientific was used; 5 µl Accu Prime buffer, 0.5 µl of each primer (100 µM), 42.2 µl H_2O, 1 µl cDNA and 0.8 µl polymerase with the following PCR cycles, shown in Table 10.

Table 10: PCR cycles

	Temperature	Duration	
Heat activation	95°C	30 sec	
Denaturation	95°C	45 sec	
Annealing	60°C	30 sec	30x
Elongation	72°C	150 sec	
Final elongation	72°C	5 min	
Hold	4°C	∞	

For further downstream analysis, the PCR-product is purified from primers, nucleotides, polymerase and salts using the QIAquick PCR Purification kit from Qiagen. The purification step was performed according to the manufacturer's protocol.

2.2.1.1 The Gateway Recombination Cloning Technology

To create the entry vector, a BP recombination reaction was carried out using the *att*B flanked DNA fragment (attB-PCR product) and the donor vector (pDONR vector 221).

Table 11: BP Reaction

BP reaction	Volume	
attB-PCR product (≥10 ng/µl, final amount 15-150 ng)	1-7 µl	
pDONR vector 221 (150 ng/µl)	1 µl	1 h, 25°C
TE buffer, (pH 8)	to 8 µl	
BP-clonase II enzyme mix	2 µl	
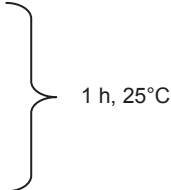
The reaction was incubated for 1 h at 25°C, afterwards 2 µg Proteinase K (Invitrogen) was added to the reaction and incubated for 10 min at 37°C. A volume of 2 µl BP reaction was transformed into 50 µl competent *E.coli* cells (One Shot TOP10, Invitrogen). The *E.coli* cells were incubated on ice for 30 min followed by a heat shock at 42°C for 30 sec. The cells were cooled down on ice prior adding 250 µl pre-warmed S.O.C. media. The cells were finally incubated for 1 h at 37°C shaking at 300 rpm. The cells were then plated on agar plates containing 100 µg/ml kanamycin and incubated overnight at 37°C. The next day, colonies were picked and the appropriate antibiotic-resistant media (kanamycin) is inoculated and incubated overnight at 37°C. A Mini prep was performed according to the manufacturer's instructions using the QIAprep Miniprep Kit from Qiagen.

Subsequently, a LR recombination reaction was maintained using the previously generated *att*L entry clone and an *att*R flanked destination vector in order to create the final expression clone containing the GOI (Table 12).

Table 12: LR Reaction

LR reaction	Volume
Entry vector (50-150 ng)	1-7 µl
Destination vector (150 ng/µl)	1 µl
TE buffer, (pH 8)	to 8 µl
LR Clonase II enzyme mix	2 µl

1 h, 25°C

After incubating the reaction mix for 1 h at 25°C, 2 µg Proteinase K (Invitrogen) was added to the reaction and incubated for 10 min at 37°C. A volume of 5 µl of the LR reaction was transformed into 50 µl competent *E.coli* cells (One Shot TOP10, Invitrogen). The cells were incubated on ice for 30 min, a heat shock (42°C for 30 sec) was performed and the cells were cooled down on ice prior adding 250 µl pre-warmed S.O.C. media. The cells were incubated for 1 h at 37°C shaking at 300 rpm, 50 µl and 100 µl were plated on agar plates containing 100 µg/ml ampicillin and the cells were incubated overnight at 37°C. The next day, colonies were picked and the appropriate

antibiotic-resistant media was inoculated and incubated overnight at 37°C. A Mini prep was performed according to the manufacturer's instructions using the QIAprep Miniprep Kit (Qiagen). In a next step, the expression clone containing the GOI was transfected into parental Flp-In T-REx HEK293 cells.

2.2.1.2 Transfection of HEK293 cells using Lipofectamine 2000

One-day prior transfection, 50.000 Flp-In T-REx HEK293 cells were seeded in a 12-well plate in antibiotic free media. Control cells were transfected either with pOG44, containing the Flp recombinase enzyme or with a pFRT_DESTFLAG/HA_eGFP vector in order to estimate the transfection efficiency. The remaining cells were transfected in a 9:1 ratio, using pOG44 (0.9 µg, maximum of 3 µl) and the expression vector (0.1 µg, maximum of 3 µl) containing the GOI. Either the pFRT_DESTHA_GOI or the pFRT_TO_DESTFLAG/HA_GOI vector was utilized respectively for creation of stable or inducible cell lines. For each transfection, 50 µl Opti-MEM was mixed with 2 µl of Lipofectamine 2000 and with the vector containing the GOI. Both reactions were properly mixed, incubated separately for 5 min at RT and then combined and incubated for 20 min at RT prior transfection.

Table 13: Transfection using Lipofectamine 2000

Reagent	Concentration/Volume	Opti-MEM		
pOG44: vector_GOI	0.9 µg:0.1 µg	50 µl	5 min, RT	20 min, RT
Lipofectamine 2000	2 µl	50 µl	5 min, RT	

The day after transfection, the media was replaced with fresh antibiotic-free media. After 48 h to 72 h the antibiotic-free media was replaced with selection media, containing hygromycin (hygro) or hygro and blasticidin (blast) for stable and inducible cell lines, respectively. Cells, transfected either with the pOG44, containing the Flp recombinase enzyme or with the eGFP vector, were also selected using hygro.

2.2.1.3 Cultivation of Flp-In T-REx HEK293 cells

Flp-In T-REx HEK293 cell lines were grown in 1x DMEM (Dulbecco's modified eagle medium) including 10% FBS, 100 U/ml penicillin, 100 µg/ml streptomycin, 2 mM L-glutamine, 15 µg/ml blast and 100 µg/ml zeocin. Inducible cells, transfected with the pFRT_to_DESTFLAG/HA plasmid were grown in 1x DMEM (Dulbecco's Modified Eagle Medium) containing 15 µg/ml blast and 100 µ/ml hygro. Inducible HEK293 cells were induced with 1 ng/µl doxycycline for approximately 24 h to 30 h according to the expression of the GOI. Stable cells, transfected with the pFRT_DESTFLAG/HA plasmid, were grown in selection media containing 100 µg/ml Hygromycin. In order to split the cells, they were previously washed with PBS (phosphate-buffered saline), trypsinized and centrifuged for 5 min at 400 x g. Cells were cultivated at 37°C with 5% CO_2. Cells were frozen in cryo vials in FBS and DMSO with a 9:1 ratio and stored either at -80°C or for long-term storage in liquid nitrogen.

2.2.1.4 Nucleolin constructs

Stable and inducible HEK293 cells expressing either full length NCL with a FLAG/HA or a GFP affinity tag, and NCL lacking the N terminus (ΔN, delta N) were generated (Figure 12).
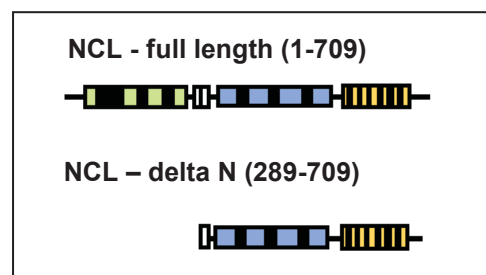
Figure 12: Nucleolin constructs

HEK293 cells were generated expressing full length Nucleolin (1-709) and Nucleolin lacking the N-terminal domain (ΔN, 289-709).

2.2.2 SDS-PAGE and Western Blot analysis

Cells were harvest, washed once with PBS and centrifuged at 400 x g for 5 min. The supernatant was discarded, the pellet was resuspended in 2% NP40 lysis buffer and incubated on ice for 20 min. The cells were centrifuged at 4°C, at 13.000 rpm for 30 min. The supernatant was transferred into a new 1.5 ml tube and the pellet containing cell debris was discarded. Quantitative determination of the protein concentration was performed with a BCA assay or a Bradford assay according to the manufacturer's instructions (Pierce® BCA Protein Assay Kit, Thermo Scientific or Bradford Reagent, Bio-Rad).

Electrophoretic separation of the protein samples was performed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and a maximum voltage of 200 V was applied. The samples were mixed either with 4x SDS-PAGE loading buffer or 2x Laemmli buffer and denaturated at 90°C for 2 min. The prestained protein ladder (Thermo Scientific) was utilized as a reference for the molecular weight (MW). In order to make the proteins accessible for further antibody detection, the separated proteins within the polyacrylamide gel were transferred to a protein binding nitrocellulose membrane, using either a wet tank or a semi-dry procedure. In general, for electrophoretic transfer of proteins onto the membrane, a sandwich layer is assembled with the gel and the PVDF-membrane embedded in between. According to the transfer direction of the proteins, from the negative charged cathode to the positive charged anode, the protein transfer was carried out at 200 mA for 2 h.

Optional, the transfer can be visualized using Ponceau red. In order to block protein binding sites and prevent unspecific antibody binding, the membrane was incubated in 5% milk powder in TBS-T for 30 min at RT. The membrane was then incubated with the appropriate primary antibody (1:1,000) in 5% BSA in TBS-T for 1 h at RT or overnight at 4°C. As a standard and internal control β-actin or β-tubulin (1:10,000) was

used in 5% BSA in TBS-T. After incubation of the primary antibody, the membrane was washed three times in PBS-T (0.1% Tween) to remove non-specifically bound reagents and to reduce the background. The membrane was subsequently incubated for 1 h at RT with the corresponding secondary antibody (1:5,000) conjugated to horseradish peroxidase (HRP) in 5% BSA in TBS-T or overnight at 4°C. After incubation of the secondary antibody, the membrane was washed three times in PBS-T, 10 min each. According to the company's instructions the membrane was incubated for 5 min in ECL (Enhanced Chemi-Luminescence) solution and subsequently exposed using the LAS Mini-3000 system.

2.2.3 NoD, a nucleolar localization signal prediction tool

The prediction tool NoD (nucleolar localization sequence detector), is a web server (http://www.compbio.dundee.ac.uk/nod) and a command-line program, which provides computer predictions and analyzes the presence of a nucleolar localization signal (NoLS) within a protein sequence. The web server requires a FASTA formatted protein sequence, it computes a confidence score and returns a prediction for a NoLS along the entire protein sequence. An output figure illustrates an overview of the entire protein sequence and shows a predicted score for each residue of the sequence and shows the protein regions, which are close to the cut-off threshold and could be predicted as a NoLS. The computational prediction tool was accomplished on a manually curated set of 46 human experimentally validated NoLS. Several hundreds of proteins that are not localized within the nucleolus serve as a negative data set (Scott et al., 2010; Scott et al., 2011).

2.2.4 Immunohistochemistry and RNA Fluorescent In-Situ Hybridization

A total of 50.000 Flp-In T-REx HEK293 cells were seeded per well in a 4-well chamber slide. Optional, the cells were treated with inhibitors or inducible cell lines were induced with 1 ng/µl doxycycline prior fixation. The cells were washed once with 500 µl 1x TBS-T for 1 min at 4°C.

For immunohistochemistry (IHC), cells were fixed in acetone for 10 min and washed twice with ice cold PBS. In order to block unspecific binding of the antibodies, cells were incubated in blocking solution containing 2% BSA in PBS-T (0.1% Tween) for 30 min. Then, the cells were incubated with the first antibody in blocking solution for at least 2 h at RT or overnight at 4°C. Cells were washed with PBS and subsequently incubated with the secondary antibody (1:1,000) and DAPI (4',6-diamidino-2-phenylindole) (1:1,000) in blocking solution for 2 h at RT. Cells were washed three times for 5 min, prior mounting the coverslip with Mowiol.

For RNA-FISH experiments, cells were fixed in 4% PFA in 1x TBS for 1 h 45 min at 4°C. Fixed cells were then washed once with TBS-T for 3 min at RT on a rocking shaker. Cells were pre-hybridized with hybridization buffer containing 50% formamide (FA) either for 15 min at RT or for several hours at 4°C. Hybridization of the appropriate RNA-FISH probes was generated over night, for a max. of 14 h in a hybridization oven at 40°C. Cells were washed twice with 500 µl wash buffer I for 5 min, once with wash buffer II for 3 min and once with 1x TBS-T for 3 min at RT. Prior antibody incubation, the cells were blocked with goat serum (GS) in TBS-T for 20 min. All washing and blocking steps were performed at RT on a rocking shaker. Cells were incubated with the first antibody (anti-C23 or anti-FLAG antibody for NCL) in blocking solution (1:1,000) for 2 h 30 min at RT. Cells were washed three times with 500 µl 1x TBS-T for 3 min. The secondary, a fluorescent-conjugated antibody (1:1,000) was incubated simultaneously with DAPI (1:1,000) in blocking solution for 1 h 30 min at RT.

Subsequently, the cells were washed three times with 1x TBS-T for 3 min prior mounting the coverslip with Mowiol.

2.2.5 TRIzol - Extraction of RNA

TRIzol is applied for the isolation of high quality RNA from cells or tissues. The amount of TRIzol is based on the area of the cell culture plate and does not correlate with the number of cells. Insufficient amount of TRIzol can lead to DNA contamination. If not specified otherwise, 1 ml TRIzol is directly added to cells in a 10 cm^2 cell culture dish. To ensure complete homogenization, cells were resuspended and transferred to a tube and incubated for 5 min at RT. Per 1 ml TRIzol, 200 µl chloroform was added to the homogenized sample. Samples were vortexed vigorously for 15 sec and incubated for 3 min at RT before centrifugation at 12,000 x g for 15 min at 4°C. The aqueous phase, containing the RNA was transferred to a new tube and the TRIzol-chloroform extraction was repeated to ensure pure and high quality RNA. The RNA was precipitated in 1 ml isopropyl alcohol per 1 ml TRIzol, incubated at RT for 10 min and centrifuged at 12,000 x g for 10 min at 4°C. The RNA was washed twice in 1 ml 75% ethanol per 1 ml TRIzol and the pellet was air dried and dissolved in an appropriate volume (~ 100) µl H_2O.

2.2.6 The 5' and 3' end labeling of RNA

The 3' as well as the 5' end of the RNA were labeled with radioactive nucleoside triphosphate. The 5' end was marked using either ^{32}P γ-ATP or γ-GTP and T4 PNK. The T4 PNK catalyzes the transfer of the γ-ATP (or γ-GTP) to the 5' hydroxyl terminus of the RNA. This reaction is also known as a kinase or phosphorylation reaction. Nucleic acids containing a 5' OH group can directly be submitted to a kinase reaction, whereas nucleic acids carrying a 5' phosphate group need to be dephosphorylated with

CIP prior RNA labeling. Removal of the 5' phosphate of the RNA is described in detail in the following section 2.2.9.3.

Table 14: The 5' end labeling reaction of RNA

Reagent	Concentration/Volume
RNA	0.1-100 pmol
γ ^{32}P-ATP	25 pmol
10x Kinase buffer	2 µl
Nuclease free H_2O	up to 20 µl
T4 PNK (10 U/ml)	1 µl

The reaction was denatured at 90°C for 30 sec, directly placed on ice prior adding 1 µl T4 PNK and incubated at 37°C for 30 min. For further 5'adapter ligation, it is recommended to add 100 µM up to 1000 µM non-radioactive ATP for 5 min at 37°C to ensure a complete phosphorylation reaction. For the 3' labeling, ^{32}P-pCp and T4 RNA ligase 1 was used. The reaction was incubated overnight at 4°C. Usually, only 10% of the radioactive pCp is incorporated in the 3' end labeling reaction.

Optional, for both end-labeling reactions unincorporated γ ^{32}P-ATP or ^{32}P-pCp can be purified by applying the samples to a RNase-free Sephadex G25 column (GE Life Science) or separating the samples on a 15% denaturing 8 M urea polyacrylamide gel (UreaGel system from National Diagnostics) for 45 min at 30 W.

Table 15: The 3' end labeling reaction of RNA

Reagent	Concentration/Volume
RNA	50-100 pmol
^{32}P-pCp	50-100 pmol (equimolar to RNA)
10x T4 RNA Ligase buffer	2 µl
Nuclease free H_2O	up to 18 µl
T4 RNA Ligase 1 (10 U)	1 µl

2.2.7 Deprotection of a 2'-ACE protected RNA oligo

All custom synthesized RNA oligos from Thermo Scientific, Dharmacon or siRNA oligos synthesized by G. Bennett (Tuschl lab, Rockefeller University) were delivered in the 2'-ACE protected form.
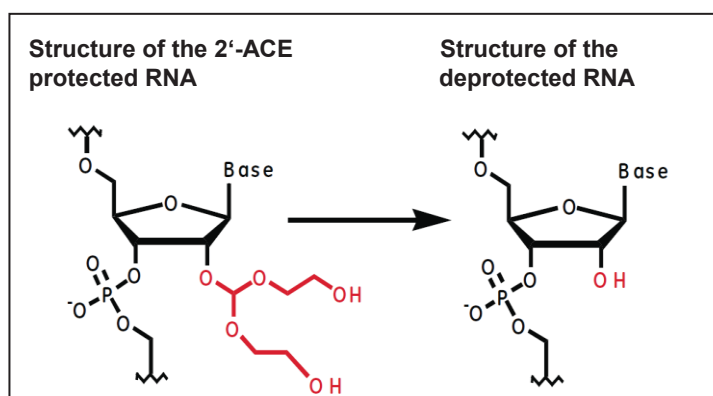
Figure 13: Deprotection of a 2'ACE protected RNA

The chemical reaction of the 2'ACE protected RNA to the deprotected RNA is indicated by the red residues. Figure: Thermo Scientific.

The protected groups of the RNA are resistant to nuclease degradation and assist to reduce secondary structures (Thermo Scientific). For deprotection, the RNA was briefly centrifuged and the RNA pellet was dissolved in 360 µl of 2'-deprotection buffer (100 mM TEMED-Acetate, pH 3.8). The RNA was vortexed and centrifuged briefly for 10 sec and incubated at 60°C for 1 h at 650 rpm. A total volume of 40 µl of 4 M NaCl was added, vortexed and up to 1.2 ml ethanol was added to precipitate the RNA. The RNA was centrifuged at 4°C at 13,000 rpm for 10 min. The supernatant was discarded. For further processes, the RNA was dried using a SpeedVac and dissolved in 100 µl to 300 µl H_2O. For RNA quality and quantity, the concentration and the 260/280 ratio was measured using a Nano Drop, expecting a 260/280 ratio of 2.0 for RNA.

2.2.8 Inhibition of the Casein Kinase II phosphorylation reaction

CKII Inhibitor I, known as TBB (4,5,5,7-Tetrabromobenzotriazole) and CKII Inhibitor III, also named TBCA (2,3,4,5-Tetrabromophenylacrylic acid) are commercially available from Calbiochem Milipore and were tested in a corresponding CK II phosphorylation reaction. TBB is a cell-permeable compound and an ATP and GTP competitive inhibitor of CKII. TBCA is a cell-permeable compound that acts as a ATP competitive inhibitor of CKII (Fritz et al., 2009; Sarno et al., 2001; Zien et al., 2005). The inhibitors were dissolved in DMSO and a final concentration of 12.5 µg TBB was used. The co-immunoprecipitated RNA was labeled with ^{32}P γ-ATP and dephosphorylated using CIP. Subsequently, western blot and phosphorimage analysis of the inhibited kinase reaction were performed.

2.2.9 Photoactivatable-ribonucleoside enhanced crosslinking and immunoprecipitation

To identify and characterize transcriptome-wide binding sites and RNA recognition elements of RNA-binding proteins, photoactivatable-ribonucleoside enhanced crosslinking and immunoprecipitation (PAR-CLIP) was carried out (Hafner et al., 2010a). The standard PAR-CLIP procedure was performed as shown by Hafner et al. (Hafner et al., 2010b) and described by Spitzer et al. (Spitzer et al., 2014). However, in order to generate successful NCL PAR-CLIP libraries the procedure was optimized and small changes were applied to the protocol. These are illustrated and marked in bold in Figure 14.

In general, HEK293 cells stably expressing the RNA-binding protein of interested with a FLAG/HA tag were incubated for 12-16 h with 4-thiouridine (4SU), a photoactivatable nucleoside, which incorporates into newly transcribed RNA and results in a T to C conversion in the corresponding cDNA. The T to C conversions indicate the

crosslinking and binding sites of the RBP. Cells incubated with 4SU are irradiated with 365 nm UV-light to crosslink the protein-RNA complex followed by cell lysis, two partial RNAse T1 digests, IP of the RBP using an anti-FLAG antibody, size fractionation of the complex, electroelution and isolation of the crosslinked and co-immunoprecipitated RNA, adapter ligation to the 3' and 5' end of the RNA for subsequent PCR amplification, preparation of a cDNA library and sequencing (Illumina), these are the main steps of the standard PAR-CLIP protocol.

Within this thesis, small changes were applied to the original protocol, in order to avoid over digestion of the RNA only one, the first RNAse T1 digest was performed. Over digestion can cause RNA segments, which might be too short to be mapped against the reference transcript. According to Spitzer et al., a second RNAse T1 digest after the IP is performed using 100 U/µl. For the final purification step, instead of the electroelution, the protein-RNA complex within the gel is transferred to a nitrocellulose membrane, which ensures that any remaining free, non-crosslinked RNA passes through the membrane. An additional optional step of re-radiolabeling the purified RNA prior 3'adapter ligation was also added to the original PAR-CLIP procedure. This ensures sufficient visualization of the 3' and 5' ligated RNA products.

4-thiouridine (4SU) (100 µM, 14 h, 37°C, 5% CO_2)
Crosslinking: 365 nm UV (1500 µJ x 100/cm^2)
Cell lysis (2% NP40 lysis buffer)
RNAse T1 digest (1U/µl, 15 min, 22°C)

Immunoprecipitation (0.25 mg/ml anti-FLAG antibody) of the RBP of interest with magnetic Dynabeads (10 µl beads/ml cell lysate)

Calf intestinal alkaline phosphatase (CIP) treatment (0.5 U/µl, 10 min, 37°C, 800 rpm)
γ ^{32}P-ATP labeling (0.1 µCi/µl) with T4 PNK (1 U/µl, 30 min, 37°C, 800 rpm)

SDS-PAGE (Novex Bis-Tris 4-12%, 1 h 15 min, 200 V in 1x MOPS buffer)
Transfer to membrane (Semi-dry, 2 h, 200 mA)

Proteinase-K digest (1.2 mg/ml, 30 min, 55°C)
RNA isolation (acidic phenol/chloroform/isoamyl alcohol)

Optional: prior adapter ligation, relabeling of isolated RNA with γ ^{32}P-ATP (0.1 µCi/µl) and T4 PNK (1 U/µl, 30 min, 37°C, 800 rpm)

3'adapter ligation (100 µM pre-adenylated adapter) using Rnl2 (1-249) K227Q ligase (1 µg/µl, overnight, ice)
15% denaturing urea polyacrylamide gel (UreaGel system, 30W, 45 min in 1x TBE buffer), RNA size fractionation, elution and isolation

5'adapter ligation (100 µM adapter) using T4 RNA ligase 1 (10 U/µl, 1 h, 37°C)
12% denaturing Urea polyacrylamide gel (UreaGel system, 30 W, 45 min in 1x TBE buffer), RNA size fractionation, elution and isolation

Recovery of the 3'- and 5'- adapter ligated products using a complementary 3' PCR primer as a carrier (overnight, 4°C, 800 rpm)

cDNA library preparation;
reverse transcription of the RNA (Superscript III, 42°C, 2h)

Small and large scale PCR amplification (5 U/µl *Taq* polymerase) Purification of the cDNA (QIAquick gel extraction kit, Qiagen) Quality and quantity control (Agilent Bioanalyzer/Tape Station)

cDNA for Next-Generation Sequencing (NGS), Illumina

Figure 14: Overview of the PAR-CLIP procedure

For the identification of cellular RNA targets of RNA-binding proteins, PAR-CLIP is performed and includes the following steps: 4SU incubation of HEK293 cells, UV-crosslinking at 365 nm, cell lysis, FLAG-Immunoprecipitation of the RNA-binding protein of interest, radioactive labeling of the co-immunoprecipitated and crosslinked RNA, size fractionation of the RNA, 3'- and 5'-adapter ligation, cDNA library preparation and subsequent deep sequencing. Small changes compared to the original protocol published by Hafner and Spitzer are marked in bold.

2.2.9.1 4SU labeling and UV crosslinking at 365 nm

Flp-In T-REx HEK293 cells stably expressing NCL were grown in the appropriate selection medium in 15 cm dishes until ~80% confluence. In total, approximately 50 x 15 cm dishes were used to collect a ~5 ml cell pellet. The cells were incubated with a final concentration of 100 µM 4-thiouridine (4SU, $C_9H_{12}N_2O_5S$) for a maximum of 14 h at 37°C in a 5% CO_2 incubator. 4SU is incorporated into newly synthesized RNA. Cells were washed once with ice-cold PBS and crosslinked on ice with UV at 365 nm (1,500 µJ x 100/cm^2). The collected cells were centrifuged at 500 x g for 15 min at 4°C.
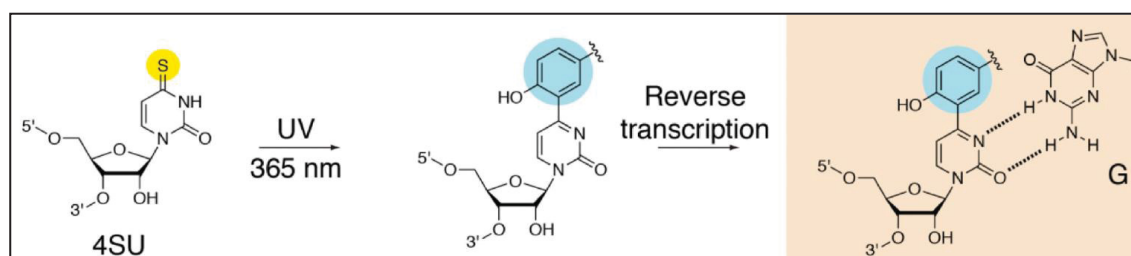
Figure 15: Incubation of 4SU and subsequent UV-crosslinking at 365 nm

4-thiouridine (4SU) is a photoactivatable nucleoside that incorporates into newly transcribed RNA. Upon 365 nm UV-light it induces a change in the conformation and during reverse transcription a G is incorporated instead of an A. This leads to T to C conversions in the cDNA, which indicates the crosslinking sites within the RNA target. Figure: (Ascano et al., 2012).

A structural change upon 365 nm UV-crosslinking of the modified nucleosides to aromatic amino acid side chains like phenylalanine (F), tryptophan (W) and tyrosine (Y) directs the incorporation. The photochemical reactivity of the thio group at position four of the uridine analogue represents increased efficiency of the photo-induced-crosslinking in contrast to the natural nucleoside (Sontheimer, 1994). The chemical structure of the uridine analogue changes, this leads a T to C transitions in the cDNA during the reverse transcription step of the library preparation. Therefore, in a successful PAR-CLIP experiment, the largest fraction of error distance one, which represents one mismatch, is expected to have T to C conversions. Another photoactivatable ribonucleoside, 6-thioguanosine (6SG) incorporates into newly transcribed RNA and also introduces characteristic sequence changes in the

corresponding cDNA. Instead of a T to C conversion obtained by 4SU, it leads to a G to A transition indicating the corresponding crosslinking site and can also be used in PAR-CLIP experiments (Ascano et al., 2012; Hafner et al., 2010a).

2.2.9.2 Cell lysis and Immunoprecipitation

The cell pellet was resuspended in 2% NP40 lysis buffer containing freshly added 0.5 mM dithiothreitol (DTT) and complete protease inhibitor cocktail from Roche. Cells were incubated for 15 min on ice and centrifuged at 13,000 x g for 30 min at 4°C. The lysate was partially digested using RNase T1 with a final concentration of 1 U/µl for 15 min in a water bath at 22°C. RNAse T1 is an endoribonuclease that specifically cleaves single stranded RNA (ssRNA) at G residues. The reaction was immediately cooled on ice for 5 min. The application of an anti-FLAG antibody with a final concentration of 0.25 mg/ml conjugated to magnetic protein G beads (Invitrogen) enables the IP of the FLAG/HA tagged RBP. A volume of 10 µl protein G magnetic beads per ml lysate was used. The anti-FLAG antibody and beads were incubated in PBS for 45 min at RT on a rotating wheel. The lysate was incubated with the antibody conjugated beads for 2 h at 4°C on a rotating wheel. After the IP, the beads were washed three times with IP-wash buffer, three times with high-salt wash buffer and were subsequently resuspended in dephosphorylation buffer in order to dephosphorylate the crosslinked RNA.

2.2.9.3 Dephosphorylation and radiolabeling of the crosslinked RNA

Prior radiolabeling the 5' end of the RNA using γ ^{32}P-ATP, the RNA was dephosphorylated using calf intestinal alkaline phosphatase (CIP). CIP catalyzes the dephosphorylation of the 5' end and 3' end of RNA phosphomonoesters (NEB). The immunoprecipitated and crosslinked RNA was incubated in NEB buffer 3 with 0.5 U/µl CIP for 10 min, at 37°C, shaking at 800 rpm. The beads were then washed twice with phosphatase buffer and twice with polynucleotide kinase (PNK) buffer. The beads were

resuspended in PNK buffer containing freshly added DTT and γ ^{32}P-ATP was added to a final concentration of 0.1 µCi/µl. A concentration of 1 U/µl of T4 PNK was added to the reaction that catalyzes the transfer of the P_i from the γ position of the ATP to the 5' hydroxyl terminus of the RNA and removes the 3' phosphoryl group of the RNA (NEB). The labeling reaction was conducted at 37°C for a minimum of 30 min at 800 rpm. To chase the radioactive ATP labeling reaction, 100 µM non-radioactive ATP was added to the mixture and incubated for 5 min to ensure complete 5' phosphorylation. This step is required for further 5'adapter ligation. The beads were washed five times with PNK buffer (without DTT), resuspended in 65 µl 1x SDS loading buffer and incubated for 5 min at 90°C to elute the immunoprecipitated and crosslinked protein-RNA complex. The supernatant was transferred to a clean tube.

2.2.9.4 Radiolabeling of the 5' end of the RNA size markers using γ ^{32}P-ATP

For size fractionation of the crosslinked RNA, radiolabeled RNA size markers with a length of 19 nt, 24 nt and 35 nt were used. These RNA size markers were radiolabeled according to the standard protocol using T4 PNK and γ ^{32}P-ATP. The size markers were separated on a urea polyacrylamide gel in 1x TBE buffer and purified and eluted overnight in 0.4 M NaCl, at 4°C, shaking at 800 rpm. The next day, the supernatant containing the radiolabeled RNA was precipitated in 100% ethanol.

2.2.9.5 SDS-PAGE and transfer to a nitrocellulose membrane

The protein-RNA complex was released from the beads and loaded on a precast (4% to 12%) Bis-Tris gradient gel (Invitrogen). The gel run using 1x MOPS running buffer (Invitrogen) at 200 V for 75 min. Subsequently, the protein complex was transferred to a nitrocellulose membrane in a semi-dry blot chamber with 1x transfer buffer for 2 h at 200 mA. Prior exposing the complex to a phosphorimage screen, the protein ladder on the membrane was marked with the liquid radioactive waste collected from the

previous washing steps. Radioactive gel pieces were placed in three corners of the membrane to facilitate the alignment of the phosphorimage printout to the membrane. The radioactive gel pieces were collected from the free γ ^{32}P-ATP from the previous gels, either from labeling the RNA size markers or from the SDS gel. The membrane was then exposed for at least 30 min to a phosphorimage screen. The radioactive bands at the size of the corresponding molecular weight of the RBP representing the crosslinked RNA-protein complex were cut out and used for further digestion steps.

2.2.9.6 Proteinase-K digest and RNA isolation

The excised protein-RNA bands from the membrane were cut into little pieces in order to digest the complex with Proteinase-K. The radioactive membrane pieces were resuspended in 1x Proteinase-K buffer with 1.2 mg/ml freshly added Proteinase-K. Proteinase-K is a serine protease that hydrolyzes a variety of peptide bonds and cleaves the carboxylic sides of aliphatic, aromatic or hydrophobic amino acids. The protease is also applied in nucleic acid purification steps and also digests unwanted nucleases (NEB, Sigma and Thermo Scientific). The mixture was incubated at 55°C for 30 min shaking at 800 rpm. The supernatant was then transferred to a new tube, the RNA was recovered and the Proteinase-K was denatured using acidic phenol/chloroform/isoamyl alcohol. After a centrifugation step at 20,000 x g for 15 min at 4°C the RNA was precipitated adding 1/10 volume of 3 M NaCl, 1 µl glycogen (10 mg/ml) and 3 volumes of 100% ethanol. The RNA was dissolved in 10 µL H_2O. Optional, the RNA can be re-labeled using T4 PNK and γ ^{32}P-ATP according to the previously described procedure or can be directly applied for further adapter ligation.

2.2.9.7 The 3'- and 5'-adapter ligation

The 3'- and 5'-adapter oligonucleotides serve as a primer-binding site for the reverse transcription (RT) followed by PCR amplification and subsequent deep sequencing.

The small cDNA library preparation comprises two adapter oligonucleotides ligation steps. For each step, a different RNA ligase was utilized. The first step includes the ligation of an oligonucleotide adapter to the 3′ end of the RNA using T4 RNA ligase 2 K227Q (Rnl2, 1-249). RNA ligase Rnl2 ligates the 5′-p terminus of the donor RNA to the 3′-hydroxyl (OH) terminus of the acceptor RNA (Hafner et al., 2008; Hafner et al., 2011). Unwanted side reactions, like circularization of the RNA and the 3′ adapter can occur during this step. To prevent circularization, the 3′ adapter is chemically synthesized carrying a pre-adenylated oligonucleotide. Side reactions of the RNA can also be reduced by a dephosphorylation step using CIP, previous to the 3′ adapter ligation and re-phosphorylation for the 5'adapter ligation. Additionally, to limit side reactions such as circularization, a truncated form of RNA ligase 2 (Rnl2), the so called Rnl2 (1–249) was applied. Rnl2 (1–249) has a reduced affinity to 5′-pRNA donors resulting in a less efficient adenyl transfer from the adenylated enzyme to the donor 5′-pRNA. The mutation of the lysine residue K227 to glutamine in Rnl2 (1–249) K227Q compromised the adenyltransferase activity, while it still retains the activity of the ligation step (Hafner et al., 2008; Hafner et al., 2011). The 3'adapter ligated products were separated on a 15% denaturing 8 M urea polyacrylamide gel (UreaGel system from National Diagnostics) for 45 min at 30 W in 1x TBE buffer.

The second adapter-ligation step comprises the linkage of the 3′-OH of the 5′-oligonucleotide adapter to the 5′ end of the RNA, which is already ligated to the 3′ adapter. The 5' ligation carries no side reactions. Therefore, conventional ligation with T4 RNA ligase 1 (Rnl1) and γ ^{32}P-ATP can be used (Hafner et al., 2008; Hafner et al., 2011). The 5'adapter ligated products were separated on a 12% denaturing 8 M urea polyacrylamide gel (UreaGel system from National Diagnostics) for 45 min at 30 W in 1x TBE buffer. After each adapter ligation, the ligated products were eluted from the gel in 350 µl (0.4 M) NaCl overnight. The RNA was precipitated in 100% ethanol. The 3'- and 5'-adapter ligated RNA products can be applied for subsequent cDNA synthesis.

2.2.9.8 cDNA library preparation

After 3′ and 5′ adapter ligation of the RNA, the cDNA library was generated using Superscript III reverse transcriptase and a primer that is complementary to the 3′ adapter sequence. Secondary structures of the small RNAs like stem-loop structures, can inhibit uniform RT of all types of the small RNA pool. In order to reduce these effects, a thermostable RT enzyme, like the Superscript is preferably utilized for this reaction. The 3' and 5' ligated products were previously eluted in 0.4 M NaCl and 1 µl PCR primer (100 µM). The primer serves as a carrier to facilitate the recovery and is necessary for further RT-PCR. The Superscript mix (1.5 µl 0.1 M DTT, 3 µl 5x first strand buffer, 4.2 µl 10x dNTPs) was prepared for each sample. The total reaction mix of 8.7 µl was added to 5.6 µl denatured RNA in H_2O. The RNA was incubated at 50°C for 3 min. Subsequently, 0.75 µl of the Superscript enzyme was added and the reaction and was incubated for 2 h at 42°C. The reverse transcribed cDNA was diluted in 80 µl H_2O. A final PCR amplification step is required for cDNA library preparation for deep sequencing. To determine the minimum number of PCR cycles for the large scale PCR, 10 µl of the cDNA were used for the small scale or pilot PCR.

Table 16: PCR cycles for cDNA library preparation

PCR step	Temperature	Duration
Denaturation	94°C	45 sec
Annealing	50°C	85 sec
Elongation	72°C	60 sec

The PCR product was precipitated in NaCl and 100% ethanol, followed by a centrifugation step at 20,000 x g, at 4°C for 30 min. The pellet was dissolved in 1x DNA loading dye in TBE and separated on a 2.5% low-melting agarose gel for 2 h 30 min at 120 V. The separated cDNA was excised at the expected size of 95 nt to 110 nt and undesired 5'adapter - 3'adapter PCR products were removed. The cDNA was purified from the gel using the QIAquick gel extraction kit from Qiagen and eluted in 30 µl in

elution buffer. Quantity and quality of the purified cDNA was controlled with the Bioanalyzer or TapeStation. Subsequently, the cDNA library was submitted for Illumina HiSeq 2000 sequencing to the Genomics Resource Center (Rockefeller, University).

2.2.9.9 Analysis of the PAR-CLIP library

The raw Illumina sequencing data comprise a FASTAQ format that includes the sequencing reads and quality scores. The first processing step of the raw PAR-CLIP data are trimming the 3'- and 5'-adapter sequences, which were ligated during the experimental PAR-CLIP procedure. The adapter sequences were removed with cutadapt (Marcel, 2011). The data are then converted into a FASTA format. The reads were mapped and aligned to the human transcriptome and hg19 using Bowtie. Bowtie is an ultrafast, memory-efficient alignment program, which aligns short DNA sequence reads to large genomes (Langmead et al., 2009). Processing and annotation of clusters to the genome was performed using the PARalyzer software (Corcoran et al., 2011). The method of PAR-CLIP is characterized by the incorporation of 4SU into newly transcribed RNA that causes a T to C conversion in the corresponding cDNA. The transition refers to the crosslinked binding site of the RNA. The frequency of the T to C conversions, resulting from crosslinked RNA, can be distinguished from non-crosslinked RNA that lack these conversions and represent background. The remaining sequence reads were aligned to the reference hg19 and clustered according to the number of reads. The overlap between each other and the amount of T to C conversions within one cluster were analyzed. To fulfill optimal requirements, the minimum number of reads within a cluster should be five and usually 20% of the reads in one cluster should contain T to C conversions. Computational tools like PARalyzer and Bowtie align these sequence reads and allow one or two mismatches in total (Corcoran et al., 2011). We refer to these mismatches as distance (d). No mismatch refers to distance 0 (d=0), one mismatch to distance 1 (d=1) and two mismatches to

distance 2 (d=2). Since the primary 47S rRNA transcript is not completely assembled in the hg, critical exceptions in processing and analyzing the data for pre rRNA need to be considered. Data analysis and bioinformatic support was provided by A. Garcia and P. Morozov from the Tuschl lab, Rockefeller University, NY.

2.2.10 RNA Immunoprecipitation-sequencing

Immunoprecipitation of a RBP, followed by RNA extraction and high-throughput RNA sequencing (RNA immunoprecipitation-sequencing), determines the expression of RNA targets that interact with the RBP. A non-immunoprecipitated lysate sample serves as a control. The protein-RNA complex is not previously crosslinked with UV. The RBP was immunoprecipitated using an anti-FLAG antibody conjugated to magnetic Dynabeads. The RNA associated with the RBP was directly isolated and eluted using TRIzol. The RNA was purified with the RNeasy purification kit from Qiagen and subsequently submitted for HiSeq sequencing by Illumina. Since no photoactivatable ribonucleoside like 4SU or 6SG was utilized, a specific binding site cannot be analyzed and identified (Ascano et al., 2013).
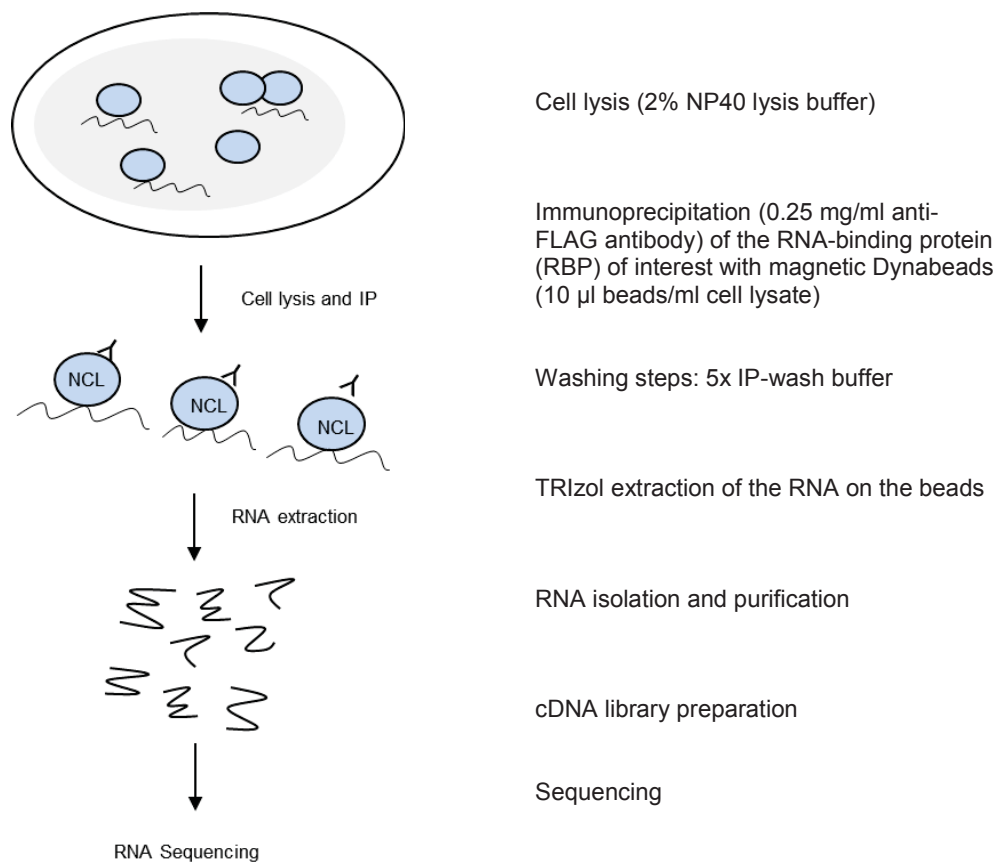
Figure 16: Overview of the RIP-Seq procedure

For the identification of cellular RNA targets of RNA-binding proteins, RIP-Seq is performed and includes the following steps: cell lysis, FLAG-immunoprecipitation of the FLAG/HA tagged RBP of interest, TRIzol extraction and isolation of the RNA, cDNA library preparation and subsequent deep sequencing.

2.2.10.1 Cell lysis and Immunoprecipitation

The cells were lysed and immunoprecipitated as previously described in the following section 2.2.9.2. However, the cells were not digested and treated with RNAse. After the IP, the beads were washed five times with IP-wash buffer prior RNA extraction using TRizol.

2.2.10.2 RNA extraction, library preparation and Next-Generation Sequencing

The co-immunoprecipitated RNA was extracted and isolated from the beads using 1 ml TRIzol. According to the manufacture's instruction, 200 µl chloroform per 1 ml TRIzol was added for subsequent RNA extraction. TRIzol and chloroform was also added to

the lysate control. A more detailed procedure of the RNA extraction using TRIzol is described in the previous chapter 2.2.5. The upper aqueous phase containing the RNA was transferred and purified using the miRNeasy kit from Qiagen. A quality check of the extracted RNA was performed with the Bioanalyzer. Subsequently, the RNA was submitted for deep sequencing by Illumina. The cDNA was barcoded with the Illumina multiplexing sample preparation oligonucleotide kit and analyzed on an Illumina HiSeq 2000 in a 100 base-pair single-end sequencing run by the Genomic Resource Center (Rockefeller, University).

2.2.11 DREME: A motif discovery algorithm to identify binding motifs

To further elucidate and identify preferential binding motifs of NCL, the discriminative regular expression motif elicitation (DREME) was utilized. DREME is a motif discovery algorithm designed by T. L. Bailey. Computational RBP motif discovery is often performed by DNA motif finding methods (Li et al., 2014) like DREME, which was recently generated in order to find short, core DNA-binding motifs of eukaryotic transcription factors.

The computational tool discovers short, ungapped motifs of 4 nt up to 8 nt in length that are enriched in the positive input sequence set compared to a negative input control sample. DREME is able to create the negative control set by shuffling the sequences of the primary input sample set. However, a negative control sample set can also be generated independently from DREME, then, the control set should contain the same length distribution as the positive set. Another PAR-CLIP library, like the empty vector data set could also serve as a control sample, as long as it has the same length distribution. In order to statistically determine the significance of a discriminative motif, DREME uses the Fisher's Exact test. The Fisher's Exact test computes the significance of relative enrichment of each motif in two given sets of sequences and uses the

analysis of 2 x 2 contingency tables. DREME requires a FASTA format of the sequences (Bailey, 2011). The motif discovery algorithm, DREME is available as part of the MEME Suite of motif-based sequence analysis tools (http://meme-suite.org/tools/dreme). The NCL motif analysis of the PAR-CLIP data using DREME version 4.9.1 was supported by R. Dellen.

2.2.12 Biochemical and structural binding studies

Two different NCL constructs containing either the first two RNA recognition motifs (RRM) or all four, respectively named NCL RRM1-2 and NCL RRM1-4 (Figure 17) were generated using the GenScript CloneEZ PCR Cloning Kit (GenScript).

NCL RRM1-2 (298-468, 19 kDa)
NCL RRM1-4 (298-648, 39 kDa)
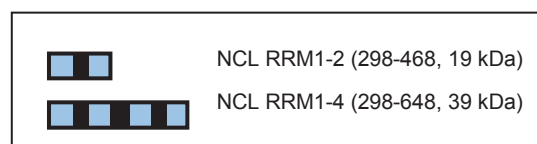
Figure 17: Recombinant Nucleolin constructs

Recombinant Nucleolin constructs containing the first two RRMs (RRM1-2) and all four RRMs (RRM1-4) were used for biochemical binding studies in electrophoretic mobility shift assays (EMSA), gel-filtration chromatography (GFC), also known as size-exclusion chromatography (SEC), isothermal-titration calorimetry (ITC) and crystallography of Nucleolin.

The constructs were cloned into a pET-28 vector containing a 6 His-SUMO (small ubiquitin-related modifier) affinity tag at the N terminus. The affinity tags were utilized for further purification on an Äkta using a nickel-column, which binds the SUMO tag and a Q-column that clears the constructs from interacting nucleic acid. The tags were finally cleaved with Proteinase-K. The NCL RRM1-2 construct was additionally purified and cleared from nucleic acids with a heparin column. In a final step, a gel-filtration column (Superdex) was used to purify the constructs. The NCL constructs were applied for further biochemical binding analysis including gel-filtration chromatography (GFC), also known as size-exclusion chromatography (SEC), electrophoretic mobility shift

assay (EMSA), isothermal-titration calorimetry (ITC) and crystallography in order to determine and confirm binding affinity and structural analysis of the NCL-RNA complex. EMSA were performed in the lab in Düsseldorf (Department of Pediatric Oncology, Hematology and Immunology). Biochemical and structural studies like ITC, SEC and crystallography of the NCL-RNA complex were performed with M. Teplova from the Memorial Sloan Kettering Cancer Center (MSKCC). Material and methods for the biochemical and structural studies were provided by the Patel lab at the MSKCC, in New Yok.

2.2.12.1 Electrophoretic mobility shift assay

Electrophoretic mobility shift assays were used to verify binding of the protein-RNA complexes. In conventional assays, proteins and nucleic acids are combined and the resulting mixtures are subjected under native conditions using agarose or polyacrylamide gel electrophoresis (Hellman and Fried, 2007). The non-denaturing gel for the NCL-RNA complex is composed of polyacrylamide in TBE-buffer. The gel pre-run in 0.25x TBE buffer for approximately 45 min at 300 V. To prevent interference, the complexed samples were not mixed with loading dye. The RNA oligos were labeled with γ ^{32}P-ATP as described in more detail in the previous section 2.2.6.

The complex formation of various RNA oligos with recombinant Nucleolin (NCL delta N, NCL RRM1-2, NCL RRM1-4) was performed for at least 45 min on ice with a total volume of 20 µl in siliconized 1.5 ml reaction tubes. A protein concentration with a wide range of 0 nM up to 1000 nM and a RNA concentration with a range of 2.5 nM to 500 nM was tested. Either a non-radioactive RNA oligo, poly (A) oligos or yeast tRNA were used as a competitor in the binding reaction.

Table 17: Nucleolin-RNA complex formation reaction in electrophoretic mobility shift assays

Reagent	Concentration/Volume
γ ^{32}P-ATP labeled RNA oligo	2.5-500 nM
recombinant NCL	0-1000 nM
DTT	1 mM
RNasin ribonuclease inhibitor	1 U/µl
yeast tRNA	100-1000 ng
1x EMSA buffer	final volume: 20 µl

After complex formation, the samples were loaded to the gel and run at ~260 V at RT for 75 min. The gel was dried in a vacuum-dryer at 70°C for 1 h and finally exposed to a phosphorimage screen for 1 h at RT or at 4°C overnight.

2.2.12.2 Isothermal-Titration Calorimetry

Isothermal-titration calorimetry (ITC) is a biophysical technique to detect the thermodynamic interaction of an interactive system. It allows the characterization of the binding process between a protein (P) and a small molecule, ligand (L) (Figure 18).

$$P + L \rightleftharpoons PL$$

$$K_d = \frac{[P]*[L]}{[P*L]}$$

P = protein, L = ligand
PL = protein-ligand complex
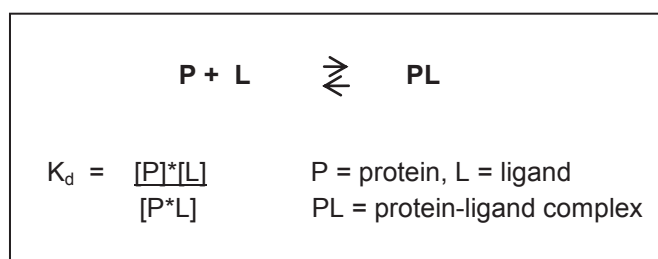
Figure 18: The dissociation constant (K_d)

The dissociation constant describes a simple and reversible binding reaction of a Protein (P) and a Ligand (L) forming a Protein-Ligand complex (PL) and gives information about the binding affinity of a protein to a ligand (Damian, 2013).

The binding reaction of a protein to the ligand results either in heat absorbance or heat release. ITC uses heat as a signal to determine the stoichiometry (N), the affinity constant (K_a) and the enthalpy changes (ΔH) of the complex. The requirements for ITC measurements are correctly folded and an active protein, and accurate determination

of the protein and the ligand concentration. It is very important that the protein is as pure as possible, therefore, it is recommended to dialyze the samples prior ITC measurements. The ligand should always be dissolved in the same ITC buffer containing the same pH and salt concentrations as the protein. Otherwise, impure or inactive compounds or inaccurate concentration can have a strong impact on the determination of the ITC parameters (ΔH, N, K_a) (Damian, 2013; Feig, 2007). The stoichiometry of the complex reveals the average number of protein binding sites in the solution. If N is equal to 1 it refers a 1:1 binding ratio of protein to ligand (RNA), if N is equal to 0.5 it refers to a 2:1 binding ratio. The dissociation constant (K_d) is the inverse of the affinity constant and is also known as the association constant. It describes the affinity between the protein and the ligand. A K_d in the nM to pM range is characterized as a specific binding, whereas a K_d in the mM range indicates weaker interaction of the complex (Damian, 2013; Wiseman et al., 1989).

In this thesis, the binding affinity of NCL (NCL RRM1-2 and RRM1-4) and a RNA target was analyzed by ITC. For the measurements, a MicroCal iTC$_{2000}$ was used (Malvern). In a standard ITC experiment, a known concentration of RNA is titrated to a known concentration of recombinant protein. Usually, if the K_d of the complex is unknown a ratio of 1:10 of protein to the ligand (RNA) is recommended. The RNA was injected to the syringe, the protein to the sample cell and both compounds were dissolved in same ITC buffer. The binding reaction was performed at 25°C. For the first measurements, 0.2 mM RNA and 0.02 mM recombinant NCL was used. For further experiments, 0.06 mM and 0.02 mM of recombinant NCL RRM1-2 and RRM1-4, respectively was injected to the cell, whereas 0.32 mM RNA was used. Each peak in the raw ITC data marks a titration point, indicating automatic titration of RNA to protein and resulting in complex formation. The injection volume can vary between instruments. In our case, 1.6 µl RNA was injected to the sample cell at each titration point for 25 times. At the beginning of the experiment there is an excess of protein, consequently the injected RNA is

completely bound by the protein. The RNA was titrated gradually to the protein until the binding equilibrium is reached. Negative peaks indicate an exothermic reaction (heat release), whereas positive peaks represent an endothermic reaction (heat absorption). If a complex formation shows endo- and exothermic reactions, it most likely refers to a two-site model. Prior testing the binding affinity of the complex, ITC buffer was titrated to the protein and to the RNA to determine the background. Small peaks, implicating no binding, were subtracted from the real binding peaks in the analysis.

2.2.12.3 Size-Exclusion Chromatography

Size-exclusion chromatography (SEC) is a method to separate proteins and other compounds according to their molecular size by filtration them through a gel (Hong et al., 2012). Proteins with high molecular weight are not able to enter the pores of the gel and consequently elute first. Compounds with lower molecular weight are able to enter the gel, penetrate through the pores and elute in later fractions.
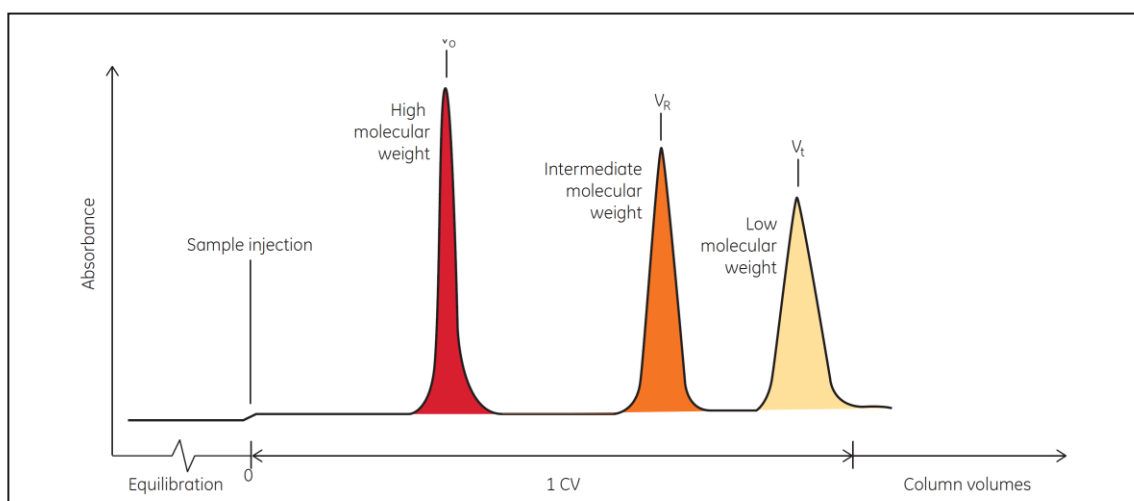
Figure 19: Size-Exclusion Chromatography

The size-exclusion chromatography diagram shows the sample injection and the corresponding elution of the sample ranging from high molecular weight to low molecular weight according to absorbance and column volumes. Figure: GE Healthcare (Size-Exclusion Chromatography, Principles and Methods).

In this thesis, the NCL-RNA complex was separated using an analytic Superdex 75 HR 10/30 column in SEC buffer. The Superdex 75 is a prepacked column for high performance gel filtration of proteins and nucleic acids in the molecular weight range of 3 to 70 KDa (Amersham Bioscience). First, recombinant NCL either NCL RRM1-2 (19 kDa) or NCL RRM1-4 (39 kDa) was loaded to the column and eluted according to the molecular weight. The 25 nt RNA sequence (7 kDa) (oligo 1) was separately loaded to the column and size eluted. The eluted fractions of the free protein and the free RNA were collected. A total volume of 40 µl RNA (88 µM), 5 µl RRM1-2 or RRM1-4 and 160 µl ECM-buffer were incubated for complex formation. Subsequently, the complex was loaded to the column and size fractioned.

2.2.12.4 Crystallography

For structural analysis of the NCL-RNA complex, crystallography was carried out. Several conditions including different salt concentrations, pH and buffer composition, various protein concentrations and ratios of RNA-protein were tested. First, a 1:1 ratio of protein to RNA with a protein concentration of 10 mg/ml was analyzed. In a second step, the protein concentration was increased up to 20 mg/ml in a 1:1 and 2:1 ratio for RRM1-2 and RRM1-4, respectively. Several different crystallization conditions using commercial available kits (Qiagen) like Classics, Classics II, JCSG Core I, JCSG Core II, PEG I, PEG II, PACT and Cove II were tested. To avoid precipitation and formation of salt crystals, first the protein and the buffer were mixed and the RNA was added at the end. A total volume of 100 µl is needed to pipette a single drop of the complex to each well of a 96 well plates containing different crystal-condition from the commercially available kits. A Mosquito robot pipetted the sample into the wells of a 96 well plate.

2.2.13 Mass-Spectrometry

In proteomics, mass-spectrometry (mass-spec or MS) is a very sensitive and powerful quantitative technique that determines protein-protein interactions and enables to distinguish real interactors from background (Keilhauer et al., 2015). The analysis of protein-protein interactions is very important for the understanding of biological processes within a cell. The MS analysis was essentially performed in order to identify and characterize protein interaction partners of NCL. Within this thesis, a single-step affinity enrichment was carried out in Flp-In T-REx HEK293 cells stably expressing FLAG/HA tagged bait proteins like NCL, NPM1, Y Box Binding Protein 1 (YBX1) or an empty vector control, that lacks the bait protein. The IP samples were used for further downstream analysis in label-free quantitative (LFQ) LC-ESI tandem MS/MS. NPM1, a nucleolar protein that is also involved in ribosome biogenesis, serves as a positive control, while YBX1 and the empty vector function as negative controls.

Cell lysis (2% NP40 lysis buffer)

Immunoprecipitation of the protein of interest (0.25 mg/ml anti-FLAG antibody) with magnetic Dynabeads (10 µl beads/ml cell lysate)

Cell lysis and IP

anti-FLAG antibody

Washing steps: 3x IP-wash buffer (with NP40), 2x IP-wash buffer (without NP40)

Elution: FLAG peptide (2 h, 4°C)

Elution with a FLAG-peptide

SDS-PAGE: Novex Bis-Tris (4% - 12%, 1 h 15 min, 200 V) in 1x MOPS buffer

Silver staining

Silver staining of the protein complex

Bands were cut out, destained and digested with trypsin

Proteolysis of proteins

Peptide separation using an Ultimate 3000 Rapid separation liquid chromatography system

Injection of the peptides to the liquid chromatography and separation on an analytical column

Mass-Spectrometry analysis of peptides

Mass-spectrometry: Orbitrap Elite high resolution instrument equipped with a nano electrospray ionization source

Mass-spectrometry analysis: quantification and identification of the protein-protein interaction

Data analysis

Figure 20: Overview of the Mass-Spectrometry procedure

Mass-spectrometry includes a FLAG-immunoprecipitation of the protein of interest expressed in HEK293 cells, elution with a corresponding FLAG-peptide, SDS-PAGE and silver staining of the protein complex, digest of the protein using trypsin and peptide separation on an Ultimate 3000 Rapid Separation liquid chromatography system. The Orbitrap elite high resolution instrument is equipped with a nano electrospray ionization source. The identified proteins were analyzed using SAINT*express* (Significant Analysis of INTeractome express).

2.2.13.1 Cell lysis, Immunoprecipitation and FLAG-peptide elution

FLAG/HA-tagged NCL stably expressed in HEK293 cells were lysed and immunoprecipitated as previously described in section 2.2.9.2. The beads were washed three times with IP-wash buffer containing NP40 and twice with IP-wash buffer without NP40 to exclude any side effects in the downstream mass-spectrometry analysis. The bound proteins were eluted under non-denaturing conditions using 3x FLAG-peptide, the beads were collected in 42 µl PBS, 40 µg FLAG-peptide was added and the mixture was incubated for 2 h on a rotating wheel at 4°C. The supernatant, containing the immunoprecipitated sample was applied for further downstream analysis.

2.2.13.2 SDS-PAGE and silver staining

The IP samples were diluted in 4x SDS-PAGE loading buffer and separated on a NuPAGE Novex Bis-Tris 4% to 12% gel and silver stained using the basic staining protocol according to the manufactory manual (Invitrogen). Prior staining, the gel was washed once in ultrapure water and fixed in 40% ethanol, 10 % acetic acid and H_2O at RT for either 20 min or overnight. The fixation step removes interfering ions and detergents from the SDS-gel and reduces the movement of the proteins in the gel. After fixation, the gel was washed 10 min in 30% ethanol to remove excess of the fixative. The gel was incubated for 10 min in the sensitizing solution containing 30% ethanol, 10% sensitizer and H_2O. This incubation step increases the sensitivity and the contrast of the subsequent silver staining. The gel was subsequently washed 10 min in 30% ethanol, then in ultrapure H_2O to remove excess of the sensitizer solution and to rehydrate the gel for further staining. The gel was incubated for 15 min in the staining solution containing 1% stainer. The staining solution binds the silver ions to the proteins. After the staining procedure, the solution was decanted and the gel was washed for 30 to 60 sec in H_2O in order to remove excess of the silver ions prior

incubation in the developing solution. The developing solution contained 10% developer, 1 drop of developer enhancer and H_2O and was applied to the gel in order to reduce silver ions to metallic silver at the protein bands. The silver stained protein bands became visible. Subsequently, the stopper solution complexes with free silver ions and prevents further reduction and stops the reaction (Invitrogen). The silver stained SDS gel serves as a control for equal protein concentration, used as an input for further downstream analysis.

2.2.13.3 Liquid chromatography-electrospray ionization Mass-Spectrometry

Mass-spectrometry of four different IP samples (groups) including NCL, NPM1, YBX1 and an empty vector control, each with five biological replicates per group (n=5) was performed. IP-samples were digested with trypsin, the resulting peptides were separated using a liquid chromatography coupled electrospray ionization mass-spectrometry (LC-ESI MS). First, the IP samples were shortly separated (5 mm) in a precast 4-12% polyacrylamide Bis-Tris gel. Protein bands were silver stained, cut out and destained in destaining buffer, reduced in DTT and ammonium bicarbonate (reducing buffer) and finally alkylated (alkylation buffer). The proteins were digested in 0.1 µg trypsin, overnight. Trypsin is mainly used in proteomic based approaches for the digestion of proteins to analyzable peptides. The protease cleaves exclusively after arginine or lysine residues (Olsen et al., 2004). The tryptic digest was stopped with trifluoroacetic acid (TFA) and the resulting peptides were extracted from the gel and separated on an Ultimate 3000 Rapid Separation liquid chromatography system (Dionex/Thermo Scientific). Within a next step, the peptides were reconstituted in 0.1% TFA, injected to the instrument and pre-concentrated on an Acclaim PepMap100 trap column (3 µm C18 particle size, pore size of 100 Å, inner diameter of 75 µm, 2 cm in length) for 10 min. A flow rate of 6 µl/min using 0.1% (v/v) TFA as the mobile phase. Subsequently, the peptides were separated on an analytical column (Acclaim

PepMapRSLC, 2 µm C18 particle size, 100 Å pore size, 75 µm inner diameter, 25 cm length, Dionex/Thermo Scientific) at 60°C by a 2 h gradient from 4-40% solvent B (solvent A: 0.1% (v/v) formic acid in H_2O, solvent B: 0.1% (v/v) formic acid, 84% (v/v) acetonitrile in H_2O) at a flow rate of 300 nl/min. In a next step, the separated peptides were directly injected in an online coupled Orbitrap elite high resolution hybrid mass-spectrometer (Thermo Scientific) with a nano electrospray ionization source equipped with a distal coated silica emitter. The mass-spectrometer was used in a positive mode, capillary temperature was set to 275°C and a voltage of 1.4 kV was applied. First, survey scans were carried out in the Orbitrap analyzer over a mass range of 350-1,700 m/z (mass-to-charge ratio) at a resolution of 60,000 (at 400 m/z). The target value for the automatic gain control was set to 1,000,000 and a maximum (max) fill time of 200 ms was applied. The 20 most intense peptide ions with a minimum signal intensity of 500 were isolated, transferred to a linear ion trap and fragmented with collision induced dissociation. Subsequently, peptide fragments were analyzed with a max fill time of 300 ms and automatic gain control target value of 10,000. The available mass range was set to 200-2,000 m/z at a resolution of 5,400. Once ions were fragmented, they were excluded for 45 sec from the next fragmentation step. Material and methods were performed and provided by G. Poschmann from the Proteomic facility at the BMFZ in Düsseldorf.

2.2.13.4 Data analysis of Mass-Spectrometry using SAINT*express* and SAM

The aim of this mass-spec data analysis is the identification of specific NCL and NPM1 co-purified proteins across four different groups (NCL, NPM1, YBX1, empty vector), each with five biological replicates. Affinity-purified proteins, known as baits, interact with proteins commonly referred to as preys (Nesvizhskii, 2012). The major computational challenge in analyzing MS data remains in the identification of specific protein interactors. The simplest strategy represents a subtractive approach of the

data, in which all proteins identified within a negative control experiment were subtracted from the list of specifically identified co-purified proteins. Usually, a negative control contains the affinity tag (FLAG/HA), while it lacks the expression of the bait protein (empty vector). However, this data analysis would not use any statistical modeling. Another challenge in data analysis remains in the diversity of the data, depending on the number of baits, negative controls, replicates, interconnectivity of the baits and the purification-affinity procedure (Nesvizhskii, 2012). Therefore, it is critical to choose the right strategy for MS data analysis aiming to distinguish true binding partners from false-positives, noise and background.

Within this thesis, the first filtering steps of the data set comprise a minimum of two identified peptides and five quantitative values in at least one of the four sample groups in order to reduce the list of obvious non-specific interactors and false-positive proteins. Database protein searches and quantification was performed using the software MaxQuant (version 1.5.2.8) (Cox et al., 2014) enabling database searches of spectra, intensity determination and normalization of intensity data. Data analysis was carried out in the 20,200 human Swiss-Prot entries (UniProt Knowledgebase release 2015_01). The false discovery rate (FDR), which estimates the ratio of proteins and peptides that were likely identified as false-positives by chance (Larsson et al., 2005) was set to 1% on peptide as well as on protein level. Tryptic cleavage specificity was considered after any K or R residue (Olsen et al., 2004) and a max of two missed cleavage sites was allowed.

In a second step, the remaining proteins were included for further analysis using the computational tool significant analysis of INTeractome *express* (SAINT*express)* (version 3.6.1), a software, which is able to handle multiple testing problems. The software is available online (http://saint-apms.sourceforge.net). SAINT assumes that a prey protein (interactor), identified within an affinity-purified IP sample is either a true or a non-specific interactor of the bait protein (NCL, NPM1) (Teo et al., 2014). SAINT

builds a probability-based model that utilizes label free quantification information for each bait-prey pair to assign confidence scores (Choi et al., 2011).

Spectral counts, a parameter that is conducted from the mass-spec data are defined as the total number of spectra that were identified for a protein in proteomic based approaches (Lundgren et al., 2010). In SAINT, spectral counts are normalized to the length of the protein and to the total number of spectra. Upon normalization, spectral counts can therefore be used as a quantitative value to measure relative protein abundance within the affinity purified IP sample (Choi et al., 2011). Normalized spectral count distributions are compared for each prey across all purified samples. SAINT utilizes the spectral counts for each prey that is identified in a bait sample (NCL and NPM1) and computes the probability of a true prey (i) - bait (j) interaction ($true$|X$_{ij}$) using the Bayes rule (Choi et al., 2011; Nesvizhskii, 2012). Spectral counts for each interaction pair (bait-prey) are modeled with a mixture distribution for true and false binding partners, which are specific for each bait-prey pair. Probability distributions of (X_{ij}|$true$) and (X_{ij}|$false$) are utilized to compute the probability of true interactors ($true$|X$_{ij}$) (Choi et al., 2011; Teo et al., 2014). To filter false-positives and non-specifically binding proteins remains a computational challenge in mass-spec data analysis. The statistical model of SAINT is a two-component mixture model that considers negative controls, which are generated during the experimental procedure, like the empty vector and the YBX1 control. Using negative controls, the method estimates the spectral count distribution for false interactions (X_{ij}|$false$) directly from these controls. This model refers to a semi-supervised approach (Choi et al., 2011; Choi et al., 2012b). In general, true interactions are present in significantly higher levels than within the negative control samples and contain higher spectral counts in a small number of baits (Teo et al., 2014). SAINT also estimates a Bayesian false discovery rate (BFDR) using the probabilities that have been calculated for all prey-bait pairs (Choi et al., 2011; Teo et

al., 2014). Within this thesis, preys were accepted with a BFDR of 5% and were defined as significantly co-purified proteins, which were used for further data analysis.

The SAINT probability score computes the probability of the binding between a prey-bait pair in the presence of spectral counts. The higher the SAINT probability score, the higher the probability of a true prey-bait pair. Additionally, SAINT also provides a fold change (FC) of the spectral counts, which is defined as the average spectral count of the tested sample (NCL) divided by the average spectral counts of the negative controls (empty, YBX1) (Choi et al., 2011; Teo et al., 2014). If the spectral counts are zero within a negative control, the value is replaced by 0.1 (Teo et al., 2014).

Furthermore, since NCL and NPM1 are both associated with ribosome biogenesis and share common binding partners, a significance analysis of microarrays (SAM) method was performed in order to determine significant binding partners only for these two baits. SAM is a statistical method that was originally established by Tusher et al. to detect significant changes in differential gene expression between sets of samples identified by microarrays (Tusher et al., 2001). SAM comprises non-parametric statistics and compares relative difference (Larsson et al., 2005), SAM uses permutations to calculate significance and to estimate and control the FDR (Tusher et al., 2001) (SAM manual). SAM is a resampling-based procedures that corrects the test statistics by adding S_0, a small, strictly positive parameter that is also known as a fudge factor, to the denominator (Lin et al., 2008).

The SAM analysis was calculated using the 'Siggenes' package from H. Schwender (Schwender, 2012) with the R environment version 3.1.1 (R Core Team, 2014). The package was modified that missing values were not imputed. SAM analysis was calculated using a S_0 of 0.1 and significant different proteins were accepted with a FDR of 1%. Data analysis was carried out by G. Poschmann from the Proteomic facility at the BMFZ in Düsseldorf.

2.2.14 RNA interference

SiRNA mediated knockdown of NCL was performed in the parental Flp-In T-REx HEK293 cell line using the transfection reagent Lipofectamine 2000 (Invitrogen). Cells, which were transfected only with the transfection reagent were used as a negative control (mock control). Total RNA was extracted with TRIzol reagent according to the manufacturer instructions. A time-dependent knockdown experiment (24-72 h) was tested. The isolated RNA was subsequently deep sequenced (Illumina) and either enriched for poly (A) in order to analyze the knockdown of the gene and other ribosome biogenesis factors or total RNA without ribo-depletion for pre rRNA analysis.

2.2.14.1 SiRNA transfection using Lipofectamine 2000

One-day prior transfection, 10^5 Flp-In T-REx HEK293 cells per ml were seeded in a 6-well plate containing 2 ml antibiotic free DMEM media. The next day, cells were transfected with 100 pmol siRNA and Lipofectamine 2000. If the transfection reagent seems to have a toxic effect on the cells, the media can be changed 24 h to 48 h after transfection.

Table 18: SiRNA transfection in a 6-well plate using Lipofectamine 2000

Reagent	Concentration/Volume	Opti-MEM		
siRNA	100 pmol	250 µl	5 min, RT	mix, 20 min, RT
Lipofectamine 2000	5 µl	250 µl	5 min, RT	

Cells were incubated at 37°C in a CO_2 incubator for 72 h prior further downstream analysis. The sufficient knockdown of the target was tested using three different siRNAs against NCL (NCL I-III) and different time points (24 h, 48 h and 72 h). Western blot analysis with a specific primary antibody for NCL and an internal control like β-tubulin or β-actin was used and performed.

2.2.14.2 RNA extraction and next-generation sequencing

Cells were washed once with PBS, and 500 µl TRIzol was directly applied to the cells in each well. Cells were resuspended and incubated for 5 min at RT prior addition of 100 µl chloroform. Cells were vigorously vortexed for 15 sec and incubated for 3 min at RT prior centrifugation at 12,000 x g for 15 min at 4°C. The aqueous phase containing the RNA was transferred to a new tube and the TRIzol-chloroform extraction was repeated. The RNA was precipitated using 500 µl isopropyl alcohol and incubated at RT for 10 min and centrifuged at 12,000 x g for 10 min at 4°C. The RNA was washed twice with 75% ethanol, the RNA pellet was air dried and subsequently dissolved in H_2O and incubated for 10 min at 60°C. Purified RNA from knockdown experiments was submitted for total RNA-Seq and for poly (A). The cDNA library was performed with the TruSeq version 1.5 kit (Illumina). The cDNA was analyzed on an Illumina HiSeq 2000 in a 100 base-pair (bp) single-end sequencing run.

2.2.14.3 Analysis of the RNA sequencing data

The RNA samples were poly (A) sequenced with 100 bp single-end sequence reads (SR). The resulting files in FASTQ format were subjected to cutadapt (Marcel, 2011), a program that removes the remaining 5'- and 3'-adapter sequences. The data was converted to a FASTA format and aligned to the human transcriptome and human genome. Bowtie (Langmead et al., 2009) is a read alignment software with settings allowing up to two mismatches and was used to align reads to the transcriptome and genome. The transcripts were divided into categories including mRNA, rRNA, pre rRNA, snoRNA and other RNA types. Each read was assigned to the most probable category following category abundance. The RNA-Seq data was specified in total number of reads mapped to the particular transcript and measured in RPKM units. RPKM stands for reads per kilobase (kb) per million mapped reads. RPKM was

calculated as the total number of reads mapped to the transcript divided by the total number of mapped reads per transcript in millions and the length of the transcript in kb. The R environment was mainly used for visualization purpose. Data analysis of the RNA-Seq data was provided and performed by P. Morozov (Tuschl lab, Rockefeller University).

3 RESULTS

3.1 Nucleolar localization signal prediction of Nucleolin

Nucleolin is localized within the three compartments of the nucleolus (Escande et al., 1985). However, the precise mechanisms of the nucleolar localization of NCL remain unknown. NCL contains a bipartite nuclear localization signal (NLS) in the N-terminal domain, whereas a nucleolar localization signal (NoLS) was not predicted for NCL by NoD (nucleolar localization sequence detector). NoD is a web server that allows the analysis of NoL signals in proteins (Scott et al., 2010; Scott et al., 2011). The computational prediction tool was accomplished on a manually curated set of 46 experimentally validated NoLS in humans. Several hundreds of proteins that are not localized within the nucleolus served as a negative data set (Scott et al., 2010; Scott et al., 2011). NoD computes a confidence score for a specific protein sequence, a score below 0.8 predics no NoLS, whereas a score of 0.8-1 shows a specific NoLS within the tested protein sequence.

For the entire protein sequence of NCL a score below 0.8 was computed using NoD (Figure 21). Peaks showing a score of $\geq$ 0.7 were observed in the NLS (279-298) of NCL, and at position ~360, at the end of RRM1 (RRM1: 307-383).
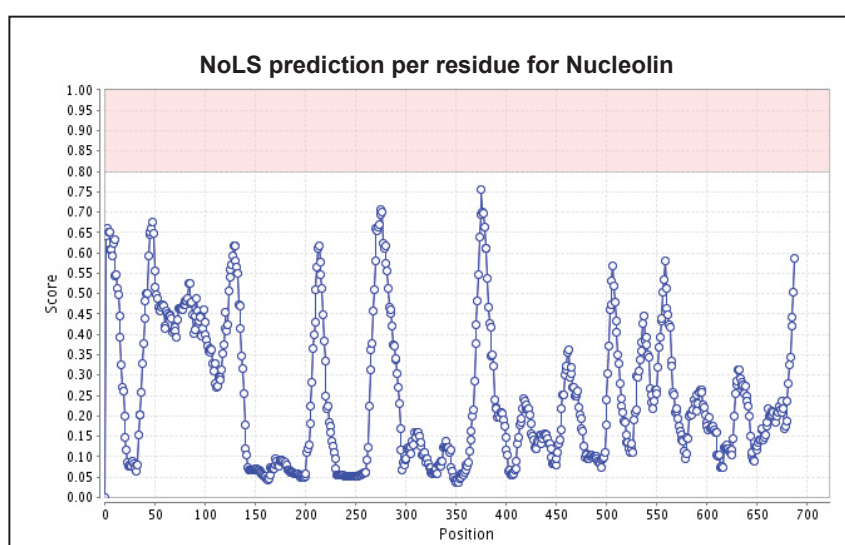
Figure 21: NoLS prediction of Nucleolin using NoD

The nucleolar localization signal prediction tool NoD computes a confidence score for each position in order to predict a nucleolar localization signal (NoLs) in Nucleolin. A score above 0.8 per residue predicts a NoLs, whereas a score below 0.8 predicts no nucleolar localization signal, which was shown for the entire sequence of Nucleolin. (Scott et al., 2011).

3.2 Localization of Nucleolin

3.2.1 Localization of full length and ∆N Nucleolin

The N terminus of NCL contains several acidic stretches of D/E residues, these acid blobs were proposed to play a role in transcription activation (Hanakahi et al., 1997). To test, if the N-terminal domain of NCL has an impact on the localization, immunohistochemistry of full length NCL versus the deletion variant of ∆N NCL, which lacks the acidic stretch and part of the bipartite NLS was performed. Full length NCL (Figure 22) and ∆N NCL (Figure 23) are both localized within the nucleolus in HEK293 cells.
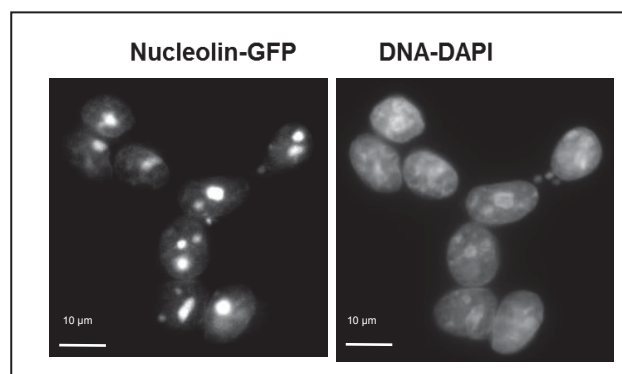
Figure 22: Localization of full length Nucleolin

HEK293 cells inducibly expressing GFP-tagged full length NCL were tested in IHC. Cells were induced with doxycycline for 30 h. The nucleus is stained with DAPI (1:1,000), while GFP tagged NCL shows nucleolar localization.

Since the ∆N NCL variant is also localized within the nucleolus (Figure 23), the first part of the bipartite NLS and the acidic stretch of the N-terminal domain are not required for nucleolar localization.
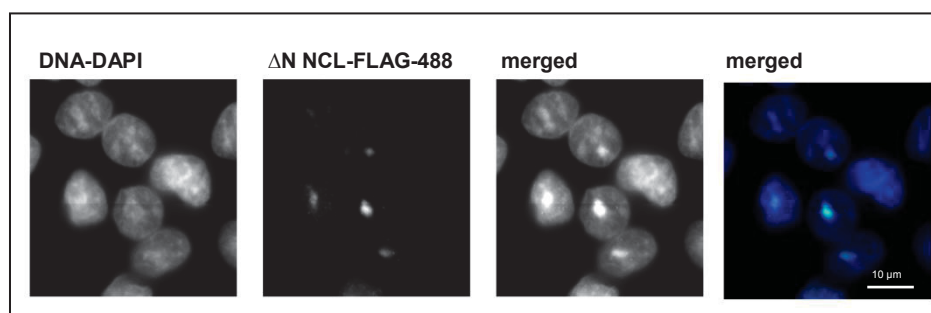
Figure 23: Localization of Nucleolin lacking the N terminus

HEK293 cells stably expressing FLAG/HA tagged ∆N NCL were fixed and stained with an anti-FLAG antibody. Nucleolin missing the N-terminal domain (∆N) is localized in the nucleolus. DAPI was used to stain the DNA.

To exclude that the GFP or the FLAG/HA affinity tags have any effect on the localization of NCL, exogenous NCL containing either a GFP or a FLAG/HA tag were tested in comparison to endogenous NCL using either an anti-FLAG antibody or a specific anti-C23 (NCL) antibody, respectively. All three constructs showed the same nucleolar localization (not shown). Therefore, the affinity tags have no impact on the nucleolar localization of NCL and can be applied for further downstream analysis.

3.2.2 Localization of Nucleolin in the presence of Actinomycin-D

It was shown that Actinomycin-D treatment influences the localization patterns of several nucleolar proteins as well as the integrity of the nucleolus (Brodska et al., 2015; Chan, 1992; Chen and Jiang, 2004). Therefore, the localization of NCL in the presence of Act-D that inhibits Pol I transcription was also tested using IHC and RNA-FISH experiments. HEK293 cells inducibly expressing GFP tagged NCL were tested with low doses (0.4 µg/ml) of Act-D for 6 h and were analyzed by fluorescence microscopy.
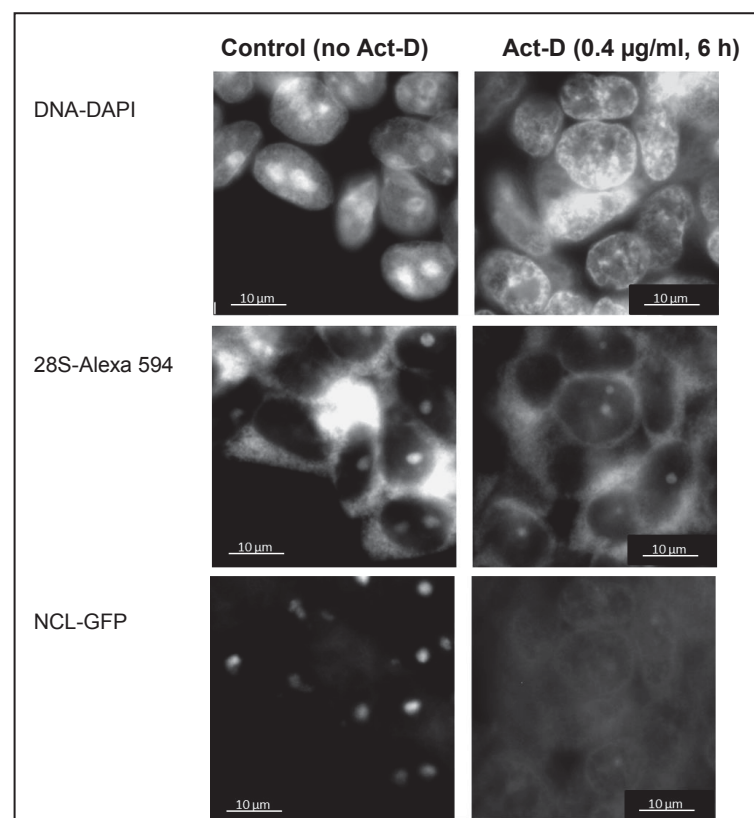
Figure 24: Localization of GFP tagged Nucleolin in the presence of Actinomycin-D

HEK293 cells inducibly expressing GFP tagged Nucleolin were incubated with 0.4 µg/ml Act-D for 6 h and tested in RNA-FISH. Cells were grown in chamber slides, induced with doxycycline for 30 h prior fixation with 4% PFA for 1 h 45 min at 4°C. Cells were hybridized overnight with a 28S rRNA probe, the DNA was stained with DAPI (1:1,000). GFP tagged Nucleolin relocalizes to the nucleoplasm in the presence of Act-D treatment.

HEK293 cells were hybridized with FISH probes targeting the 28S rRNA to visualize the nucleolus, DAPI was used to stain the DNA within the nucleus. In the presence of Act-D for 6 h NCL diffuses and relocalizes to the nucleoplasm (Figure 24). To exclude any side effects of the GFP affinity tag upon Act-D treatment, HEK293 cells expressing FLAG/HA tagged NCL were also tested. The previous result was confirmed (Figure 25), NCL diffuses and relocalizes to the nucleoplasm upon Act-D treatment.

Control (no Act-D)   Act-D (0.4 µg/ml, 6 h)

10 µm

Nucleolin-488 and DNA-DAPI, merged
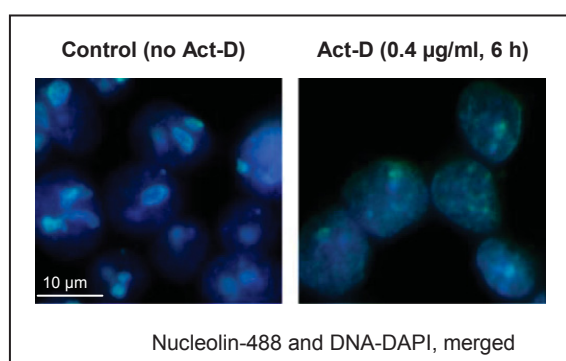
Figure 25: Localization of FLAG/HA tagged Nucleolin in the presence of Actinomycin-D

HEK293 cells expressing FLAG/HA tagged NCL were treated with 0.4 µg/ml Act-D for 6 h and tested in IHC. Cells were fixed in 4% PFA for 1 h 45 min at 4°C, incubated with an anti-FLAG antibody and stained with a corresponding secondary antibody GAM-488 (1:1,000) and DAPI (1:1,000) in TBS-T containing goat-serum for 2 h at RT. Under these conditions, Nucleolin relocalizes to the nucleoplasm in the presence of Act-D.

3.3 4-thiouridine treatment in HEK293 cells

To identify the RNA targets of NCL, PAR-CLIP experiments were performed. In a PAR-CLIP experiments, 4-thiouridine (4SU), a photoactivatable nucleoside is incorporate into newly transcribed RNA and results in a T to C conversion upon UV crosslinking in the corresponding cDNA after reverse transcription (Figure 15). To exclude any side-effects and to test if 4SU has an impact on the localization of NCL or causes a nucleolar stress response, HEK293 cells were exposed to various 4SU concentrations and were tested by fluorescence microscopy and in western blot analyses.

It was shown that GFP tagged NCL remains localized in the nucleolus upon treatment of 4SU with concentrations of 50 µM up to 400 µM for 16 h (Figure 26) and shows no difference to the untreated control cells.

| | 4SU, 16 h | | |
| Control (untreated) | 50 µM | 100 µM | 400 µM |

Nucleolin-GFP   DNA-DAPI   Nucleolin-GFP   DNA-DAPI   Nucleolin-GFP   DNA-DAPI   Nucleolin-GFP   DNA-DAPI

20 µm
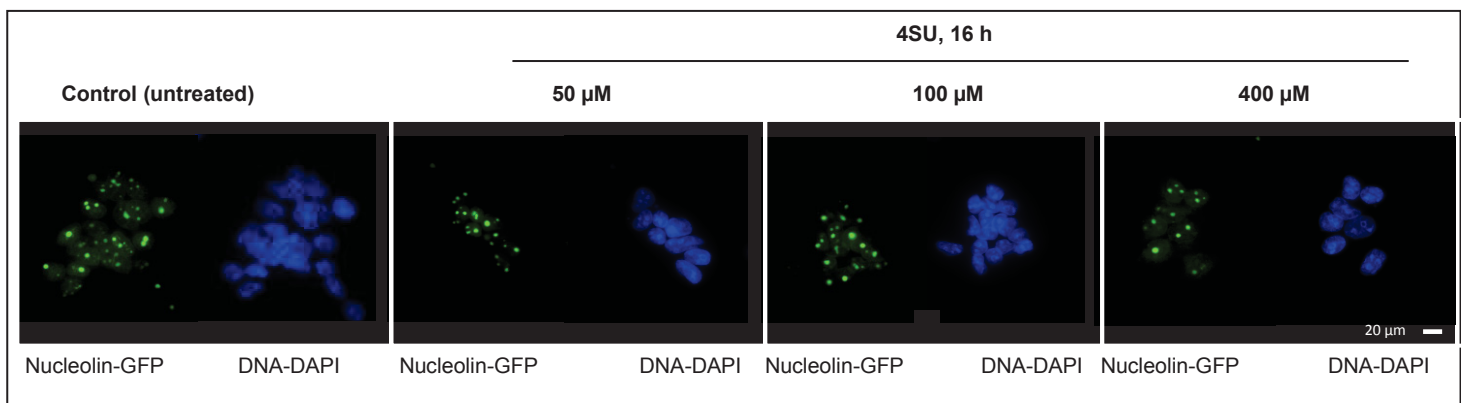
Figure 26: Localization of Nucleolin in the presence of 4SU

HEK293 cells expressing GFP tagged NCL were incubated with 50 µM, 100 µM and 400 µM 4SU for 16 h and tested in IHC. 4SU is dissolved in DMSO (0.26 mg/µl). Cells were grown in chamber slides, induced with doxycycline for 30 h prior fixation with 4% PFA and the nucleus is stained with DAPI (1:1,000). Nucleolin remains localized in the nucleolus even in the presence of 400 µM 4SU.

Additionally, to test if 4SU induces a nucleolar stress response regarding inhibition of rRNA processing, the expression of the tumor suppressor p53 was determined by western blot analysis. High levels of p53 correlate with the inhibition of rRNA processing and indicate nucleolar disruption. Therefore, p53 levels were examined in HEK293 cells, which were treated with 50 µM, 100 µM, 200 µM and 400 µM 4SU for 16 h. According to the expression of NCL, levels of p53 did not increase upon 4SU treatment (Figure 27). To summarize, 4SU does not cause a nucleolar stress response in HEK293 cells and can be used for further downstream analysis in PAR-CLIP experiments.
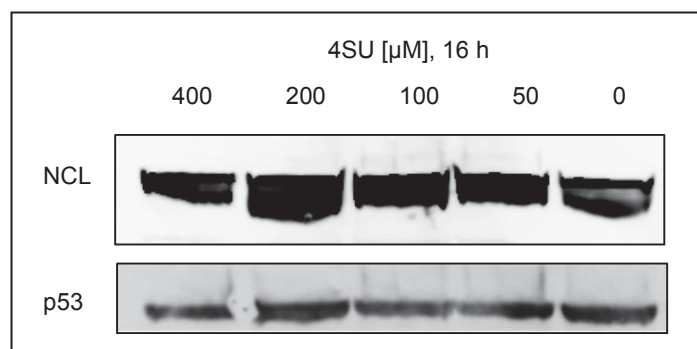
Figure 27: Stable expression of p53 upon 4SU treatment

HEK293 cells stably expressing FLAG/HA tagged NCL were incubated with various concentration of 4SU, ranging from 50 µM up to 400 µM for 16 h at 37°C. Cells were harvest, protein concentrations were determined using BCA, equal protein concentrations were loaded to the SDS gel and subsequently transferred to a nitrocellulose membrane. Protein expression of p53 (anti-p53, 1:1,000) and NCL-FLAG/HA (anti-FLAG, 1:1,000) were detected with a specific primary antibody and a corresponding HRP coupled secondary antibody. P53 levels remained stable in the presence of 4SU compared to NCL. Therefore, 4SU does not cause a nucleolar stress response.

3.4 Phosphorylation of Nucleolin by Casein Kinase II

In order to characterize and identify RNA targets of NCL, initial PAR-CLIP experiments were performed. Usually, a crosslinking band is only expected in cells that were treated with 4SU and subsequently crosslinked with 365 nm UV. A series of control experiments in the presence and absence of 4SU and UV-crosslinking were tested in NCL-FLAG immunoprecipitated HEK293 cells (Figure 28). In all conditions, protein-bound RNAs were isolated and radiolabeled with γ ^{32}P-ATP using T4 PNK. Phoshorimaging of NCL showed a radiolabeled band in every sample (Figure 28), in the absence of 4SU and UV, in the presence of 4SU and UV, in the presence of 4SU and absence of UV and vice versa.
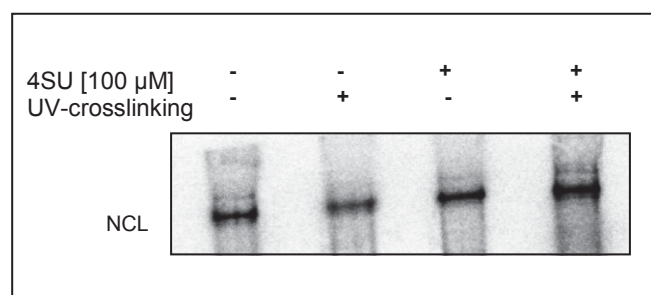
Figure 28: Phosphorimage of immunoprecipitated Nucleolin in the presence and absence of 4SU and 365 nm UV-crosslinking

HEK293 cells stably expressing FLAG/HA NCL were either incubated with or without 100 µM 4SU for 16 h and either with or without UV-crosslinking at 365 nm. The RNA is labeled using γ ^{32}P-ATP and T4 PNK. Nucleolin shows a crosslinking signal in all of the tested control samples, even in the absence of 4SU and UV-crosslinking.

To further analyze the radioactive signal, the phosphorylation reaction was performed in the absence and presence of T4 PNK (Figure 29). T4 PNK catalyzes the transfer of the γ-ATP to the 5' hydroxyl terminus of the RNA. NCL shows a phosphorylation signal in the sample without 4SU and UV crosslinking at 365 nm and even in the absence of T4 PNK. The empty vector, used as a negative control, does not show any crosslinking signal (Figure 29). These results demonstrate that NCL either comprises a kinase activity, interacts with a protein kinase that facilitates the phosphorylation reaction or NCL simply interacts directly with ATP.
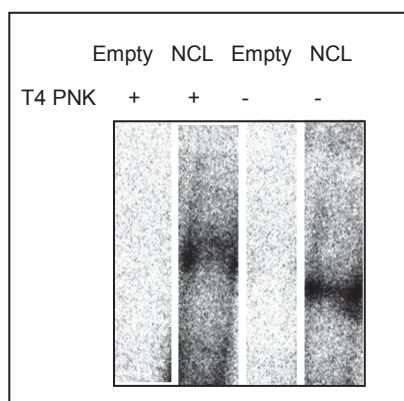
Figure 29: Phosphorimage of immunoprecipitated Nucleolin in the presence and absence of T4 PNK

Nucleolin shows a phosphorylation signal even in the absence of T4 PNK. The empty vector control, lacking FLAG/HA tagged NCL is used as a negative control and does not show any radioactive phosphorylation band.

Since γ ^{32}P-ATP is used for the labeling reaction of the crosslinked RNA within the PAR-CLIP procedure, in a next step, it is tested if NCL interacts directly with ATP. Therefore, the labeling reaction was carried out using either α ^{32}P-ATP, γ ^{32}P-ATP or γ ^{32}P-GTP. If NCL binds directly to ATP, a radioactive band is expected in the reaction with α as well as with γ ^{32}P-ATP, whereas no band should occur in the reaction with γ ^{32}P-GTP. However, no signal was observed in the labeling reaction performed with α ^{32}P-ATP, whereas a radioactive band in the labeling reaction using γ ^{32}P-ATP and γ ^{32}P-GTP was examined. This implicates that NCL does not interact with ATP, but could rather bind to a protein kinase, which is able to utilize ATP and GTP for the phosphorylation reaction (Figure 30).
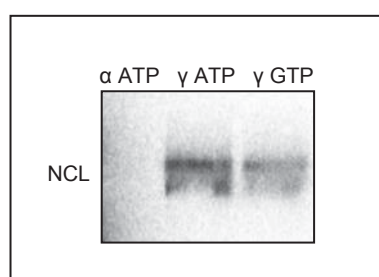
Figure 30: Phosphorimage of immunoprecipitated Nucleolin using ^{32}P α-ATP, γ-ATP or γ-GTP

A FLAG-IP using HEK293 cells stably expressing FLAG/HA NCL is performed. Subsequently, the labeling reaction was carried out using either ^{32}P α-ATP, γ-ATP or γ-GTP. The complex is loaded to a SDS gel, transferred to a membrane and exposed to a phosphorimage screen. No signal was observed in the labeling reaction using α ^{32}P-ATP, whereas a radioactive band was shown in the labeling reaction with γ ^{32}P-ATP and γ ^{32}P-GTP. Consequently, NCL does not interact directly with ATP.

CKII is one of the only protein kinases that utilizes ATP as well as GTP for its phosphotransferase reaction. To test if CKII could be responsible for the phosphorylation reaction, co-purification analysis of NCL was performed and CKII was determined by western blot analysis within the lysate, supernatant and FLAG-IP samples in HEK293 expressing FLAG/HA tagged NCL (Figure 31). The western blot indicates that CKII is co-immunoprecipitated by NCL. For western blot analysis, a specific anti-CKII α antibody was used. Consequently, it can be assumed that NCL recognizes the α-subunit of CKII, which represents the catalytic subunit of the enzyme.
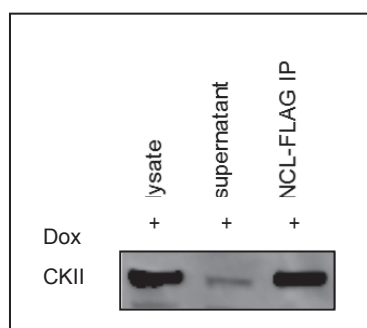
Figure 31: Co-immunoprecipitation of Casein Kinase II by Nucleolin

Western blot analysis of CKII of a lysate, supernatant and a NCL-FLAG IP sample in doxycycline induced HEK293 cell, expressing FLAG/HA tagged NCL is performed. CKII is co-immunoprecipitated by Nucleolin, implicating interaction of NCL and CKII.

To further confirm if CKII is responsible for the phosphorylation reaction of NCL, tetrabromobenzotriazole (TBB) a CKII inhibitor was tested in the labeling reaction using γ ^{32}P-ATP. Western blot analysis of NCL and CKII and phosphorimages were analyzed. The phosphorylation signal was completely abolished in the presence of TBB, in contrast to the control sample (without inhibitor). Western blot analysis detected a band for CKII and FLAG in the presence and absence of TBB (Figure 32).
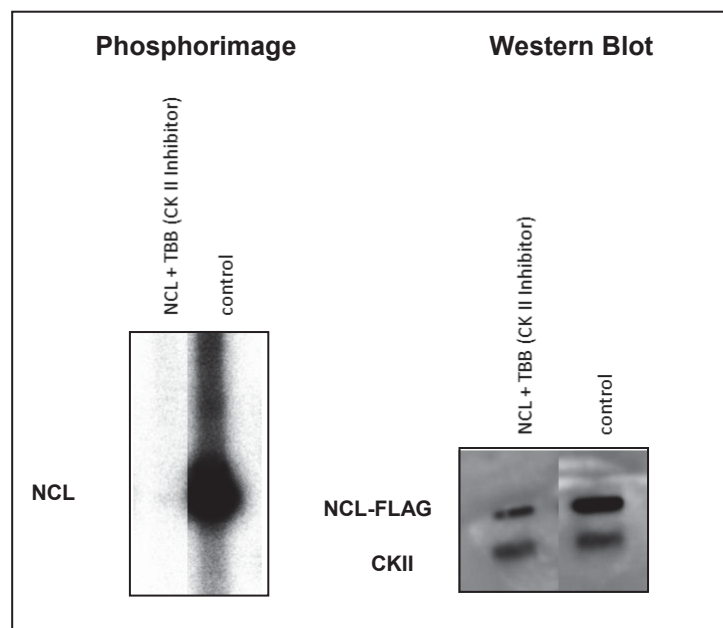
Figure 32: Inhibition of the Casein Kinase II phosphorylation reaction with TBB

HEK293 cells stably expressing FLAG/HA tagged NCL were immunoprecipitated with an anti-FLAG antibody. Cells were washed with IP-wash buffer, high-salt buffer and CIP treated. The labeling reaction was performed with T4 PNK in the presence and absence of TBB, a CKII inhibitor. A phosphorylation signal is observed in the absence (control) of TBB, whereas the signal is completely abolished in the presence of TBB. Western blot analysis detected FLAG-NCL and CKII in both samples, in the absence and presence of TBB.

In summary, NCL does not interact with ATP (Figure 30), NCL is recognized by the α-subunit of CKII and is phosphorylated. CKII is one of the only kinases, which can utilize ATP as well as GTP for the phosphorylation reaction that is completely inhibited in the presence of TBB.

3.5 Interaction of Nucleolin with RNA

3.5.1 UV-Crosslinking of Nucleolin

To identify the RNA targets of NCL, PAR-CLIP experiments were performed. Therefore, HEK293 cells, stably expressing FLAG/HA tagged NCL were incubated with 4SU and crosslinked at 365 nm UV. After a FLAG-IP, the immunoprecipitated and NCL-crosslinked RNA was dephosphorylated with CIP and labeled with T4 PNK and γ ^{32}P-ATP. The radiolabeled RNA-protein complexes were resolved via SDS-PAGE (Figure 33). Phosphorimaging allowed the detection of a radioactive signal at ~110

kDa, the size range of NCL. Western blot analysis using an anti-FLAG antibody confirmed the successful IP of NCL. As a negative control, HEK293 cells stably expressing only the FLAG/HA tag were used. Subsequently, these cells are referred to as 'empty' or 'empty vector control'. These cells were treated in the same way as described for the NCL-FLAG/HA expressing cells, however no radioactive signal at the NCL size range could be detected (Figure 33).
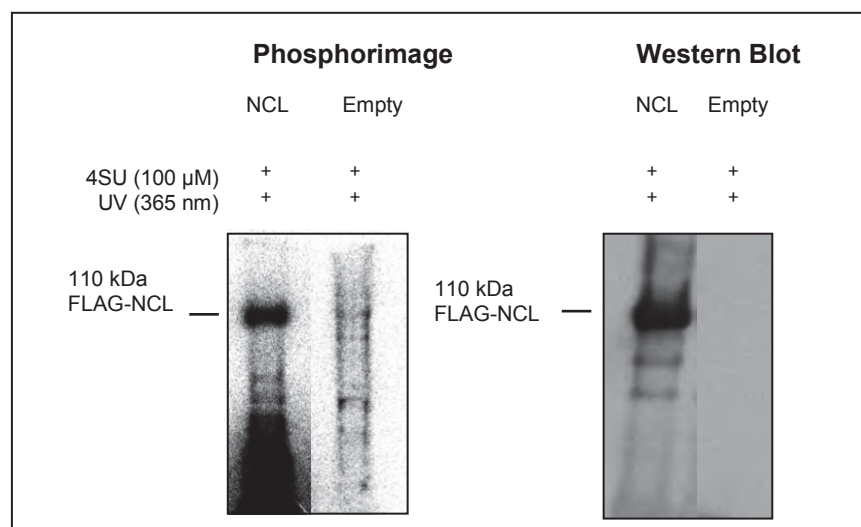
Figure 33: PAR-CLIP experiment of HEK293 cells stably expressing FLAG/HA tagged Nucleolin

HEK293 cells stably expressing FLAG/HA tagged NCL were incubated for 14 h with 100 µM 4SU and crosslinked with UV at 365 nm. The phosphorimage shows the SDS gel of the resolved RNA-protein complexes. A crosslinking band of NCL to RNA was detectable at the MW of NCL. The corresponding western blot detects a FLAG signal in the NCL sample at 110 kDa. The empty vector expresses a FLAG/HA tag but lacks the GOI, and serves as a negative control. The control shows no crosslinking signal.

In a next step, it was tested if the crosslinking signal of NCL obtained in PAR-CLIP experiments results partially from RNA and does not only result from phosphorylated NCL residues by CKII. Therefore, in addition to the labeling of the 5' end of crosslinked RNA that is usually performed in PAR-CLIP experiments and conventionally labeled using γ ^{32}P-ATP and T4 PNK. The 3' end of the crosslinked RNA was labeled using T4 RNA ligase and ^{32}P-pCp, a substrate that cannot be utilized by CKII. Consequently, a crosslinking band due to the 3' end labeling using pCp results from crosslinked RNA rather than from phosphorylated protein residues by CKII.

Since the labeling reaction of 3' ends of the RNA is not as efficient as the 5' end labeling, it has to be considered that the intensities of the two signals are not comparable in terms of RNA quantity. HEK293 cells induced with dox showed a radioactive phosphorylation signal at the molecular weight of NCL in the 3' and 5' labeling reaction. The phosphorylation signal of the 3' end indicates only crosslinked and co- immunoprecipitated RNA by NCL, whereas the 5' ATP labeling reaction shows crosslinked and co- immunoprecipitated RNA as well as phosphorylated protein residues by CKII. Cells that were not induced with dox, did consequently not express FLAG/HA tagged NCL and served as a negative control. As expected, these control cells did not show any signal (Figure 34).
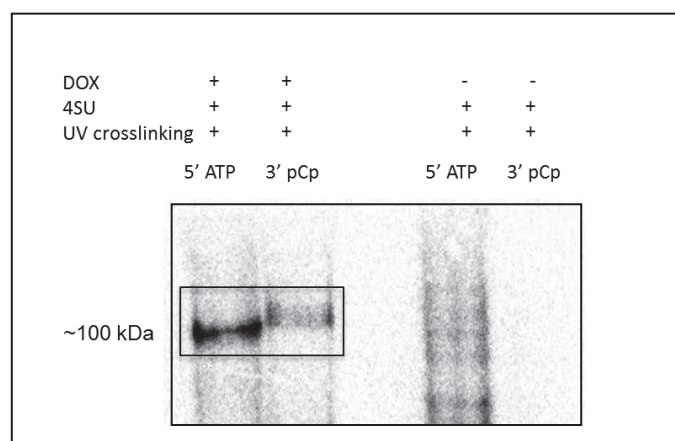
Figure 34: Initial PAR-CLIP experiments of Nucleolin using 3' or 5' end labeling of the crosslinked RNA

Prior FLAG-IP, HEK293 cells were incubated with 4SU and UV crosslinked at 365 nm. The crosslinked RNA is subsequently labeled either at the 5' or the 3' end of the RNA, using γ ^{32}P-ATP or ^{32}P-pCp, respectively.

In summary, NCL interacts with RNA, either labeled at the 3' or 5' end, showing a radioactive band at the molecular weight of NCL. Subsequently, PAR-CLIP analysis of NCL was performed in a large scale to isolate and identify the NCL bound RNAs.

The isolated RNA was treated with RNAse T1 and ligated with 3'- and 5'-adapters for reverse transcription and subsequent sequencing. Finally, successful cDNA PAR-CLIP libraries of NCL were submitted for deep sequencing. PAR-CLIP allows a precise

identification of the position of the crosslinking by mutation, which makes it possible to separate these sequences from background that are derived from abundant cellular RNAs (Hafner et al., 2010a; Hafner et al., 2010b). cDNA libraries prepared by PAR-CLIP contain a characteristic thymidine to cytidine transition using 4SU (Figure 15).

3.5.2 PAR-CLIP analysis of Nucleolin

Within this thesis, two cDNA libraries of NCL (full length and delta N) and one empty vector control library were submitted for sequencing in order to further characterize and analyze the cellular RNA targets of NCL. A second cDNA PAR-CLIP library of NCL was previously generated by Jessica Höll. The HuR PAR-CLIP library was used as a control from already published datasets (Mukherjee et al., 2011). HuR, also known as ELAVL1, is a nuclear RBP that contains RRMs like NCL. Since NCL and HuR share a common snoRNA target, MBII-52 (Soeno et al., 2010), HuR was used as a positive control for the interaction with snoRNAs. HuR also binds to mRNA and stabilizes mRNA targets (Abdelmohsen et al., 2011; Lebedeva et al., 2011), therefore, this PAR-CLIP library set was additionally used as a control for mRNAs interaction.

During 365 nm UV crosslinking of the photoreactive 4SU with aromatic amino acids of the protein, the chemical structure of the uridine analogue changes conformation. In the reverse transcription step the cDNA library of the crosslinked and isolated RNA fragment is prepared, the RT enzyme incorporates a G instead of an A at the corresponding position, which results in a T to C conversion (Figure 15). This event makes PAR-CLIP a very powerful technique, since T to C transitions enable the identification of specific crosslink events and distinguishes specific RNA-protein interactions from background RNA. Therefore, in a successful PAR-CLIP experiment, the largest fraction having one mismatch aligned against the genome and transcriptome should contain T to C conversions (d1 T $\rightarrow$ C) (Hafner et al., 2010a).

During the mapping procedure to the reference, zero, one or two mismatches were allowed. Subsequently, d0 implicates zero mismatch, d1 and d2 refer to one and two mismatches, respectively. Data analysis groups the libraries in different RNA categories, which were mapped in a hierarchical order according to their relative abundance, from the most abundant to the less abundant as they occur within the cell. The cDNA PAR-CLIP library analysis is performed with a minimum read length of 20 nt up to 41 nt and was compared to HuR, a publically available data set (Mukherjee et al., 2011) and an empty vector control.

The analysis of the cDNA PAR-CLIP library of NCL is shown in a bar diagram (Figure 35) that illustrates the accumulation of T to C crosslinks in snoRNA and pre rRNA targets bound by NCL. The bar diagram plots the redundant read counts on the y-axis and the read length from 20 nt up to 41 nt on the x-axis. A maximum of two mismatches (d2) depicted in black were allowed, 'd1 T → C' implicates the described T to C conversion that are depicted in red, 'd1 other' implicates one mismatch, except T to C (grey) and no mismatch is defined as d0 (white). Eye-catching are the red bars in the diagram that are enriched in snoRNAs and pre rRNA. These represent the T to C transitions and crosslinking targets sites for NCL. NCL showed good crosslinking capacity as previously observed by phosphorimaging. The resulting PAR-CLIP library showed high T to C conversions in snoRNA and pre rRNA, whereas the crosslinking efficiency for mRNA and rRNA targets is very low (Figure 35 and Table 19). However, most of the reads in total were observed in rRNA, but almost no T to C conversions were identified. The major reads within this RNA category showed zero mismatches (d0), while in mRNA most of the reads contained two mismatches (d2).
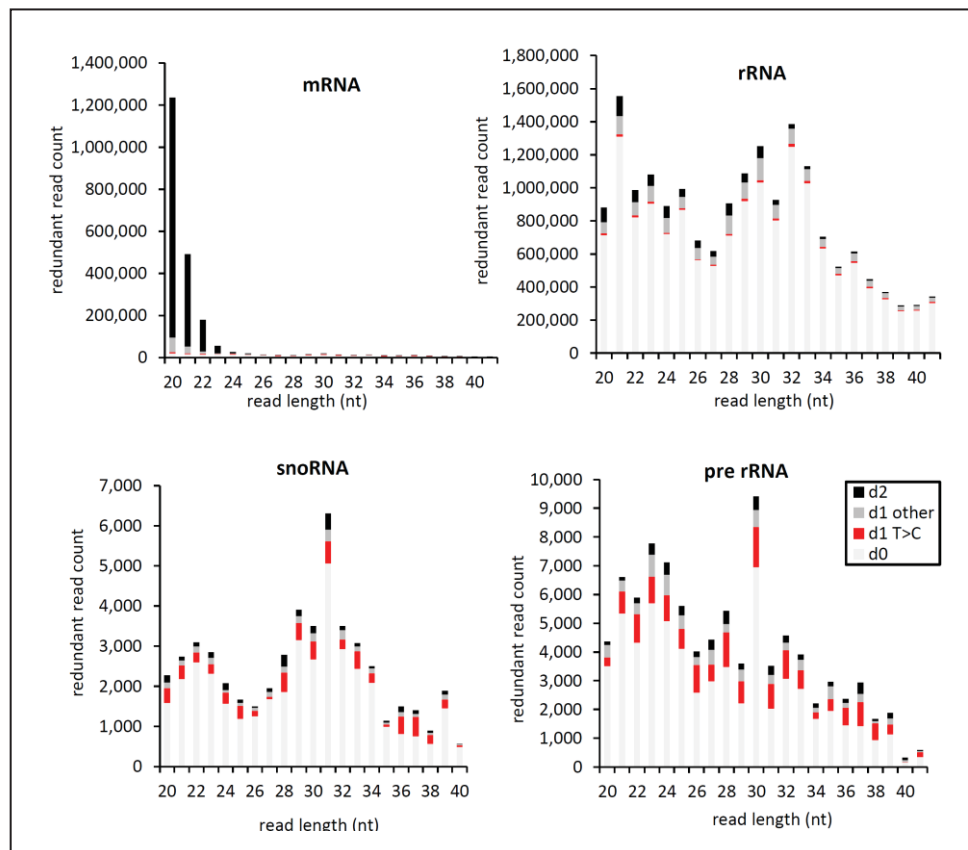
Figure 35: Read distribution of the cDNA PAR-CLIP library of full length Nucleolin

PAR-CLIP cDNA library analysis of Nucleolin shows redundant read count versus read length of 20 nt up to 41 nt for mRNA, rRNA, snoRNA and pre rRNA. Two mismatches (d2) depicted in black, d1 T>C implicates T to C conversion, which are illustrated in red, d1 other implicates one mismatch except T to C (grey) and no mismatch is defined as d0 (white). SnoRNAs and pre rRNA were identified as targets for Nucleolin, whereas mRNA and rRNA show almost no T to C conversions, which are indicated by the red bars.

In summary, for a successful PAR-CLIP experiment, the largest fraction of distance 1 mismatches is expected to be T to C transitions. NCL crosslinks to snoRNAs and pre rRNA and interacts with two RNA targets that are directly associated with ribosome biogenesis, while NCL is absent from mRNA and mature rRNA (Figure 35).

Figure 36 shows an overview of the entire human ribosome repeating unit including the IGS and the ~13 kb long primary 47S rRNA transcript (Figure 3). In order to get a more precise analysis of the identified target sites, Figure 36 illustrates log2 read count and the crosslinking sites of NCL along the entire 47S rRNA transcript. Striking, the read coverage starts immediately after the first cleavage site (01), which is marked in green and is localized in the 5'ETS. The graph below shows the crosslinking sites of NCL in

the precursor region implicated by T to C conversions that are depicted in red and one

mismatch in total (d1) depicted in black. T to C conversions, implicating crosslinking of

NCL to rRNA, are mainly found in the precursor region in the first half of the 5'ETS after

the first cleavage site and in the two ITS (ITS1 and ITS2). Therefore, it can be

supposed that NCL functions downstream of the first cleavage site, in the 45S rRNA.

The read coverage in the mature rRNA (grey boxes) is quite high, although, these

regions contain almost no T to C conversions (red lines).

In summary, NCL mainly binds to the precursor region rather than to the mature region

in the human ribosome repeating unit, as illustrated in Figure 36.
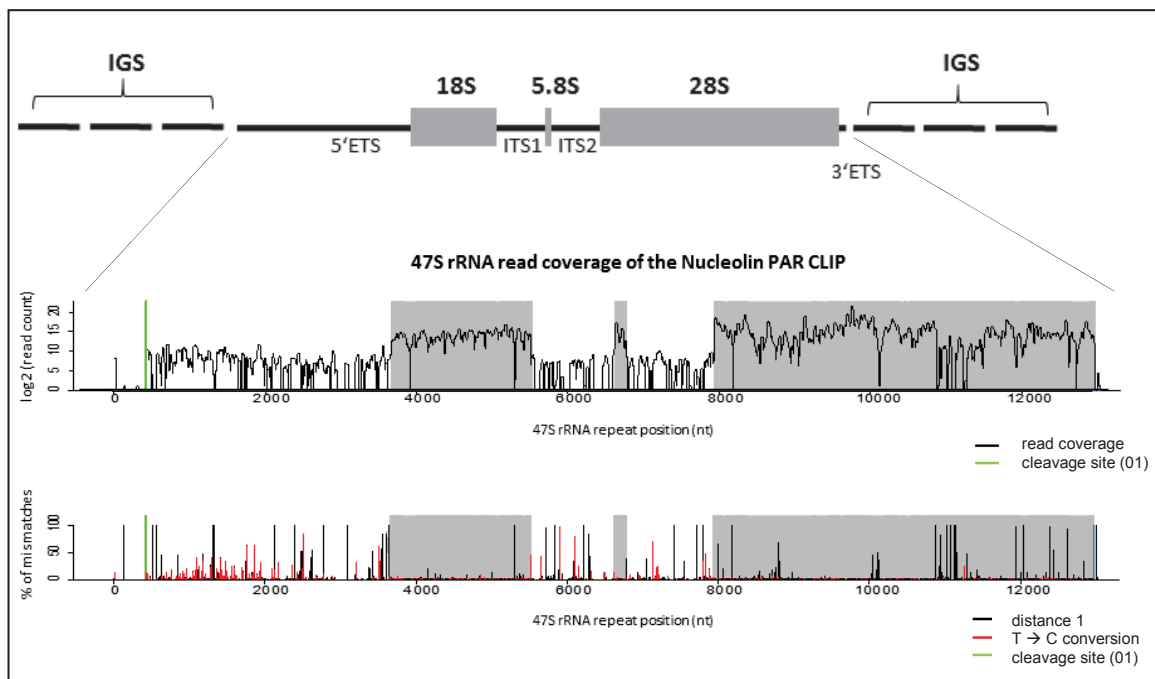
Figure 36: Crosslinking sites of Nucleolin along the 47S rRNA identified by PAR-CLIP

The cluster of the human ribosome repeating unit, composed of three mature ribosomal RNAs; 18S, 5.8S and 28S, the 5'ETS (external transcribed spacer), ITS1 and ITS2 (internal transcribed spacer), 3'ETS followed by a ~30 kb long intergenic spacer (IGS), which separates each cluster from the next 5'ETS. The 47S rRNA read coverage of the Nucleolin PAR-CLIP experiment II is illustrated. NCL shows enriched read coverage immediately downstream of the first cleavage site (01) depicted in green, in the 5'ETS. The graph below distinguishes distance 1 in total (black) from T to C conversions (red). Grey boxes represent the mature rRNAs (18S, 5.8S, 28S).

Table 19 summarizes the PAR-CLIP library analyses of NCL I, NCL II, HuR and the

empty vector control, and compares total number of reads and total annotated reads in

%. The empty vector functions as a negative control and represents background

binding to the magnetic beads that are linked to the anti-FLAG antibody. Different RNA categories within each library including mRNA, mRNA gene, rRNA, pre rRNA, snoRNA, pre snoRNA and others are listed according to annotated reads in %. The RNA classes were mapped in a hierarchical fashion, due to their relative abundance within a cell, from the most abundant to the less abundant RNAs. Additionally, the table shows zero (d=0), one (d=1) and two mismatches (d2) in %, the ratio of T to C conversions to zero mismatch (d1(T$\rightarrow$C)/d0) and to one mismatches (d1(T$\rightarrow$C)/d1). If d1(T$\rightarrow$C)/d1 is close to 1, almost 100% of the occurring mismatches contain T to C conversions. The NCL PAR-CLIP experiments I and II show a high ratio of 0.7 in pre rRNA, and 0.7 and 0.9 in snoRNAs. HuR, which is known to bind snoRNAs and interacts with the brain specific snoRNA MBII-52 (Soeno et al., 2010) shows a ratio of 0.7 in snoRNA, whereas it is absent from pre rRNA and consequently a low ratio of 0.2 is observed. HuR is a mRNA-binding protein and shows a high ratio of 0.9, almost all reads with one mismatch contain T to C conversions in mRNAs. In contrast, NCL shows only a ratio of 0.5 in mRNAs. The empty vector control indicates a cut off ratio of 0.3 in pre rRNA and 0.1 for snoRNAs, indicating almost no crosslinking (Table 19).

In summary, the PAR-CLIP experiment revealed that NCL mainly crosslinks to pre rRNA and snoRNAs. During data analyses we encountered difficulties, since rRNAs exist in multiple copies within the genome (Konig et al., 2011). This directly leads to another challenge. Current genome annotation databases, like Refseq or Ensembl, show poorly curated and incomplete rRNA annotations. Also, in the presence of the highly abundant, mature rRNAs it can be very difficult to detect expression changes within the large rRNA precursor transcript. Although, hierarchical alignment against the most abundant RNAs showed specific upregulation of pre rRNA and snoRNAs within the NCL PAR-CLIP analysis (Table 19).

Table 19: PAR-CLIP analysis of Nucleolin versus an empty vector control and HuR

Protein	total reads	total annotated reads (%)	category	annotated reads (%)	d0 (%)	d1 (%)	d2 (%)	d1 (T→C)/ d0	d1 (T→C)/ d1
NCL, Experiment I	5,463,040	40.20	mRNA	3.7900	0.5300	0.5300	2.7300	0.5	0.5
			mRNA gene	14.9500	0.4800	1.7500	12.7200	0.5	0.1
			rRNA	44.8400	31.4700	7.8000	5.5700	0.1	0.2
			pre rRNA	0.0910	0.0470	0.0320	0.0120	0.5	**0.7**
			snoRNA	0.4470	0.2170	0.1970	0.0330	0.8	**0.9**
			pre snoRNA	0.0079	0.0011	0.0045	0.0023	2.9	0.7
			genome and others	35.8741	19.0300	6.8900	9.9541	0.0	0.1
NCL, Experiment II	188,650,681	20.00	mRNA	5.8000	0.5600	0.5700	4.6700	0.3	0.3
			mRNA gene	26.5000	0.3200	2.8400	23.3400	0.9	0.1
			rRNA	47.5200	40.6100	4.5000	2.4100	0.0	0.1
			pre rRNA	0.2400	0.1700	0.0600	0.0100	0.2	**0.7**
			snoRNA	0.1400	0.1100	0.0200	0.0100	0.2	**0.7**
			pre snoRNA	0.0035	0.0018	0.0014	0.0003	0.7	0.9
			genome and others	19.7965	2.0900	3.4000	14.3065	0.2	0.1
HuR	17,231,865	53.40	mRNA	23.2100	3.4700	11.7000	8.0400	3.0	0.9
			mRNA gene	46.0200	6.0300	21.5200	18.4700	2.4	0.7
			rRNA	5.3000	2.1400	1.6000	1.5600	0.1	0.1
			pre rRNA	0.0011	0.0005	0.0003	0.0003	0.1	**0.2**
			snoRNA	0.0092	0.0022	0.0038	0.0032	1.3	**0.7**
			pre snoRNA	0.0009	0.0001	0.0006	0.0002	6.4	0.9
			genome and others	25.4588	6.2900	10.5500	8.6188	1.0	0.6
Empty vector	70,067,824	39.00	mRNA	7.7400	0.4300	0.6500	6.6600	0.1	0.1
			mRNA gene	35.5200	1.2800	3.5000	30.7400	0.2	0.1
			rRNA	15.3600	11.2700	1.9100	2.1800	0.0	0.2
			pre rRNA	0.0227	0.0197	0.0018	0.0012	0.0	**0.3**
			snoRNA	0.0640	0.0530	0.0070	0.0040	0.0	**0.1**
			pre snoRNA	0.0007	0.0000	0.0000	0.0007	/	/
			genome and others	41.2926	12.1800	7.2500	21.8626	0.1	0.1

PAR-CLIP library analysis of 20 nt to 41 nt read length for NCL experiment I and II, HuR and an empty vector control. The table summarizes the amount of total reads and total annotated reads in %. RNA categories including mRNA, mRNA gene, rRNA, pre rRNA, snoRNA, pre snoRNA, and genome and others are listed according to annotated reads in % and to the number of mismatches, including zero mismatches (d0 in %), d1 in % shows one mismatch, whereas d2 in % represents two mismatches within all reads. The corresponding ratio of T to C conversions to one mismatch in total (d1 (T → C)/d1) refers to the crosslinking sites in each RNA category. If the ratio is close to 1 almost all of the occurring mismatches (d1) contain T to C conversions, implicating crosslinking and binding sites of the RBP to the specific RNA target. Indicated in bold are precursor rRNA and snoRNAs, which show a high ratio in the NCL PAR-CLIP experiments I and II. HuR, which is known to bind snoRNAs and interacts with the brain specific snoRNA MBII-52 reveals a ratio of 0.7, indicating snoRNAs to be a target of HuR. In general, HuR is a mRNA-binding protein and consequently shows a high ratio of 0.9 of T to C conversions to total d1, whereas NCL shows a ratio of 0.5 in mRNAs. The empty vector control shows a ratio of 0.3 in pre rRNA and only 0.1 in snoRNAs. Jessica Höll performed the PAR-CLIP library for the NCL experiment I. The HuR PAR-CLIP library is a publically available dataset from Mukherjee et al. (Mukherjee et al., 2011).

3.5.3 PAR-CLIP analysis of ΔN Nucleolin

The N terminus of NCL is rich in aspartic and glutamic acids (D/E) (Figure 10). Domains with acidic stretches were characterized as negative noodles or acid blobs that are common in transcriptional activating proteins (Sigler, 1988). As a consequence, NCL might be involved in transcription activation and regulation of Pol I (Hanakahi et al., 1997). Previous studies by Ginisty et al. proposed that the N terminus is required for the interaction with the U3 snoRNP. Also, northern blots revealed that U3 interacts specifically with NCL, whereas no binding was observed in NCL lacking the N terminus (ΔN) (Ginisty et al., 1998). According to these studies, a PAR-CLIP experiment of NCL that is missing the N terminal domain was performed to test if this variant shows comparable T to C conversion like full length NCL and interacts with the same cellular RNA targets.

The analysis of the cDNA PAR-CLIP library of ΔN NCL is shown in a bar diagram (Figure 37) that plots the redundant read counts on the y-axis and the read length from 20 nt up to 41 nt on the x-axis. In general, the ΔN NCL PAR-CLIP library illustrates very low levels of T to C conversions in mRNA, rRNA, snoRNAs and pre rRNA and consequently a very low crosslinking efficiency was observed (Figure 37 and Table 20). However, most of the reads in total were found in rRNA, but almost no T to C conversions were identified. The major reads within this RNA category showed zero mismatches (d0), while in mRNA most of the reads contained two mismatches (d2) depicted in black. According to the number of reads, mRNA showed the highest ratio of d1(T→C)/d1 of 0.5, a ratio of 0.2 was observed in rRNA, and 0.4 for snoRNA and pre rRNA (Table 20). A comparison of the ΔN with the full length NCL PAR-CLIP library is summarized in Table 20.
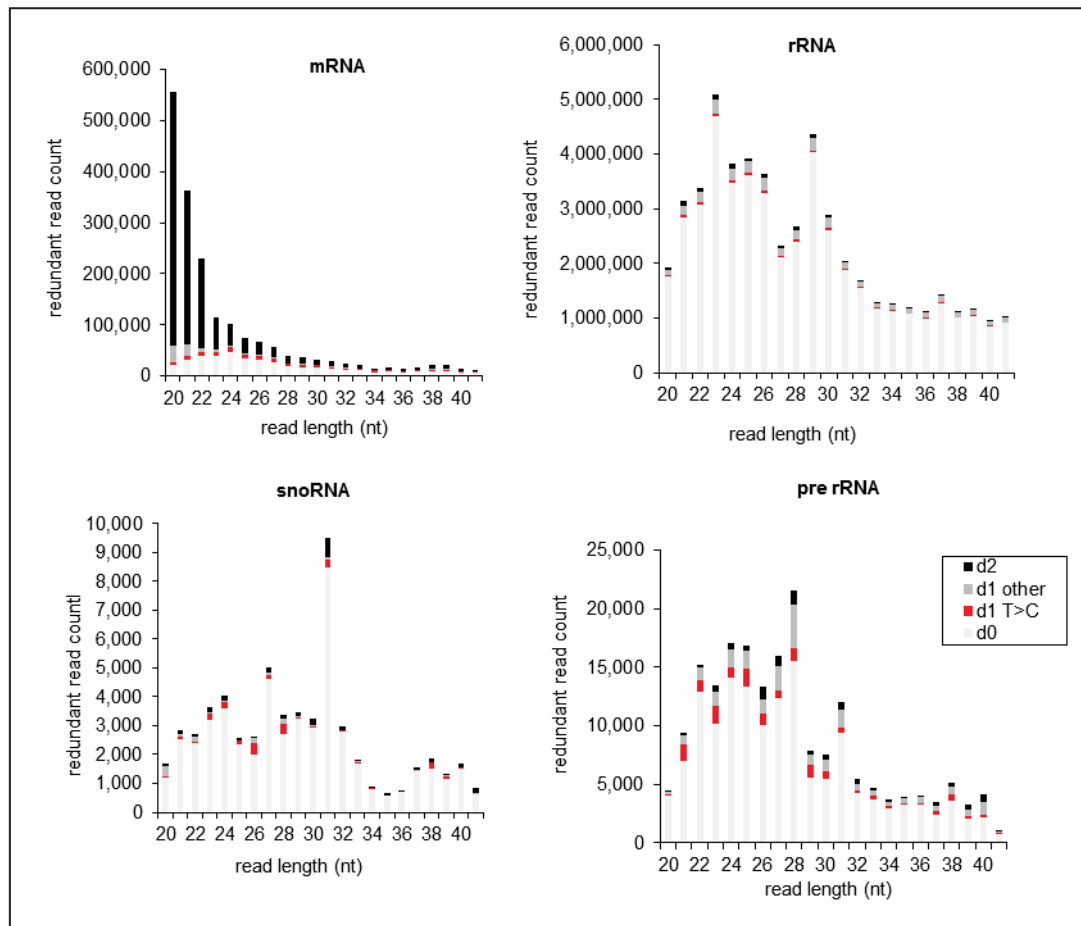
Figure 37: Read distribution of the cDNA PAR-CLIP library of ∆N Nucleolin

PAR-CLIP cDNA library analysis of delta N NCL shows read count versus read length of 20 nt up to 41 nt for mRNA, rRNA, snoRNA and pre rRNA. Two mismatches (d2) depicted in black, 'd1 T>C' implicates a T to C conversion allowing one mismatch (red), 'd1 other' implicates one mismatch except T to C conversion (grey) and no mismatch (d0) is illustrated in white bars.

3.5.4 PAR-CLIP analysis of full length Nucleolin versus ∆N Nucleolin

In order to identify differences in RNA-binding properties between NCL (full length) and

∆N NCL, the T to C conversion events in four different RNA categories (mRNA, rRNA,

pre RNA, snoRNA) in both PAR-CLIP libraries were compared to each other and are

summarized in Table 20. Full length NCL shows a higher ratio of (d1(T→C)/d1) for

snoRNAs (0.9 and 0.7) and pre rRNA (0.7) indicating crosslinking, in contrast to ∆N,

which shows a lower ratio of 0.4 in snoRNAs and pre rRNA. Consequently, it can be

assumed that the N terminus of NCL, although, it does not contain any obvious RNA-

binding domains, might play a role in RNA recognition and interaction with pre rRNA

and snoRNA targets. It was previously described by Ginisty et al. that NCL missing the N terminus shows no interactions with U3 (Ginisty et al., 1998). However, the PAR-CLIP library for the delta N variant was performed only once and shows therefore, only a tendency that needs to be confirmed in a repeat-experiment.

Table 20: Small cDNA PAR-CLIP library analysis of full length versus ∆N Nucleolin

PAR-CLIP cDNA library	category	d1 (T → C)/d1
NCL I (full length)	mRNA	0.5
	rRNA	0.2
	pre rRNA	**0.7**
	snoRNA	**0.9**
NCL II (full length)	mRNA	0.3
	rRNA	0.1
	pre rRNA	**0.7**
	snoRNA	**0.7**
∆N NCL	mRNA	0.5
	rRNA	0.2
	pre rRNA	**0.4**
	snoRNA	**0.4**

The cDNA PAR-CLIP library of Nucleolin full length I and II versus ∆N NCL, lacking the N terminus are grouped in RNA categories including mRNA, rRNA, pre rRNA and snoRNA. The corresponding ratio of T to C conversions to one mismatch in total (d1 (T → C)/d1) refers to the crosslinking sites in each RNA category. If the ratio is close to 1 almost all of the occurring mismatches (d1) contain T to C conversions, implicating crosslinking and binding sites of the RBP. Full length NCL as well ∆N NCL show low T to C conversions in mRNA and mature rRNA. Full length Nucleolin crosslinks to pre rRNA and snoRNAs, whereas ∆N NCL shows reduced T to C conversions. The PAR-CLIP library for NCL I was generated by Jessica Höll.

To summarize, the PAR-CLIP analysis of the delta N NCL variant provides evidence that the D/E rich, N-terminal domain of NCL might be important for RNA-binding.

3.5.5 RIP-Seq analysis of Nucleolin

In order to verify the PAR-CLIP results of NCL a RIP-Seq experiment was performed. The method is explained in more detail in section 2.2.10 and is summarized in Table 8. In RIP-Seq, RNA-protein complexes are immunoprecipitated without prior crosslinking and without previous addition of any photoactivatable ribonucleoside (4SU or 6SG). Bound RNAs were isolated and deep sequenced and expected to be enriched in abundance compared to a lysate control sample.

The RIP-Seq data of NCL shows enrichment of pre rRNA and snoRNAs in the IP sample versus the lysate control (Figure 38). The expression of mRNA and mature rRNA remained almost stable in the IP sample in comparison to the lysate control. The bar diagram shows a fold change (FC) of the IP sample illustrated in black to the non-immunoprecipitated lysate control in grey. To determine changes of the RNA expression levels, the FC is commonly used in the analysis of RNA-Seq data. SnoRNAs and pre rRNA show respectively almost a 3- and 4- fold increase, in comparison to the lysate control.
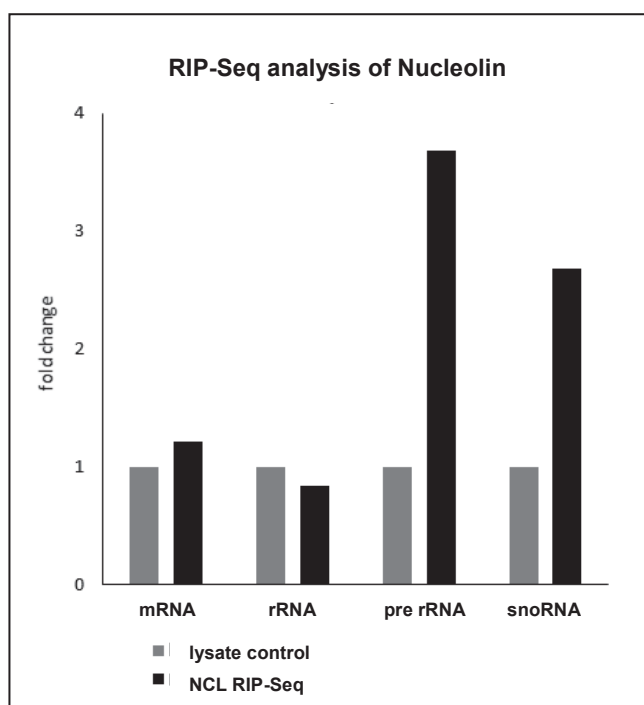
Figure 38: RIP-Seq analysis of Nucleolin

The bar diagram shows the fold change of the Nucleolin RIP-Seq sample grouped in RNA categories including mRNA, rRNA, pre rRNA and snoRNA. The RIP-Seq sample of NCL in HEK293 cells indicated by the black bars (I) revealed enrichment of pre rRNA and snoRNAs, whereas the expression of mRNA and mature rRNA remained almost stable in the IP in comparison to the lysate control sample (I). SnoRNAs show a 2.7-fold change increase, while pre rRNA are increased up to 3.7-fold.

Since the experiment was performed only once, a standard deviation is not included in the bar diagram. As a consequence, the RIP-Seq results show only a trend and provide evidence that pre rRNA and snoRNAs are the main targets of NCL, which is in agreement with the previous NCL PAR-CLIP data (Figure 35).

In order to show a detailed illustration of the read coverage of pre rRNA versus mature rRNA, Figure 39 shows the log2 coverage of reads derived from the RIP-Seq experiment. The red line implicates reads of the NCL-FLAG IP, reads of the lysate control sample are depicted in blue. The first cleavage site (01) in the 5'ETS is shown in green. In accordance to the PAR-CLIP data, read coverage of NCL starts downstream of the first cleavage site. Therefore, it can be supposed that NCL functions in the process of ribosome biogenesis within the 45S rRNA, after the first cleavage event. The NCL FLAG-IP shows enriched read coverage in the precursor region, including the 5'ETS, ITS1 and ITS2. The chart below shows the FC difference of the IP versus the control depicted in a black line. An increasing FC of two-three represents differences in abundance, and is only observed within the precursor regions (5'ETS, ITS1, ITS2) compared to the mature rRNAs, which are indicated by grey bars.
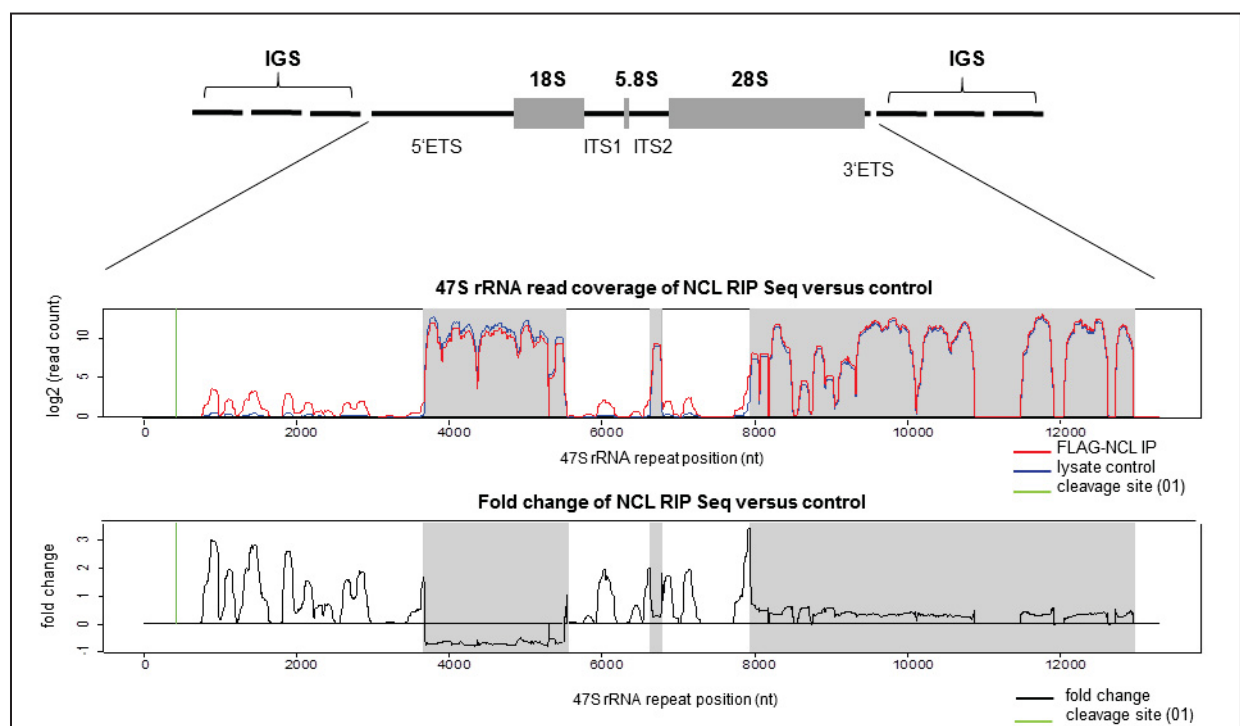
Figure 39: Binding sites of Nucleolin along the 47S rRNA in RIP-Seq analysis

FLAG-Immunoprecipitation of Nucleolin (red line) shows enrichment in the precursor rRNA in the 5'ETS, ITS1 and ITS2 versus the lysate control (blue line). Below, the fold change difference of the lysate control to the NCL-IP in terms of read coverage along the 47S rRNA. The first cleavage site of the transcript within the 5'ETS is indicated in green.

The limitation of this method lies in the analysis of a precise binding site. RIP-Seq only determines an enrichment of the target RNA in the IP sample in comparison to a lysate control sample. To summarize the RNA-interaction of NCL, enrichment of snoRNAs and pre rRNA were identified by RIP-Seq in the FLAG/HA tagged NCL IP sample in contrast to the control. Additionally, crosslinking and binding sites associated with the same cellular RNA targets (snoRNA, pre rRNA) were observed by PAR-CLIP. These results demonstrate that both methods, independently, confirmed the same results and snoRNAs and pre rRNA were identified as specific targets for NCL. NCL is therefore associated with two interactions partners of the transcriptome that are directly related to ribosome biogenesis. Furthermore, this provides evidence that NCL is involved in ribosome biogenesis.

3.5.6 Identification of snoRNAs by PAR-CLIP and RIP-Seq analysis of Nucleolin

As previously demonstrated, the PAR-CLIP and RIP-Seq data revealed snoRNAs as targets for NCL (Figure 35 and Figure 38). Figure 40 illustrates the overlap of the snoRNAs identified by PAR-CLIP and RIP-Seq analysis. A total of 152 snoRNAs were found in the PAR-CLIP data set, whereas 52 snoRNAs were identified by RIP-Seq. There are 21 overlapping snoRNAs, the majority of snoRNAs belongs to the H/ACA family (snoRA), while only five belong to the C/D box family (snoRD). The two snoRNA families are describe in more detailed in section 1.2.4.
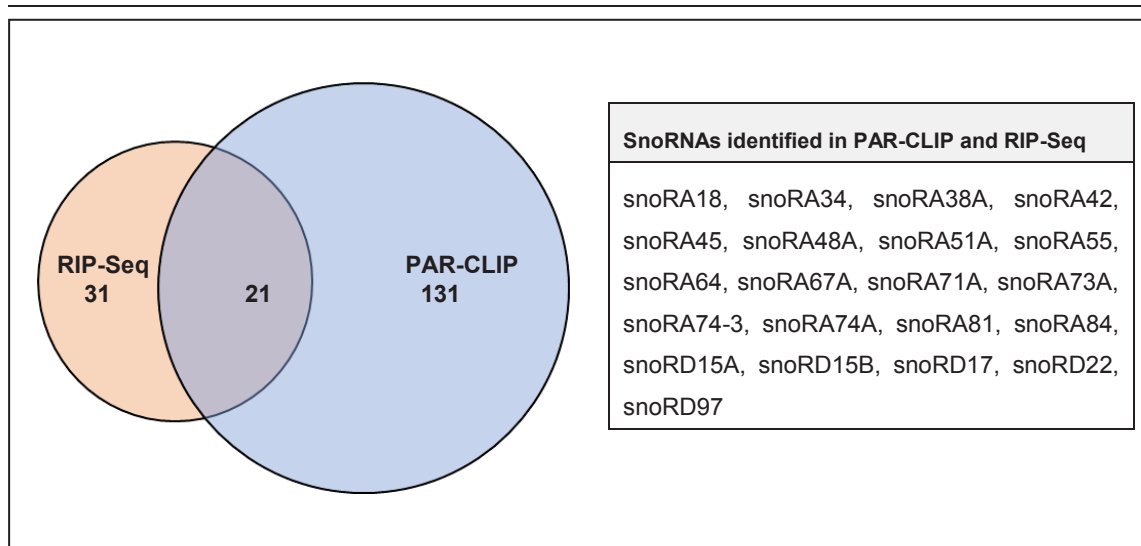
Figure 40: SnoRNAs identified by PAR-CLIP and RIP-Seq of Nucleolin

The venn diagram shows the overlap of snoRNAs that were identified in PAR-CLIP and RIP-Seq analysis of Nucleolin. In total, 152 snoRNAs were identified by PAR-CLIP, 52 snoRNAs associated with NCL were found in RIP-Seq analysis and 21 snoRNAs were identified in both data sets, by PAR-CLIP and RIP-Seq.

3.6 Nucleotide composition of the precursor rRNA transcript

Previous PAR-CLIP and RIP-Seq results of NCL (Figure 36 and Figure 39) demonstrate that NCL preferentially interacts with precursor regions of the rRNA, including the 5'ETS, ITS1 and ITS2, whereas no interaction with the mature rRNA (18S, 5.8S, 28S) and the IGS site was observed. To identify sequence specificities of NCL within the primary precursor transcript, the nucleotide composition was analyzed (Figure 41).

The primary 47S rRNA transcript comprises 35.5% cytosines (C), 36.9% guanosines (G), 15.3% thymines (T) and only 12.1% adenines (A) (Table 21). Consequently, the transcript is very rich in C/G residues and contains up to 72% cytosines and guanosines.
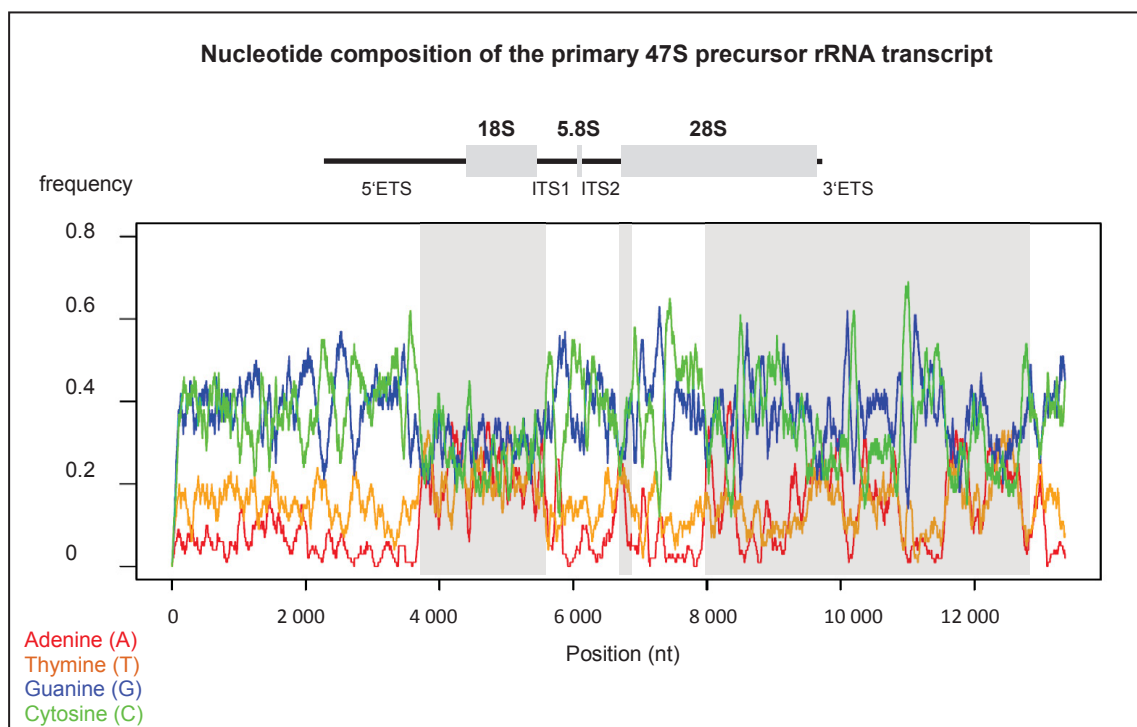
Figure 41: Nucleotide coverage plot of the 47S rRNA

The nucleotide composition of the 13.3 kb long 47S rRNA transcript (NR_046235.1) shows high levels of G and C residues. The 18S rRNA (grey box) shows equal coverage of adenine (red), thymine (orange), guanine (blue) and cytosine (green), whereas the entire 28S (grey box) is mostly rich in G and C residues. The 47S rRNA comprises 1,611 adenines (12.1%), 2,038 thymines (15.3%), 4,727 cytosines (35.5%), and 4,909 guanines (36.9%).

The precursor region including the 5'ETS, ITS1, ITS2 and 3'ETS are rich in G/C residues, whereas the 18S rRNA is equally distributed across all four nucleotides. The 5.8S rRNA is almost equally distributed in A, G and C residues, but contains less T residues, while the 28S rRNA is rich in G and C residues (Table 21).

Table 21: Nucleotide composition of the mature and the precursor rRNA

Region of the rRNA	Sequence length (nt)	A	T	G	C
5'ETS	3,656	207	553	1450	1446
18S	1,871	419	402	551	499
ITS1	1,096	85	139	441	431
5.8S	157	31	36	45	45
ITS2	1,156	59	137	463	497
28S	5,034	800	752	1811	1671
3'ETS	345	10	49	148	138
47S	13,315	1,611 (12.1%)	2,068 (15.5%)	4,909 (36.9%)	4,727 (35.5%)

According to the nucleotide composition analysis of the 47S rRNA in combination with the PAR-CLIP and RIP-Seq results, it can be assumed that NCL most likely binds to G/C rich sequences within the precursor region. This hypothesis will be tested in further biochemical and structural binding assays in 3.8.

3.7 DREME identifies a G-rich binding sequence of Nucleolin

Since it is supposed that NCL binds to G and C rich regions within the pre rRNA, in a next step, a sequence specific binding motif of NCL was determined. The discriminative regular expression motif elicitation (DREME) is a motif discovery algorithm designed by T. L. Bailey to find short, core DNA-binding motifs of eukaryotic transcription factors (Bailey, 2011). A previously conducted PAR-CLIP library of NCL (not shown) was analyzed using DREME in order to identify a specific binding motif. Analyzed by DREME, NCL binds to a G-rich motif sequence (CTGGGGH) with a computed p-value of 3.20E-07. The logo is shown in Figure 42. The bits (0-2) represent the height of each nt within the sequence motif and are proportional to the frequency.

Motif	Positives	Negatives	P-value	E-value
CTGGGGA	17/2,750	3/2,750	1.30E-03	5.80E+01
CTGGGGC	16/2,750	4/2,750	5.80E-03	2.70E+02
CTGGGGT	12/2,750	2/2,750	6.40E-03	2.90E+02
CTGGGGH	45/2,750	9/2,750	3.20E-07	1.50E-02
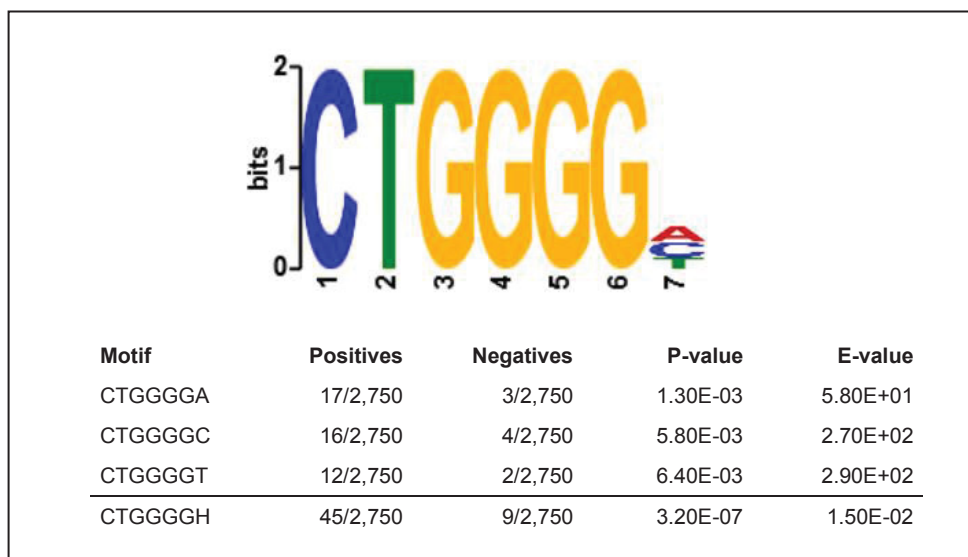
Figure 42: DREME identifies a G-rich binding motif of Nucleolin

The CTGGGGH sequence motif was identified by DREME in a Nucleolin PAR-CLIP library in 45 out of 2,750 sequences in the positive set, while 9 sequence motifs were found in the negative set. The p-value of the positive set is 3.20E-07 and was calculated by the Fisher-Exact Test using a contingency table. All motifs are significantly present in the positive set. DREME also displays the reverse strand with the following sequence motif: HCCCCAG, which is not shown here.

DREME computes the p-value using a contingency table and suggests three significant 7 nt long sequence motifs with a common 6 nt sequence of CTGGGG and a variable nt of A, C or T at position seven. The CTGGGGH sequence motif was identified 45 times out of 2,750 sequences in total in the positive set, nine motifs were found in the negative set. Consequently, the CTGGGGA motif was found 17 times in the positive set, and three times in the negative set, the CTGGGGC motif was identified 16 times in the positive set, and four times in the negative set, while the CTGGGGT motif sequence was found 12 times in the positive and only twice in the negative control set. All three motifs were identified to be significant in the positive data set.

The 6 nt G-rich motif (CTGGGG) is present within the 47S rRNA, in the ITS1, ITS2 and also in the 28S rRNA. Furthermore, this motif is tested in gel-shift assays of NCL using a three repeat sequence of the binding motif (Figure 46).

3.8 Biochemical and structural studies of the Nucleolin-RNA complex

In order to experimentally confirm binding of NCL to the pre rRNA targets identified by PAR-CLIP and the computationally determined G-rich sequence by DREME (Figure 42), biochemical studies including gel-shift assays, ITC and SEC were performed. Additionally, crystallography screens were set up to analyze a three-dimensional structure of the RNA-NCL complex.

3.8.1 Electrophoretic mobility shift assay of the identified Nucleolin-RNA complex

3.8.1.1 Electrophoretic mobility shift assay of Nucleolin testing pre rRNA targets

In order to test binding of NCL with pre rRNA targets that were identified by PAR-CLIP, NCL gel-shift assays were performed. Recombinant NCL containing either the first two RRMs (RRM1-2), all four RRMs (RRM1-4) or a GST (glutathione-s-transferase) tagged

delta N NCL (289-709) were used. The tested RNA targets were either rich in G/C residues or contained a known NCL binding motif like the NRE or the ECM motif (Figure 43). RNA oligonucleotides (oligos) were radioactively labeled and incubated with recombinant NCL for complex formation. To compare the interaction of the complex, which is represented by a shift, the same concentration for each RNA oligo and NCL RRM1-4 were used in every binding reaction. The control sample (5) implicates the radioactive labeled RNA oligo in the absence of NCL. The tested RNA oligos (1-4) (listed in Table 22) formed a complex with NCL RRM1-4, implicated by an intense signal and a size shift of the radioactive RNA oligo. Striking are the two distinct spots in each complex. According to the size, this represents two different binding ratios of the NCL-RNA complex. The complex forms a 1 to 1 binding ratio, representing a more intense spot and a less intense spot increasing in size with a 2 to 1 binding ratio of protein to RNA. Oligo 3 comprises the known ECM motif, contains 75% C/G residues and shows a complete shift, without free radioactive RNA oligo, which represents high binding affinity.
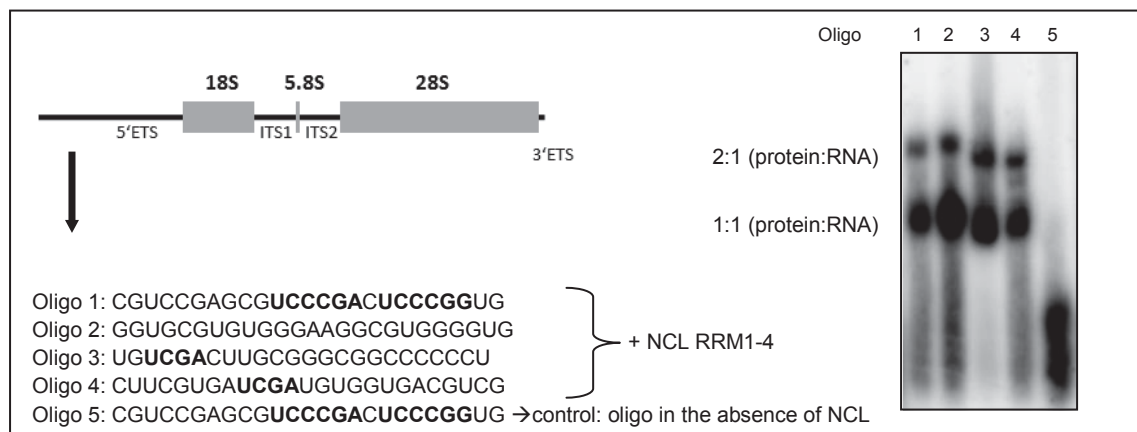
Figure 43: Electrophoretic mobility shift assay of Nucleolin with pre rRNA targets identified by PAR-CLIP

Four different pre rRNA targets identified by PAR-CLIP of NCL were tested for binding in gel-shift assays. The tested RNA oligos contained either known NCL binding motifs, the NCL recognition element (NRE) U/G CCCG A/G, the evolutionary conserved element (ECM) UCGA or G/C rich sequences. The known binding motifs are marked in bold letters. The RNA oligonucleotides were radioactively labeled with [32]P γ ATP and incubated with NCL RRM1-4. The control sample (5) shows free RNA in the absence of NCL. Complex formation of the RNA and the protein is indicated by an intense signal and a size shift of the radioactive RNA oligo (1-4). To compare the intensities of the shift, the same concentration for each RNA oligo and NCL RRM1-4 were used for each binding reaction.

In order to determine sequence specific binding preferences of NCL, the dinucleotide composition of A/U and G/C residues of the tested RNA oligos (1-4) was analyzed. Oligo 1 to 3 contain approximately 75% C/G residues, whereas oligo 4 is almost equally distributed (56% A/U and 44% C/G) and contains the known ECM binding motif (UCGA) like oligo 3. Oligo 1 contains twice the known NRE (UCCCG) motif and is rich in G/C residues (76%).

Table 22: Dinucleotide composition of RNA oligos 1-4

Sequence		A/U content	C/G content
1: CGUCCGAGCG**UCCCG**AC**UCCCG**GUG	(25 nt)	24%	76%
2: GGUGCGUGUGGGAAGGCGUGGGGUG	(25 nt)	28%	72%
3: UG**UCGA**CUUGCGGGCGGCCCCCCU	(24 nt)	25%	75%
4: CUUCGUGA**UCGA**UGUGGUGACGUCG	(25 nt)	56%	44%

To further characterize the binding affinity of the RNA target, which contains twice the known NRE motif, deletion and point mutations of the binding motif were performed and tested in gel-shift assays (Figure 44). To exclude unspecific binding to the GST affinity tag, GST in the presence of RNA was used as a control. Point mutations (G to C) within the motif, resulting in a sequence change of UCCC<u>G</u> to UCCC<u>C</u> did not show any reduced binding effect (oligo 5 versus oligo 6). Both RNA oligos bind with equal affinity to NCL that is indicated by a strong shift and intense signal of the RNA. Deletion of the first and the second NRE motif, shortens the oligo to 15 nt and completely abolishes binding (oligo 7), this implicates either a minimum length of the RNA target or both motifs are responsible for binding and complex formation. Regarding the length of the tested RNA oligo, a 25 nt poly (A) oligo was used as a competitor for oligo 5 and 6, whereas a 20 nt poly (A) oligo was used for the short RNA oligo 3, with an excess of 200- fold and 100- fold, respectively.
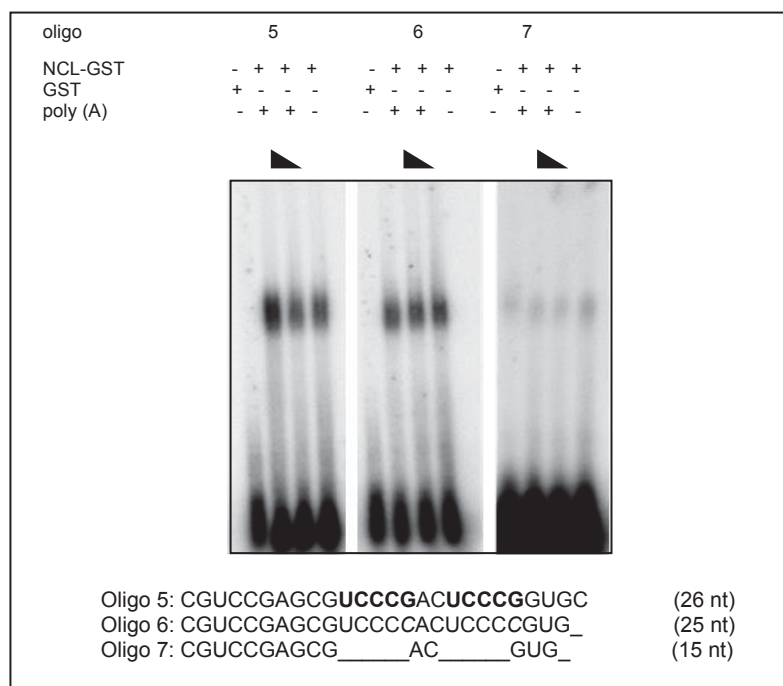
Figure 44: Deletion and mutation analysis of the NRE-motif in electrophoretic mobility shift assays of Nucleolin

For gel-shift analysis recombinant GST tagged Nucleolin lacking the N-terminal domain was used. Oligo 5 was identified in previous NCL PAR-CLIP experiments. The sequence contains twice the known NRE binding motif of NCL with a sequence motif of UCCCG marked in bold letters. Point mutation of the motif indicated in cursive letters and deletion of both motifs (___) were tested in gel-shift assays. According to the length of the tested oligo, 25 nt poly (A) was used as a competitor for oligo 5 and 6, 20 nt poly (A) for oligo 7, with an excess of 200x and 100x. 500 nM of each RNA oligo and 500 ng of GST tagged delta N NCL (289-709) was utilized for complex formation.

The dinucleotide composition of A/U and G/C residues of the previously tested RNA oligos (5-7) was determined and is shown in Table 23. The tested RNA oligos contained approximately 75% G/C residues.

Table 23: Dinucleotide composition of RNA oligos 5-7

Sequence	A/U content	C/G content
5: CGUCCGAGCG**UCCCG**AC**UCCCG**GUGC (26 nt)	23%	77%
6: CGUCCGAGCGUCCCCACUCCCCGUG (25 nt)	24%	76%
7: CGUCCGAGCG_____AC_____GUG (15 nt)	27%	73%

To further characterize the role of the first and the second NRE binding motif, both motifs were separately deleted. To determine if the length of the RNA oligo has also an effect on the binding affinity, the identified RNA sequence was shortened from the 5'

and/or the 3' end, retaining both motifs. The sequences of the tested RNA oligos are illustrated in Table 24 and Figure 45 and contain at least 75% G/C residues.

Table 24: Dinucleotide composition of RNA oligos 8-13

Sequence		A/U content	C/G content
8: _____AGCGUCCCGACUCCCG____	(16 nt)		
9: _____AGCGUCCCGACUCCCGGUGC	(20 nt)		
10: ___UCCGAGCGUCCCGACUCCCG_____	(20 nt)	25%	75%
11: CGUCCGAGCGUCCCGAC_____GUG_	(20 nt)		
12: CGUCCGAGCG_____ACUCCCGGUG_	(20 nt)		
13: CGUCCGAGCGUCCCGACUCCCGGUGC	(26 nt)	23%	77%

Oligo 13 (previously tested as oligo 5) with 26 nt in length shows the most intense shift signal compared to the other RNA oligos (8-13). Oligo 8 contains the first and the second NRE motif, although no shift and consequently no complex formation with NCL was observed. The sequence was trimmed from the 3' and 5' end to a length of 16 nt. However, oligo 9, which contains also both NRE motifs was trimmed from the 5' end, to 20 nt in length and shows a shift that indicates binding. The intensity of the shift signal of oligo 9 is comparable to oligo 11, which lacks the second NRE motif and is also 20 nt in length. Therefore, it can be assumed that the second NRE motif does not contribute to the binding affinity and a maximum length of 20 nt is responsible for interaction. Testing oligo 10 and 12, a less intense shift signal was observed compared to oligo 9 and 11. Both oligos are 20 nt in length, oligo 10 is shortened from the 3' and 5' end and contains both NRE motifs, whereas oligo 12 lacks the first NRE motif. To determine the role of the first and the second NRE binding motif, oligo 11 and 12 are compared to each other. Oligo 12, which lacks the first sequence motif shows lower affinity binding and a weaker shift signal compared to oligo 11, which lacks the second motif. Therefore, the first NRE motif might play a role in complex formation.

	- + -	- + -	- + -	- + -	- + -	- + -	non-radioactive oligo
	+ - -	+ - -	+ - -	+ - -	+ - -	+ - -	poly (A)
	8	9	10	11	12	13	oligo

Oligo 8: _____AGCG**UCCCG**AC**UCCCG**____   (16 nt)
Oligo 9: _____AGCG**UCCCG**AC**UCCCG**GUGC   (20 nt)
Oligo 10: __UCCGAGCG**UCCCG**AC**UCCCG**_____   (20 nt)
Oligo 11: CGUCCGAGCG**UCCCG**AC_____GUG_   (20 nt)
Oligo 12: CGUCCGAGCG_____AC**UCCCG**GUG_   (20 nt)
Oligo 13: CGUCCGAGCG**UCCCG**AC**UCCCG**GUGC   (26 nt)

\} + GST-NCL (289-709)
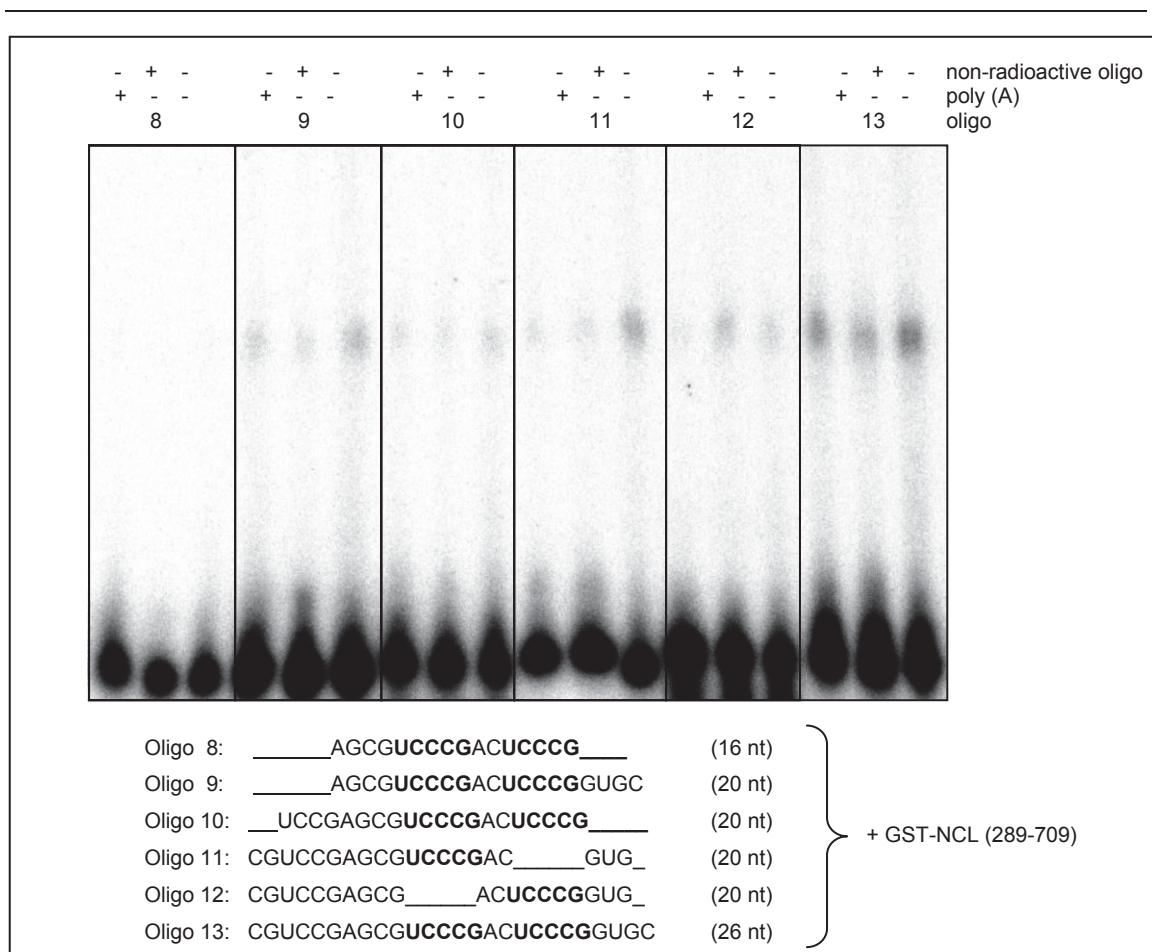
Figure 45: Deletion analysis of a RNA oligo in electrophoretic mobility shift assays

Oligo 8-13 are either shortened at the 3' and/or 5' end of the sequence retaining both known NRE motifs. Oligo 13 represents the identified full length RNA target containing both motifs, whereas the first NRE motif is deleted in oligo 12, the second NRE motif is deleted in oligo 11. The motifs are marked in bold letters, deletions are indicated as a black line (__). Three different binding reactions were tested, either in the excess of a non-radioactively labeled competitor (360x), a poly (A) oligo (20 nt or 25 nt) with a 375x excess or in the absence of any additional oligo. According to the length of the tested RNA oligos, a 25 nt poly (A) competitor was used in the binding reaction with oligo 13, whereas for the other targets a 20 nt poly (A) oligo was used. A total of 520 nM of recombinant GST tagged NCL (289-709) and 35 nM of each RNA oligo was used. Full length oligo 13 shows high binding affinity and illustrates the most intense shift signal compared to the other RNA oligos (8-12). Oligo 8 contains both NRE motifs, the sequence is reduced to a length of 16 nt and binding of the complex is completely abolished. However, oligo 9, which contains both motifs and was trimmed only from the 5' end to a length of 20 nt shows a shift signal, indicating complex formation. The intensity can be compared to the signal of oligo 11, which is also 20 nt in length. Oligo 10 and 12 show a less intense signal. Both oligos are 20 nt in length, oligo 12 lacks the first NRE sequence motif, whereas oligo 10 was shortened from the 3' and 5' end and contains both motifs.

On the one hand, it seems that the first NRE motif in contrast to the second NRE motif within the identified oligo plays an essential role in complex formation and recognition. On the other hand, the length of the sequence has also an impact on the complex formation and contributes to high binding affinity. A 16 nt RNA oligo, even in the presence of both motifs showed no interaction with NCL. In summary, both NRE motifs

alone are not sufficient for high binding affinity, additionally the length of the RNA sequence plays a crucial role. Therefore, this study provides evidence that Nucleolin might not bind sequence-specific to the well-known NRE motif as previously described.

3.8.1.2 Electrophoretic mobility shift assay of Nucleolin testing the G-rich binding motif identified by DREME

To test if the G-rich binding motif, which was previously identified by DRME and is present in the pre rRNA within ITS1, ITS2 and in the 28S rRNA binds to NCL, gel-shift assays were performed. For the gel-shift assays a recombinant GST tagged NCL (289-709) was used (Figure 46). Three RNA oligos with the same length of 18 nts were tested. The identified DREME motif with a three-repeat sequence of CUGGGG (oligo 15) interacts with NCL, a motif that lacks the G-rich repeat and contains six repeats of CUG (oligo 16) and a scrambled sequence (GGCCGGUUCCGGUUCCUU) that function as a control (oligo 14) (Figure 46 and Table 25) showed no binding to NCL. Therefore, interactions with the computationally identified G-rich motif by DREME was confirmed by NCL gel-shift assays.
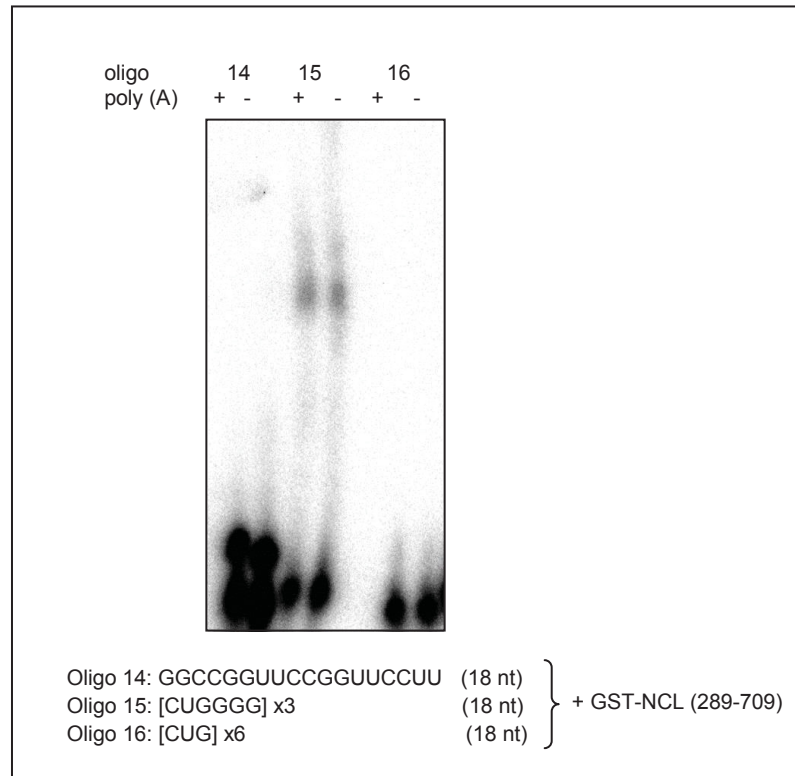
Figure 46: Electrophoretic mobility shift assay of Nucleolin using a G-rich motif identified by DREME

To test binding of the sequence motif identified by DREME, three 18 nt long RNA oligos (14-16) were tested in gel-shift assays using recombinant GST tagged NCL, lacking the N terminus (289-709). The DREME motif with a three-repeat sequence of CUGGGG, a control lacking the G-repeat, (CUGx6) and a scrambled RNA sequence (GGCCGGUUCCGGUUCCUU). Complex formation with NCL, indicated by a shift was only observed in the motif identified by DREME.

The dinucleotide composition of A/U and G/C residues of the tested RNA oligos (14-16) is summarized in Table 25. The DREME motif (oligo 15) contains 83% G/C residues, whereas the other two oligos comprise only 67% G/C residues.

Table 25: Dinucleotide composition of RNA oligos 14-16

Sequence	A/U content	C/G content
14: GGCCGGUUCCGGUUCCUU (18 nt)	33%	67%
15: [CUGGGG] x3 (18 nt)	17%	83%
16: [CUG] x6 (18 nt)	33%	67%

In summary, the pre rRNA targets identified by PAR-CLIP, the G-rich binding motif computed by DREME, the performed binding assays and the nucleotide composition of the 47S rRNA provide evidence that NCL does not bind sequence-specific to a certain motif, but preferentially binds to G-repeat and C/G- rich sequences.

3.8.1.3 Electrophoretic mobility shift assay of Nucleolin testing repeat sequences

To confirm the hypothesis that NCL preferentially interacts with G-rich sequences, G/U and G/C repeat sequences were tested in gel-shift assays using recombinant NCL RRM1-4 (Figure 47). The dinucleotide composition of the RNA oligos regarding A/U and G/C residues was analyzed (Table 26). All of the tested oligos (17-20) bind with high affinity to recombinant NCL. Oligo 17 shows a complete shift and comprises 67% G/C residues, oligo 18 shows a smear of free RNA and contains only 33% G/C-residues, oligo 19 contains 100% G/C residues and oligo 20 contains two of the known NRE motifs and up to 76% G/C residues. Oligo 20 was previously tested as oligo 1 in gel-shift assays and serves as a positive control in this experiment.

Table 26: Dinucleotide composition of RNA oligos 17-20

Sequence		A/U content	C/G content
17: [GGU]x6	(18 nt)	33%	67%
18: [GUU]x6	(18 nt)	67%	33%
19: [GGC]x6	(18 nt)	0%	100%
20: CGUCCGAGCG**UCCCG**AC**UCCCG**GUG	(25 nt)	24%	76%

All of the tested repeat sequences interact with NCL RRM1-4, which is indicated by an intense shift signal (Figure 47). As shown before, oligo 20 (Figure 43) forms a distinct 1 to 1 and 2 to 1 binding ratio of the protein-RNA complex.
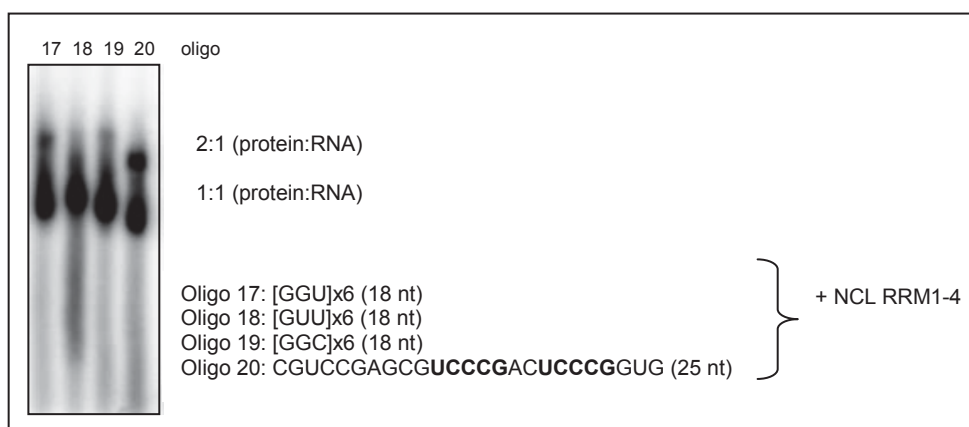
Figure 47: Electrophoretic mobility shift assay of Nucleolin using G-rich sequences

The gel-shift assays were performed using Nucleolin RRM1-4. Different RNA oligos with a length of 18 nt were tested. The previously tested RNA oligo contains twice the NRE motif (UCCCG) and serves as a positive control (oligo 20).

In summary, NCL RRM1-4 binds to G/C as well as to G/U repeat sequences.

To further determine if NCL also binds to poly (A) or A/U rich repeat sequences, gel-shift assays with poly (A), [ACU] and [AUU] repeat oligos were performed using NCL RRM1-4. Table 27 shows the dinucleotide composition of the tested oligos (21-26).

Table 27: Dinucleotide composition of RNA oligos 21-26

Sequence		A/U content	C/G content
21: poly (A)	(20 nt)	100%	0%
22: [ACU]x6	(18 nt)	67%	33%
23: [AUU]x6	(18 nt)	100%	0%
24: [GU]x9	(18 nt)	50%	50%
25: [GUU]x6	(18 nt)	67%	33%
26: [GGC]x6	(18 nt)	0%	100%

No complex formation of NCL RRM1-4 is observed using A/U rich oligos (21-23) (Figure 48), whereas the previously tested G/U and G/C rich oligos (24-26) show an intense shift signal indicating binding.
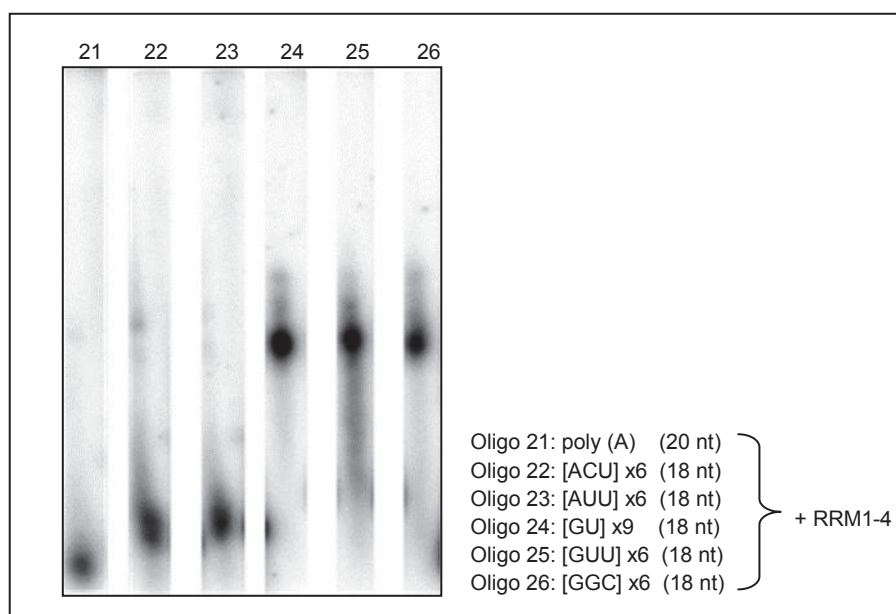
Figure 48: Electrophoretic mobility shift assay of Nucleolin with RNA repeat sequences

To determine if Nucleolin also binds to A/U rich repeat sequences, six different repeat sequences were tested in gel-shift assays using NCL RRM1-4. Poly (A) as well as A/U repeating sequences did not bind to NCL RRM1-4, whereas G/U and G/C rich sequences showed a complete shift indicating high binding affinity.

The GUU-repeat sequence binds to NCL, whereas the AUU-repeat oligo showed no binding, therefore, it can be supposed that a G-residue is crucial for binding and NCL preferentially binds to G or C residues instead to A and U residues. This is also in accordance to the nucleotide composition of the ribosomal precursor region and the identified DREME motif. In summary, besides the identified DREME motif, which is more a G-rich target, NCL does not interact with a sequence-specific binding motif. NCL binds to G-repeats, G/C and G/U rich sequences, rather than to A/U rich sequences. As previously mentioned, the length of the RNA target might also play a crucial role, consequently all of the tested repeat sequences were at least 18 nt in length. To further characterize sequence preferences and binding affinity of the first two RRMs (RRM1-2) and all RRMs (RRM1-4) gel-shift assays, SEC and ITC were performed. Three different G/U repeats (Table 28) with 18 nt in length using RRM1-2 and RRM1-4 were tested in gel-shift assays (Figure 49).

Table 28: Dinucleotide composition of RNA oligos 27-29

Sequence		A/U content	C/G content
27: [GU]x9	(18 nt)	50%	50%
28: [GGU]x6	(18 nt)	33%	67%
29: [GUU]x6	(18 nt)	100%	0%

The tested oligos showed a complete shift and equal binding affinity between RRM1-2 and RRM1-4 was observed.
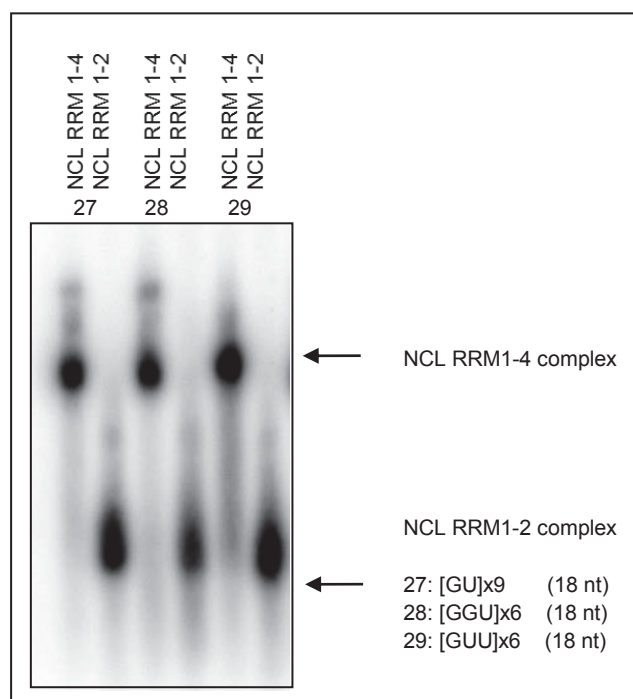
Figure 49: Electrophoretic mobility shift assay of Nucleolin RRM1-2 versus RRM1-4 using G/U-repeat sequences

To determine binding characteristics of the four RRMs in Nucleolin, recombinant RRM1-2 and RRM1-4 were generated and tested with G/U repeat sequences in gel-shift assays. The shift of the RRM1-2 and RRM1-4 complex are indicated with arrows. All of the tested RNA oligos (27-29) show a complete shift in RRM1-2 and RRM1-4.

3.8.2 Size-exclusion chromatography of the Nucleolin-RNA complex

3.8.2.1 Size-exclusion chromatography of Nucleolin testing a pre rRNA target

Previous gel-shift analyses revealed a NCL-pre rRNA binding ratio of two-to-one in oligo 1-4 and 20, whereas a binding ratio of two-to-one was not observed in oligo 18 and 19 (Figure 43 and Figure 47). To further determine the binding ratio of the NCL-RNA complex, size-exclusion chromatography (SEC) also known as gel-filtration chromatography, was performed. In general, in a SEC experiment, RNA-bound protein complexes migrate faster through the column in contrast to free protein or RNA. Consequently, components with higher MW elute in earlier fractions than components with lower MW. Therefore, SEC is able to measure the stoichiometry of the complex. For the NCL SEC experiment two different variants of the purified recombinant NCL

proteins, RRM1-2 and RRM1-4 were separately mixed with oligo 1 (CGUCCGAGCGUCCCGACUCCCGGUG). Oligo 1 is a target sequence of NCL that was identified by PAR-CLIP and contains twice the well-known NRE motif, which is known to recognize the first two RRMs of NCL.

Table 22 shows the dinucleotide composition of the tested RNA oligo. Figure 50 and Figure 51 illustrate the results of the SEC experiments, the absorbance in mAU is plotted on the y-axis, while, the volume of the fraction in mL according to the molecular weight of the eluted components is plotted on the x-axis. Three peaks were observed for the protein-RNA complex (Figure 50). The peaks represent a different MW form of the complex and a distinct binding ratio of three-to-one, two-to-one and one-to-one of RRM1-2 to the RNA target sequence, respectively. The highest peak indicates a two-to-one binding ratio, which implicates that two recombinant NCL RRM1-2 protein variants are required for complex formation with the tested RNA oligo. Since the well-known NRE motif recognizes the first two RRMs (RRM1-2), it can be supposed that each NRE motif within the tested RNA sequence (oligo 1) binds only to the first two RRMs and consequently a two-to-one binding ratio was observed.
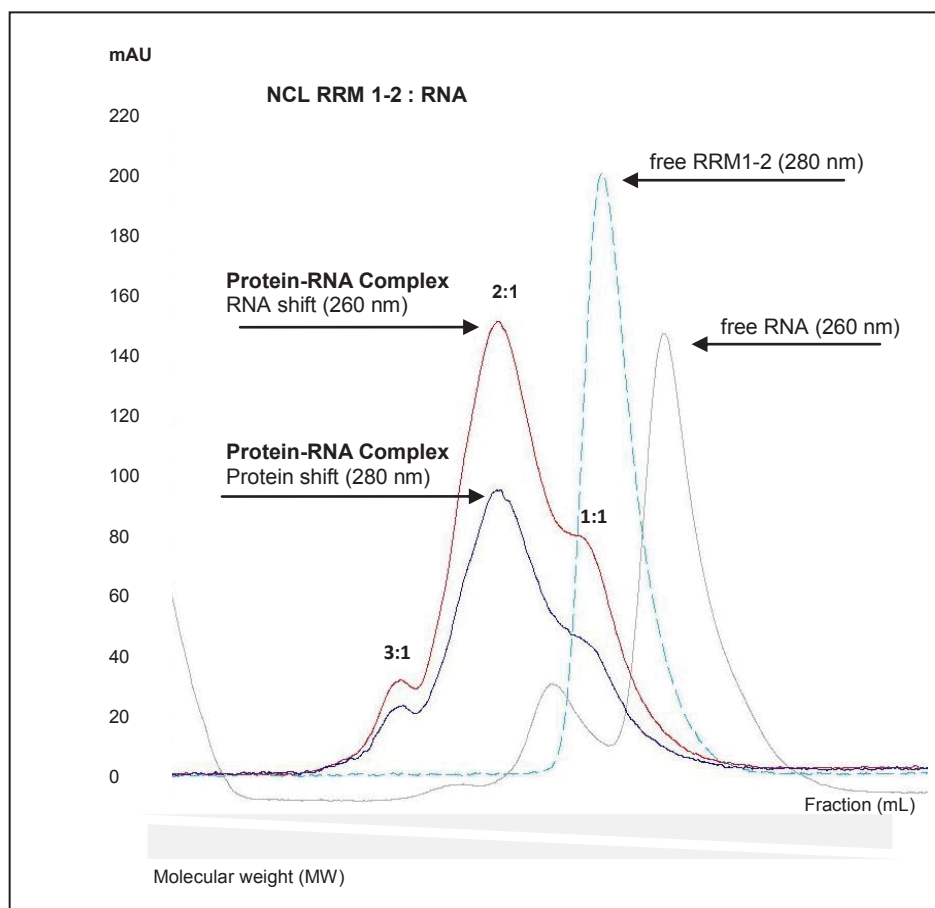
Figure 50: Size-exclusion chromatography of the Nucleolin RRM1-2-RNA complex

The absorbance spectrum of SEC using NCL RRM1-2 and the 25 nt long RNA sequence; CGUCCGAGCGUCCCGACUCCCGGUG (oligo 1), containing twice the NRE binding motif was identified by PAR-CLIP analysis. The protein, the RNA and the protein-RNA complex were analyzed separately according to the absorbance spectrum at 280 nm and 260 nm, for protein and RNA, respectively. The complex shows three peaks, each represents a distinct ratio of 3 to 1, 2 to 1 and 1 to 1 of protein to RNA.

To further test the binding ratio of all four RRMs of NCL with the same RNA target (oligo 1), SEC experiments with the RRM1-4 variant were performed. The complex is indicated by two equally distributed binding peaks by the RNA and the protein (Figure 51). Each peak represents a distinct binding ratio of two-to-one and one-to-one. This implicates that two NCL RRM1-4 variants can bind to one RNA sequence, which would be in accordance with the previous result (Figure 50) indicating that the NRE motif is recognized by RRM1-2. The one-to-one binding ratio supposes that one NRE motif interacts with the first two RRMs as described, while the second motif could interact with RRM3-4 of NCL.
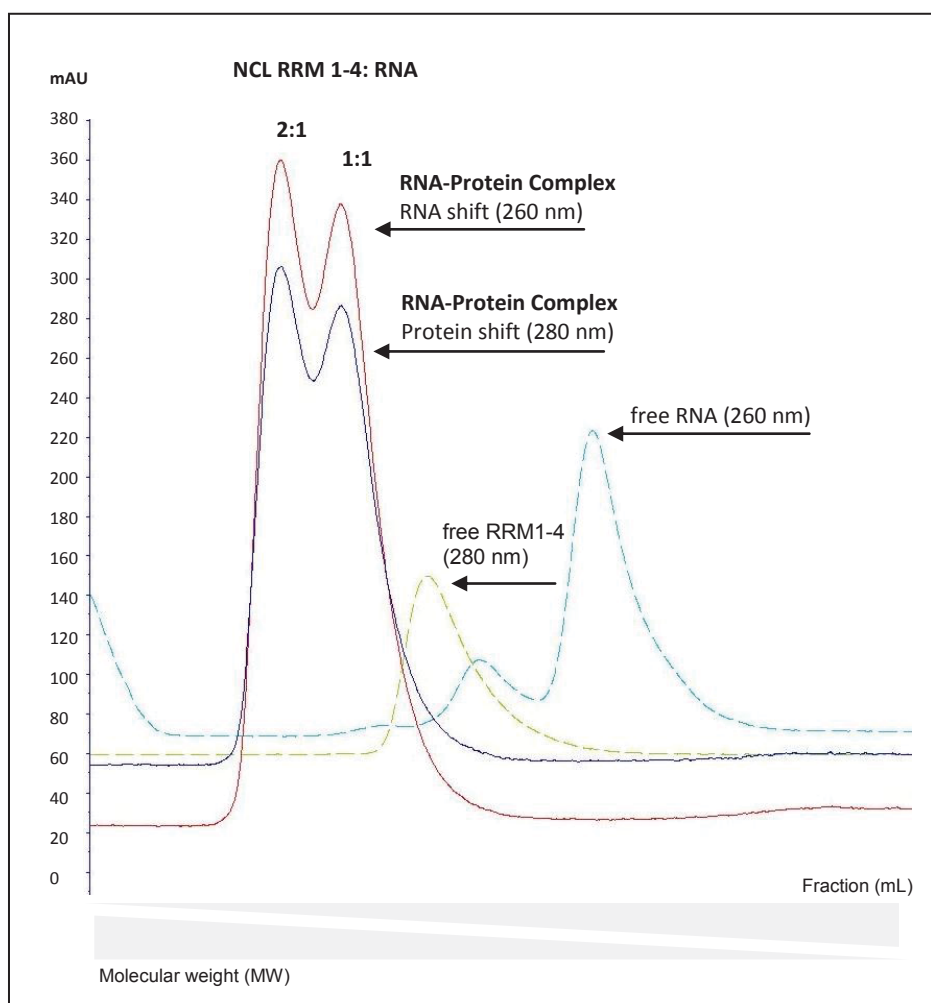
Figure 51: Size-exclusion chromatography of the Nucleolin RRM1-4-RNA complex

The absorbance spectrum of SEC using NCL RRM1-4 and the 25 nt long RNA sequence, CGUCCGAGCGUCCCGACUCCCGGUG (oligo 1), which was identified by PAR-CLIP analysis. The protein, the RNA and the protein-RNA complex were analyzed separately according to the absorbance spectrum at 280 nm and 260 nm, for protein and RNA, respectively. The complex shows two peaks, each represents a distinct ratio of 2 to 1 and 1 to 1 of protein to RNA. The two different binding ratios are equally distributed.

3.8.3 Isothermal-Titration Calorimetry of the Nucleolin-RNA complex

Isothermal-Titration Calorimetry (ITC) detects the thermodynamic interaction of an interactive system and allows the characterization of a binding complex between a protein and a small molecule. The binding reaction of a protein to its ligand results either in heat absorbance or heat release. ITC is used to determine the K_d value and the stoichiometry (N) of three different RNA oligos (oligo 1, 30, 31) using the two

different variants of NCL, containing either the first two RRMs (RRM1-2) or all four RRMs (RRM1-4). Two pre rRNA target sequences (oligo 1 and 30) that were identified by PAR-CLIP and an AC-repeat sequence (oligo 31) were tested in ITC measurements.

3.8.3.1 Isothermal-Titration Calorimetry of Nucleolin testing a pre rRNA target sequence

ITC measurements of NCL RRM1-2 and RRM1-4 using the previously tested 25 nt long pre rRNA sequence (oligo 1) that contains twice the well-known NRE binding motif showed a two-site binding model (Figure 52). A two-site model implicates an exothermic and endothermic binding reaction, which is implicated by negative and positive peaks upon titration in ITC measurements. The protein-RNA complex using RRM1-2 shows a K_d of 662 nM, which implicates lower binding affinity than the RRM1-4 - RNA complex with a K_d of 165 nM. The stoichiometry (N) represents the binding sites of RRM1-2, which is 0.55 $\pm$ 0.72 and 0.38 $\pm$ 0.008 for RRM1-4. This implicates a two-to-one binding ratio of the protein to RNA. If N is 1, the ratio of protein to RNA is equally distributed in a one-to-one binding ratio, whereas if N is 0.5 it refers to a two-to-one ratio. These results are in agreement with the previous SEC results (Figure 50 and Figure 51), that showed also different binding ratios of NCL to the tested RNA.

NCL - RRM1-2

CGUCCGAGCGUCCCGACUCCCGGUG (25 nt)

NCL - RRM1-4

CGUCCGAGCGUCCCGACUCCCGGUG (25 nt)

Model: Two sites
N2: 0.550 ± 0.717 Sites
K2: 1.51E6 ± 4.04E5 M -1
H2: -3.147E4 ± 7.14E4 cal/mol
S2: -77.3 cal/mol/deg

Model: Two sites
N2: 0.384 ± 0.00766 Sites
K2: 6.04E6 ± 7.90E5 M -1
H2: 4.058E4 ± 1.46E3 cal/mol
S2: 167 cal/mol/deg
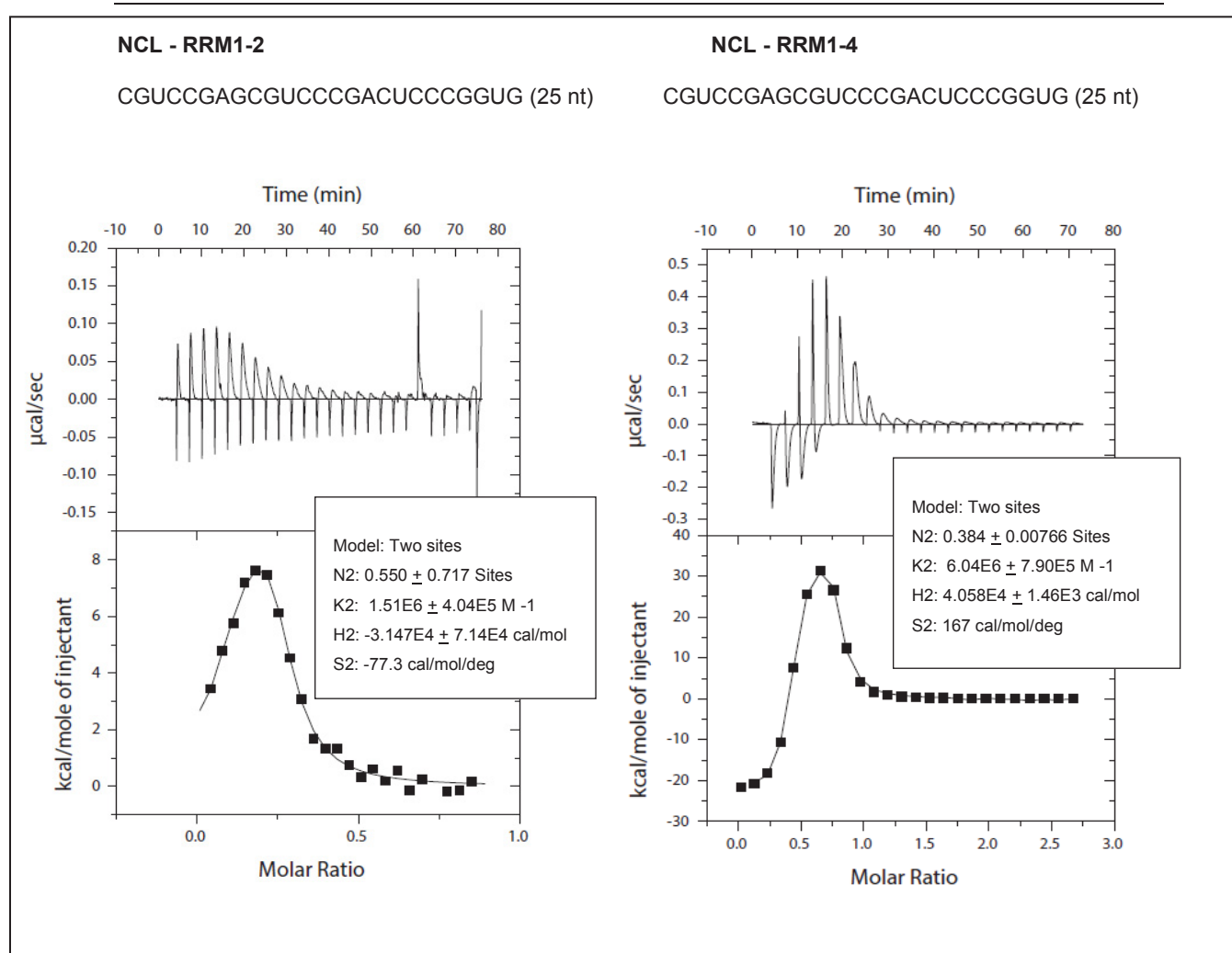
Figure 52: Isothermal-titration calorimetry of Nucleolin testing a pre rRNA sequence identified by PAR-CLIP

ITC measurements of NCL RRM1-2 and RRM1-4 using a 25 nt long RNA sequence, previously identified in PAR-CLIP analysis, show a two-site binding model. This is indicated by heat release and absorbance and shows negative and positive peaks upon titration. The stoichiometry (N) for the RRM1-2-RNA complex is 0.55 ± 0.72, whereas for the RRM1-4-RNA complex N is 0.38 ± 0.008. If N is 0.5, it represents a two-to-one ratio of protein to RNA. If N is 1, the ratio of protein to RNA is equally distributed. The RRM1-2-RNA complex shows a K_d of 662 nM, whereas the RRM1-4-RNA complex shows a lower K_d of 165 nM and therefore, a higher binding affinity to the tested RNA sequence.

ITC measurements of NCL RRM1-2 and RRM1-4 using a 23 nt long G-rich pre rRNA sequence shows a one-site binding model (Figure 53) and consequently only an exothermic binding reaction, which is indicated by heat release and negative peaks upon titration. The NCL RRM1-2 -RNA complex has a K_d of 584 nM implicating lower binding affinity than the RRM1-4-RNA complex with a very low K_d of 19 nM, which represents very high binding affinity to the tested RNA sequence. The stoichiometry of

the complex is 0.58 ± 0.01 for RRM1-2 and 0.66 ± 0.0005 for the RRM1-4 (Figure 53),

which indicates almost a two-to-one binding ratio.
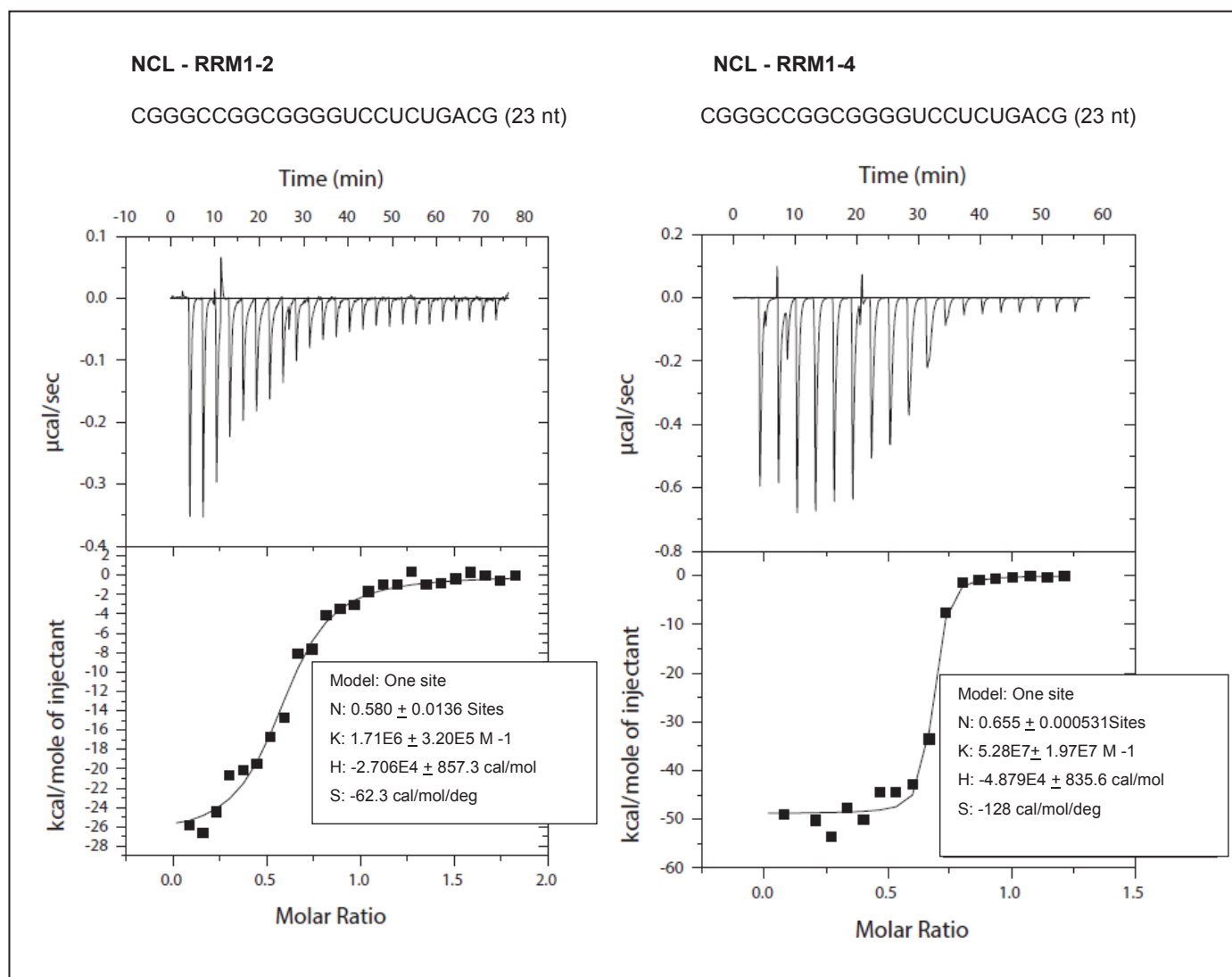
Figure 53: Isothermal-titration calorimetry of Nucleolin testing a G-rich pre rRNA sequence

ITC measurements of NCL RRM1-2 and RRM1-4 with a 23 nt long G-rich RNA sequence show a one-site binding model and an exothermic binding reaction upon titration, which is indicated by negative peaks and heat release. The stoichiometry (N) for the RRM1-2-RNA complex is 0.58 ± 0.01, while for the RRM1-4-RNA complex N is 0.66 ± 0.0005. If N is 0.5, it represents a two-to-one ratio of protein to RNA. If N is 1, the ratio of protein to RNA is equally distributed. The protein-RNA complex of RRM1-2 shows a K_d of 584 nM, whereas the complex of RRM1-4 shows a lower K_d of 19 nM and therefore, a much higher binding affinity to the tested RNA sequence.

3.8.3.2 Isothermal-Titration Calorimetry of Nucleolin testing a AC-repeat sequence

ITC measurements of NCL RRM1-2 and RRM1-4 testing a 18 nt long AC-repeat sequence (oligo 30) shows an one-site binding model and an exothermic binding reaction upon titration. The protein-RNA complex of RRM1-2 has a K_d of 4,500 nM indicating a very low binding affinity, in contrast to the RRM1-4-RNA complex, which shows a K_d of 118 nM and represents therefore, a much higher binding affinity. The stoichiometry of the RRM1-2-RNA complex is 0.62 $\pm$ 0.04 and 0.46 $\pm$ 0.003 for the RRM1-4-RNA complex (Figure 54), indicating a two-to-one binding ratio.
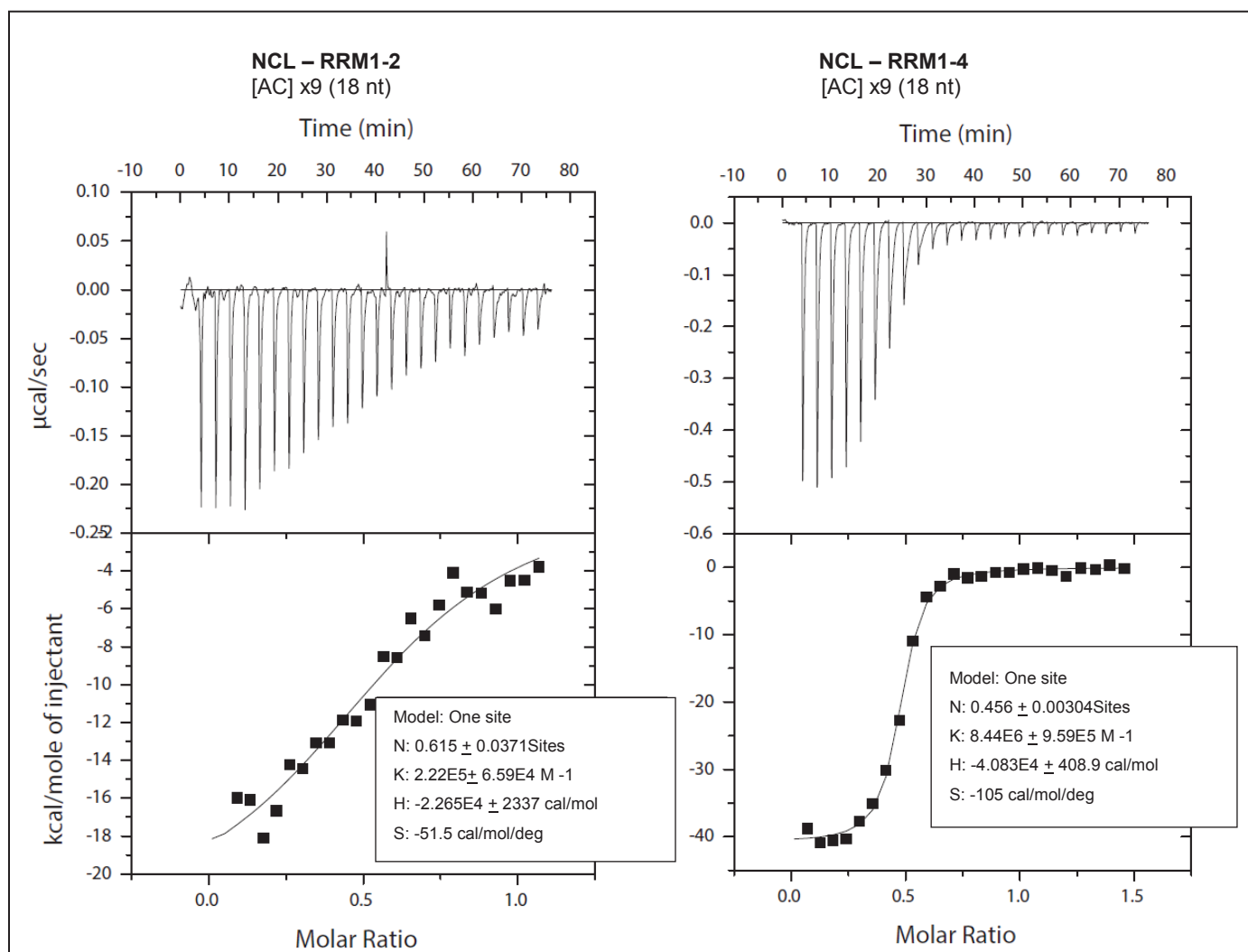
Figure 54: Isothermal-titration calorimetry of Nucleolin testing a [AC]-repeat sequence
ITC measurements of NCL RRM1-2 and RRM1-4 with a 18 nt AC repeat sequence show a one-site binding model and an exothermic binding reaction, which is indicated by negative peaks and heat release. The stoichiometry (N) for the RRM1-2 protein-RNA complex is 0.62 $\pm$ 0.04, whereas for the RRM1-4 RNA complex N is 0.46 $\pm$ 0.003. If N is 0.5, it represents a two-to-one ratio of protein to RNA. If N is 1, the ratio of protein to RNA is equally distributed. The protein-RNA complex of RRM1-2 shows a K_d of 4,500 nM, the complex of RRM1-4 shows a much lower K_d of 118 nM and therefore, a higher binding affinity to the tested repeat sequence.

Table 29 summarizes the ITC results, including K_d and N of the tested RNA sequences and either with NCL RRM1-2 or RRM1-4. In general, RRM1-4 shows a higher binding affinity than RRM1-2. The G-rich sequence (oligo 30) and RRM1-4 show the lowest K_d (19 nM), which therefore, represents the highest binding affinity of the tested RNA targets. The stoichiometry for RRM1-2 and RRM1-4 of the pre rRNA sequence (oligo 1) is 0.55 $\pm$ 0.72 and 0.38 $\pm$ 0.008, respectively. This is in agreement with the previous SEC and gel-shift analysis, which showed a two-to-one binding ratio of NCL to the pre rRNA sequence. Although, the tested RNA targets vary in its sequence length, in summary it can be supposed that NCL preferentially binds to G- and C- rich RNA sequences. In general, the G-rich pre rRNA target binds with higher binding affinity to NCL (RRM1-2 and RRM1-4) than oligo 1, which contains twice the well-known NRE motif. RRM1-4 binds with high affinity to the AC-repeat sequence, while RRM1-2 shows very low binding affinity to the same repeat sequence. Consequently, it can be assumed that the last two RRMs of NCL (RRM3-4) are responsible for the interaction with the A/C repeat sequences. To confirm the ITC results, the AC-motif could be further tested in gel shift assays.

Table 29: Binding affinity and stoichiometry of Nucleolin testing three different RNA sequences

Sequence	A/U content	G/C content	RRM 1-2	RRM 1-4
Oligo 1: CGUCCGAGCGUCCCGACUCCCGGUG (25 nt)	24%	76%	K_d = 662 nM N = 0.55 $\pm$ 0.72	K_d = 165 nM N = 0.38 $\pm$ 0.008
Oligo 30: CGGGCCGGCGGGGUCCUCUGACG (23 nt)	17%	83%	K_d = 584 nM N = 0.580 $\pm$ 0.01	K_d = 19 nM N = 0.655 $\pm$ 0.0005
Oligo 31: [AC]x9 (18 nt)	50%	50%	K_d = 4,500 nM N = 0.615 $\pm$ 0.037	K_d = 118 nM N = 0.456 $\pm$ 0.003

3.8.4 Crystallography of the Nucleolin-RNA complex

3.8.4.1 Crystallography of Nucleolin testing a pre rRNA target

Furthermore, structural analysis can provide crucial information about this specificity of the RNA-binding protein and the interaction with the RNA targets. Based on the results obtained by electrophoretic mobility shift assay (Figure 43), SEC (Figure 50 and Figure 51) and ITC measurements (Figure 52), crystallography trials of the NCL-RNA complex were performed. Initial trials were carried out using the pre rRNA target sequence (oligo 1), which was identified by PAR-CLIP and previously tested for interaction in gel-shift assays, SEC and ITC analysis. Commercially available crystallography kits with various conditions including different salt concentration, buffer compositions and pH were used for initial trails for structural analysis of the complex. In order to let the crystals grow, the complex was incubated for several weeks, before they were sent to the synchrotron for high-resolution structure determination. Usually, crystals generate diffraction patterns when exposed to x-rays, these diffraction data were then considered to extract structural information. Unfortunately, crystals for the tested NCL-RNA complex did not diffract. The complex is disordered and therefore not suitable for structural analysis. This could be consistent with the two-site binding model, that was observed in ITC measurements (Figure 52) and the formation of two complexes upon titration in gel-filtration assays also known as SEC (Figure 50 and Figure 51). A three-dimensional crystal structure of the complex could not be dissolved; all of the crystal like fragments (Figure 55) showed a two-dimensional structure, which were not sufficient for structure determination and further analysis.

Condition I:
0.1 M sodium acetate pH 4.6,
2.0 M ammonium sulfate

Condition II:
0.1 M Tris·HCl pH 8.5,
2.0 M ammonium sulfate

Condition III:
0.1 M sodium phosphate, 0.1 M potassium phosphate,
0.1 M MES pH 6.5,
2.0 M sodium chloride
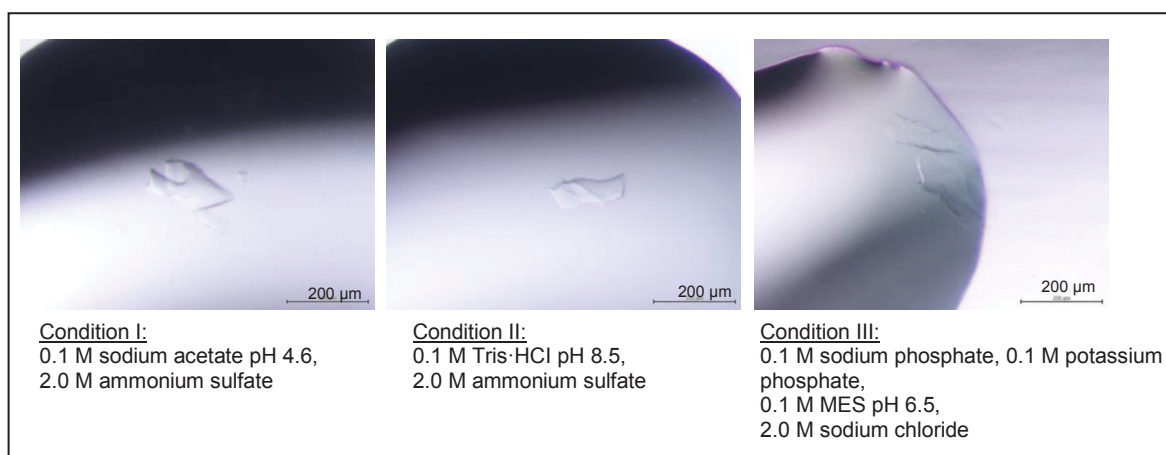
Figure 55: Crystallography of Nucleolin testing a pre rRNA sequence that was identified by PAR-CLIP

Crystallography of NCL RRM1-2 using a pre rRNA sequence (oligo1: CGUCCGAGCGUCCCGACUCCCGGUG), previously identified in PAR-CLIP analysis, was performed. Different conditions from commercially available kits (I: 0.1 M sodium acetate pH 4.6, 2.0 M ammonium sulfate, II: 0.1 M Tris·HCl pH 8.5, 2.0 M ammonium sulfate, III: 0.1 M sodium phosphate, 0.1 M potassium phosphate, 0.1 M MES pH 6.5, 2.0 M sodium chloride) were used and tested to determine the crystal structure of the NCL-RNA complex. All of the (I-III) crystal looking fragments showed a 2D structure, therefore, the structure could not be dissolved and the fragments were not sufficient for structure determination.

Similar two-dimensional crystal-like-fragments were observed in the following conditions and are listed in Table 30.

Table 30: Initial crystallography conditions for Nucleolin

Kit	Composition
Classics E5	0.1 M Tris pH 8.5, 12% (v/v) Glycerol, 1.5 M Ammonium sulfate
Classics E7	0.01 M Ferric chloride, 0.1 M tri-Sodium citrate pH 5.6, 10% (v/v) Jeffamine M-600
Classics E8	0.1 M HEPES pH 7.5, 20% (v/v) Jeffamine M-600
Classics II C1	3.5 M Sodium formate pH 7.0
Classics II G3	0.2 M Lithium sulfate, 0.1 M Bis-Tris pH 6.5, 25% (w/v) PEG 3350

3.9 The interaction network of Nucleolin analyzed by Mass-Spectrometry

In a next step, protein complexes and networks associated with NCL were characterized and identified by mass-spectrometry (MS or mass-spec). Therefore, immunoprecipitation of FLAG/HA tagged NCL, NPM1 and YBX1 using unsynchronized HEK293 cells was performed. Additionally, a FLAG-IP in HEK293 cells using an empty vector control was performed.

NCL and NPM1 are both involved in ribosome biogenesis and function as baits in the MS approach and in the corresponding data analysis using SAINT*express* and SAM. SAINT considers the empty vector and YBX1 as negative controls. The analyses are explained in more detail in section 2.2.13.4. SAINT utilizes spectral counts, which were generated by mass-spec in order to determine bait-prey pair interactions. To compute the probability of a true interaction, SAINT uses the Bayes rule and a discrete probability distributions like the Poisson distribution (Teng et al., 2015). The spectral counts of a protein generated by mass-spec analysis are countable and are expected to follow the Poisson distribution.

Additional SAM analysis was performed in order to identify significant preys between two groups, NCL and NPM1. SAM addresses the problem of multiple testing and estimates a FDR by using permutations and computes observed d-values (Tusher et al., 2001). Expected and observed d-values are plotted in a SAM diagram to display the significant preys.

Western blot analysis shows the expression of FLAG/HA tagged NCL in the corresponding HEK293 lysate in comparison to the empty vector control, while β-actin functions as an internal loading control. Additionally, the samples were silver stained to highlight all proteins in each IP sample (Figure 56) and subsequently, tested in downstream mass-spec analysis.
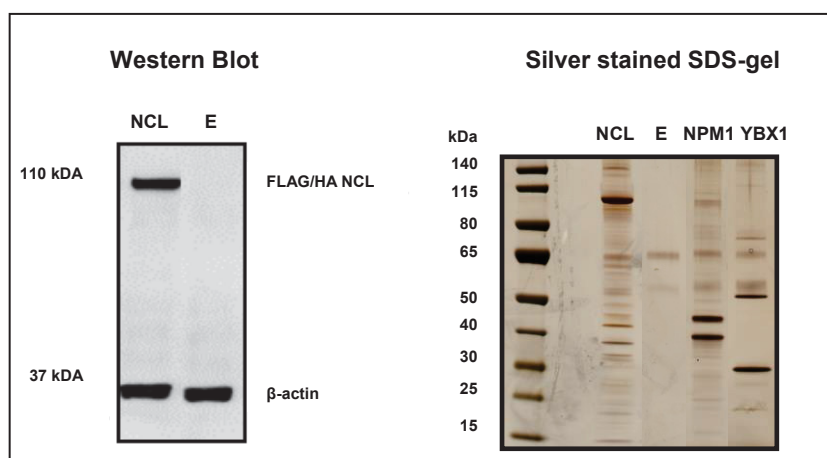
Figure 56: Western Blot analysis and silver staining of the FLAG-IP samples

Western Blot analysis of FLAG/HA tagged NCL in HEK293 cells and an empty vector, which serves as a negative control. An anti β-actin antibody was used as an internal control for normalization. The silver stained FLAG-IP samples for NCL, Empty (E), NPM1 and YBX1 were used for further downstream analysis in mass-spectrometry. HEK293 cells stably expressing FLAG/HA tagged NPM1 and YBX1 were generated by lab members from the Höll group.

NCL shows a distinct and expected band at ~110 kDa in the FLAG detected western blot and in the silver stained SDS-gel. YBX1 has a molecular weight of 36 kDa, a band at the expected MW was observed. NPM1 has an observed MW of 35-40 kDa and shows a distinct band at 40 kDa. The empty vector control shows unspecific background binding to the beads that are linked to an anti-FLAG antibody, this appears to be background in each sample (Figure 56).

3.9.1 Nucleolin interacts with factors associated with ribosome biogenesis

A total of 775 NCL co-purified proteins were identified by mass-spec. This data set contains several non-specifically binding proteins. The first filtering steps comprise a minimum of two identified peptides and five quantitative values in at least one of the four sample groups in order to further exclude false-positive proteins. Consequently, the list was reduced to 606 proteins with at least two different identified peptides and further down to 275 proteins with increasing stringency showing five quantitative intensity values. These co-purified proteins were considered for further SAINT*express* analysis (2.2.13.4). In a first step, the robust spectral count data were used to identify NCL co-purified proteins by calculating probability scores using SAINT*express* (version 3.6.1) (Choi et al., 2011; Teo et al., 2014). In this approach, NCL and NPM1 were considered as independent baits, and YBX1 and the empty vector sample were used as negative controls. In general, the statistical model of SAINT is a mixture model (Choi et al., 2012a). The spectral counts are modeled with a so called mixture distribution of two components that represents true and false interaction for each prey-bait pair. According to Choi et al. our setup refers to a semi-supervised mixture model (Choi et al., 2011).

SAINT estimates the Bayesian false discovery rate (BFDR) from the SAINT probability scores. Significant NCL binding partners were considered if the BFDR was estimated $\leq$ 5% ($p \leq 0.05$) and with a SAINT probability scores of 0.79-1. The higher the SAINT probability score, the higher the probability of a true prey-bait pair. In total, 119 proteins with higher abundance in NCL (Figure 58) and 40 proteins associated with NPM1, in comparison to the empty vector and YBX1 were identified by SAINT.

The majority of the NCL associated proteins play an essential role in ribosome biogenesis, including 47 ribosome biogenesis factors, 39 ribosomal proteins, 11 RNA-helicases, two subunits of the FACT complex (SUPT16H, SSRP1) and 3 histones of the H1 histone family, which are known as linker histones and are able to activate rDNA transcription (Leary and Huang, 2001). Additionally, a total of 17 other proteins that are not involved in ribosome biogenesis were identified. The different groups are shown and summarized in the pie chart in Figure 57.
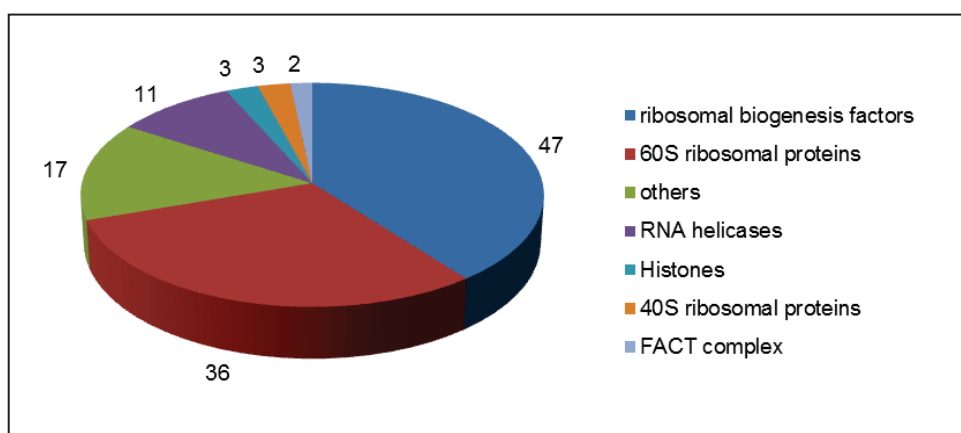
Figure 57: Interaction partners of Nucleolin identified by Mass-Spectrometry

SAINT*express* analysis of mass-spectrometry data identified 119 proteins associated with Nucleolin. The pie chart illustrates proteins grouped into ribosome biogenesis factors (47), ribosomal proteins of the small (3) and the large subunit (36), the two subunits of FACT complex (2), RNA-helicases (11), histones (3) and others (17), representing proteins that are not involved in ribosome biogenesis.

The following heat map in Figure 58 represents the mean intensity of the five replicates (n=5) of empty, NCL, YBX1 and NPM1 and the identified interaction partners (preys). The measured intensities are used for the comparison of protein abundance between the four sample groups.
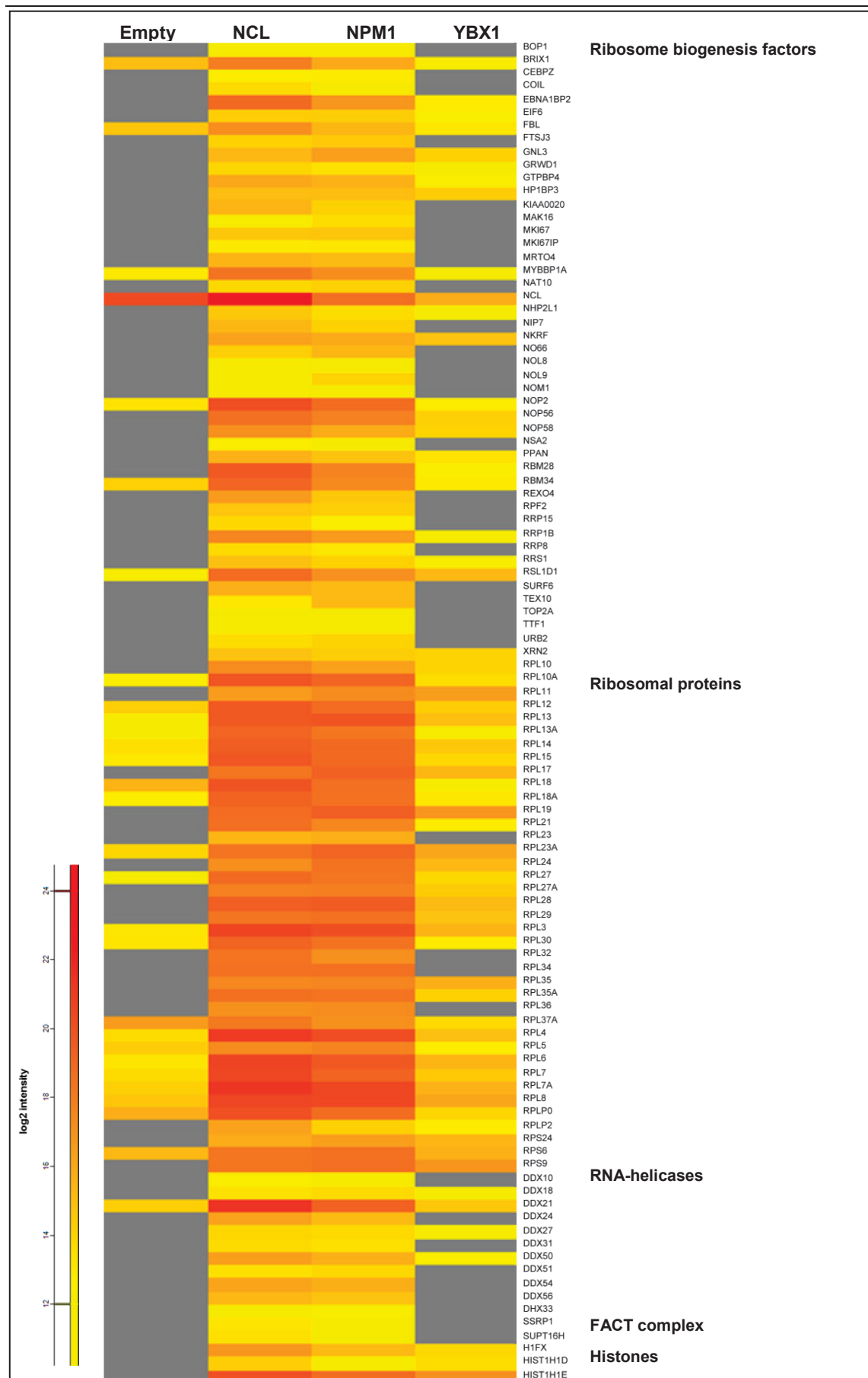
Figure 58: Heat-map of the interaction partners identified by Mass-Spectrometry

The mass-spectrometry derived mean intensities of five replicate analyses from co-immunoprecipitations are illustrated in a heat map. The empty vector and YBX1 serve as unspecific controls, whereas NCL and NPM1 are used as specific baits in the co-immunoprecipitated experiments. The heat-map indicates that NCL and NPM1 that are both involved in ribosome biogenesis share common ribosomal proteins and ribosome biogenesis factors like DDX21, BRIX1, NOP2, MYBBP1A, RBM28, RBM34, NOP56 and NOP58 with a high mean $\log_2$ intensity.

The two baits, NCL and NPM1 share most of the identified preys, especially ribosomal proteins, which are highly enriched in both baits and are indicated by orange-to-red bars. As expected, the bait NCL itself is highly enriched in the NCL-IP sample. However, NCL is also enriched in the NPM1 and YBX1 sample and unexpected high enriched in the empty vector control sample. This is caused by a carry-over effect during the measurements. Since the order of the replicates of each sample group were randomly mixed and then loaded onto the column, small sample fractions were 'carried over' to the next sample. Except a few preys, like some ribosome biogenesis factors, ribosomal proteins and the RNA helicase DDX21, which were identified in the YBX1 and the empty vector sample due to the carry-over effect, the control samples in general, showed no interaction with the identified preys. This is indicated by the grey bars in the heat map and illustrated in Figure 58.

In more detail, Table 32 in the appendix summarizes the identification and quantification of the NCL co-immunoprecipitated proteins according to protein name, gene name and UniProt prey identifier, spectral counts of the bait-prey pair and spectral control counts of the negative controls. The fold change (FC), which is calculated by the average spectral count in the test interaction (NCL) divided by the average counts in the control samples (empty and YBX1). Zero counts in the negative controls were replaced by 0.1 (Teo et al., 2014). The BFDR estimates all probability thresholds that are calculated directly from the probabilities (Choi et al., 2011; Teo et al., 2014). The binding partners are grouped according to ribosome biogenesis factors, RNA-helicases, subunits of the FACT complex, histones and others. The RPs are listed in Table 33 in the appendix. Only significant interactors of NCL, which are defined by a BFDR of $\leq$ 0.05 are summarized and further analyzed in the following tables.

Nucleolin binds to several factors involved in early and late processing steps of ribosome biogenesis, RNA-helicases, histones and other proteins. Mass-spec data reveal that NCL interacts with the nucleolar protein FTSJ3, which is involved in the

early cleavage steps and NIP7, a transacting factor that plays a role in the biogenesis of the 40S ribosomal subunit. It is assumed that both factors share functional interactions in pre rRNA processing (Morello et al., 2011a). Additional co-purified interaction partners of NCL are FBL, NOP56 and NOP58 that represent the core components of the snoRNP complex of the C/D box snoRNAs (Dupuis-Sandoval et al., 2015; Esteller, 2011). NCL is also co-purified with DDX21 and NOP2, two components of the UTP-B complex (Sloan et al., 2014) (Table 32).

Other proteins, which are not associated with ribosome biogenesis reflect only 14% of all identified proteins, counting histones to be involved in rDNA transcription. Histones are basic proteins that interact through their positive charges with DNA in the nucleus and facilitate the condensation of the chromatin. The nucleosome consists of DNA that is wrapped around eight histone proteins (histone octamer) containing a histone tetramer $(H3-H4)_2$ and two histone dimers (H1A-H2B). For DNA replication, DNA repair and transcription machineries that access the DNA, nucleosomes are disassembled (Burgess and Zhang, 2013). Mass-spectrometry analysis of NCL identified three histones (H1FX, HIST1H1D, and HIST1H1E) of the H1 histone family that function as a linker histone. Also, several RNA-helicase, which are explained in more detail in section 1.2.3.2, are co-purified with NCL. RNA-helicases function in processing, assembly and unwinding the RNA duplex structures and are associated with rRNA transcription, and with the small and large subunit biogenesis (Leary and Huang, 2001). DDX21 is associated with 28S rRNA stabilization and is involved in the 18S and 28S rRNA production, while DDX50 is an antagonist of DDX21 and facilitates maturation of 18S and 28S rRNA (Martin et al., 2013).

3.9.2 Nucleolin interacts mainly with ribosomal proteins of the large subunit

Out of 80 ribosomal proteins in total, NCL binds to 39 RPs; 36 RPs are involved in the maturation of the large subunit (RPL), whereas NCL binds only 3 RPs that belong to the small ribosomal subunit (RPS). These proteins are summarized according to protein name, gene name, UniProt prey identifier, spectral counts of the bait-prey pair, spectral control counts of the negative controls, FC, BFDR and the SAINT probability score in Table 33 in the appendix.

3.9.3 RNA-binding proteins identified by Mass-Spectrometry analysis of Nucleolin

In mass-spec analysis of NCL, out of the 119 co-purified proteins in total, 92 were identified as RNA-binding proteins and are grouped according to Figure 7 (Gerstberger et al., 2014b) to their RNA targets (Figure 59). The majority of the identified RBPs represent ribosomal proteins (41.5%), followed by rRNA- including pre rRNA-binding proteins (38.3%), mRBPs (8.5%), snoRBPs (4.3%), unknown RBPs (3.2%), snRBPs (2.1%), ncRBPs (1.1%) and diverse RBPs (1.1%). Since NCL interacts with snoRNA and pre rRNA, and RBM28 binds to mRNAs and snRNA, these two RBPs are listed in both categories in Table 34.
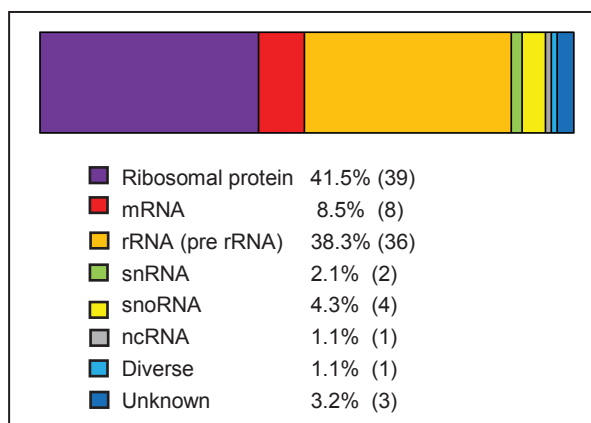
Figure 59: RNA-binding proteins identified by Mass-Spectrometry of Nucleolin
Out of the 119 co-purified proteins identified by mass-spectrometry of NCL, 91 RNA-binding proteins were identified. The majority represents ribosomal proteins (41.5%), followed by rRNA- including pre rRNA-binding proteins (38.3%), mRBPs (8.5%), snoRBPs (4.3%), unknown RBPs (3.2%), snRBPs (2.2%), ncRBPs (1.1%) and diverse RBPs (1.1%). Nucleolin interacts with snoRNA and pre rRNA, RBM28 binds to mRNAs as well as to snRNA, these two RBPs were therefore, listed in both RNA categories.

Table 34 groups the NCL co-purified RBPs according to their class of RNA-binding proteins. If the RBP is associated with a disease, it is also summarized in Table 34 in the appendix.

3.9.4 Specific binding partners of Nucleolin and Nucleophosmin

Furthermore, NCL and NPM1 are involved in ribosome biogenesis and share common binding partners, consequently a substantial overlap in the co-purified proteins was identified. Therefore, the question remained, which of the identified preys bind more specific to NCL and NPM1. Here, the mass-spectrometric intensity data were analyzed using the SAM method. For this analysis, the intensity data from 113 proteins with at least three intensity values in each group were used. SAM is a statistical method established by Tusher et al. in order to detect significant changes in differential gene expression between sets of samples (Tusher et al., 2001). The SAM analysis was calculated using the siggenes package by H. Schwender (Schwender, 2012) with the R environment version 3.1.1 (R Core Team, 2014). The package was modified so that missing values were not imputed. SAM analysis was calculated using a S_0 of 0.1 and significant different proteins were accepted with a FDR of 1%. Log2 intensity values of LFQ information were used for the SAM analysis to stabilize the variance. Consequently, 28 proteins in total were considered as significantly different between the two groups, 22 proteins showed a stronger association with NCL, whereas six proteins were identified in NPM1 that are significantly different to NCL (Figure 60 and Table 35 in the appendix).
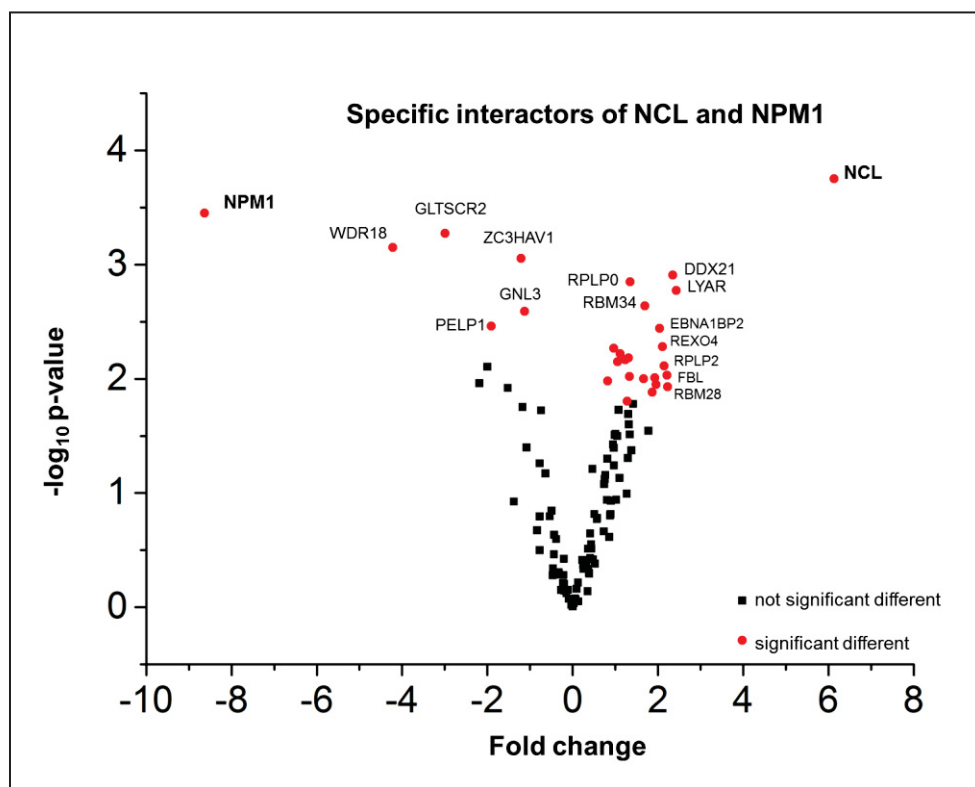
Figure 60: Specific interactors of Nucleolin versus Nucleophosmin identified by Mass-Spectrometry

The volcano plot illustrates specific binding partners of NCL and NPM1. The given fold change represents the distance of the mean log_{10} intensities of NCL and NPM1 co-purified samples. Positive fold changes represent proteins that are co-purified with NCL, while negative fold change refer to proteins that are co-purified with NPM1. SAM was used to determine significantly differentially co-enriched proteins with either NCL or NPM1 as bait proteins with a false discovery rate (FDR) of 1% and S_0 of 0.1.

Table 35 (appendix) lists the identified interacting proteins of NCL and NPM1 with a FDR of 1% in functional groups, including ribosome biogenesis factors (15), ribosomal proteins (5), RNA-helicases (3), histones (2) and others (3). A ratio of NCL to NPM1 between 0-1 indicates specific interactors of NPM1, whereas a ratio above one refers to specific co-purified proteins of NCL. Additionally, preys with negative d-values were specifically co-purified by NCL, whereas preys identified with a positive d-value bind more specific to NPM1. Nucleolin binds specifically to the following ribosome biogenesis factors including DDX21, EBNA1BP2, REXO4, BRIX1, RBM28, FBL, RBM34, RSL1D1 and DDX50, whereas NPM1 interacts with GLTSCR2, GNL3 and binds to two factors of the SENP3-associated protein complex composed of WDR18, PELP1 and TEX10 (Finkbeiner et al., 2011a; Finkbeiner et al., 2011b).

The following SAM plot illustrates 113 preys, which were identified by mass-spec of NCL and NPM1 and are depicted by black and green dots (Figure 61). The plot shows observed d-values on the y-axis versus the expected d-values plotted on the x-axis. The solid black line represents the 'observed-expected line', if preys are distributed along this line, the preys are not significant. Therefore, the wider the preys deviate from the line, the more likely they are significant, vice versa, the closer the preys are located at the line, the more likely they are not significant. The distance between the solid and the dotted line is Δ (Δ= 1.797). The green dots represent the 28 significant preys that were co-purified either with NCL or NPM1. These preys are also summarized in the appendix in Table 35. Six proteins were identified to be significantly associated with NPM1, whereas 22 significant interaction partners were identified for NCL. A cutlow of the observed d-value $\leq$ -2.55 indicates proteins that are co-purified with NCL, while a cutup of the observed d-value $\geq$ 3.53 represents preys that were co-purified with NPM1. All preys with a FDR of 1% were considered and are illustrated in Figure 61.
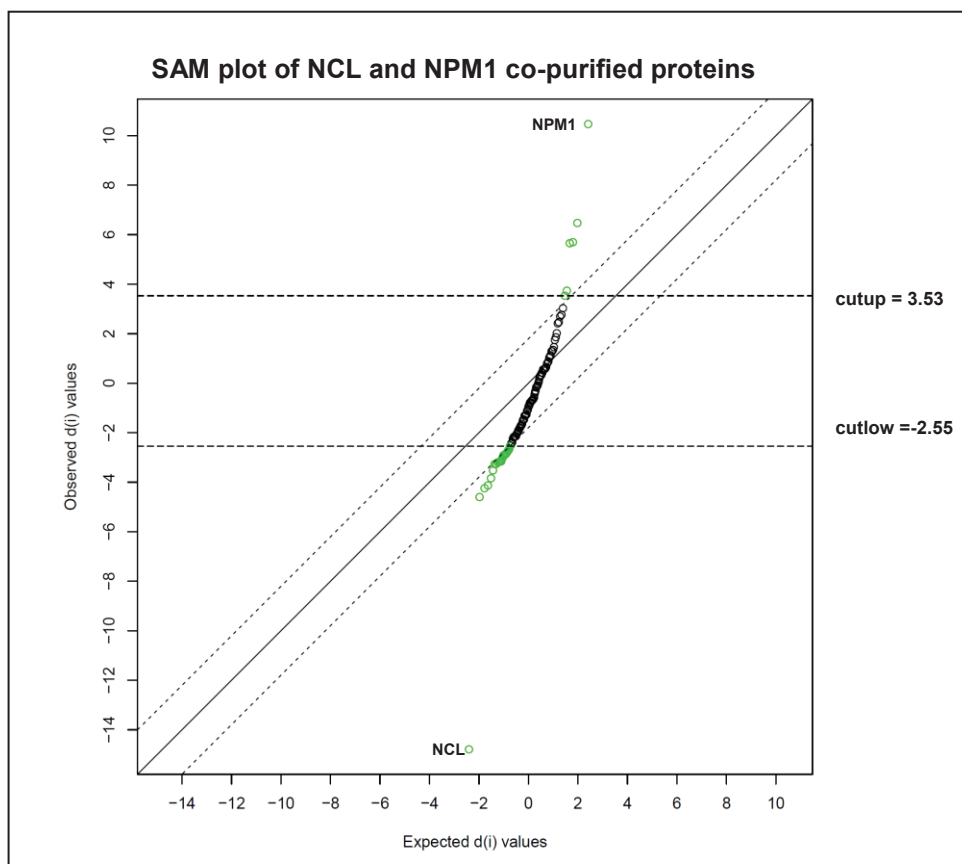
Figure 61: SAM plot of co-purified proteins by Nucleolin and Nucleophosmin

A SAM plot for delta (Δ) = 1.797 is generated. On the x-axis of the expected d(i)-values are plotted, on the y-axis the observed d(i)-values. The SAM plot shows 113 preys in total. Each dot (black or green) represents a prey of either NCL or NPM1. In total 28 significant preys (green) were identified by SAM for NCL and NPM1. Six proteins are co-purified by NPM1, whereas 22 significant interaction partners were identified for NCL. A cutlow value $\leq$ -2.55 indicates proteins that are associated and co-purified with NCL, while a cutup value $\geq$ 3.53 represents preys, which are associated with NPM1. All preys with a FDR of 1% were considered. The wider a prey that is represented by a dot deviates from the 'observed-expected line' (solid black line), the more likely the prey is significant. The distance between the two dotted lines is 2Δ, therefore, the distance from the 'observed-expected' line to the dotted line is Δ=1.797.

3.10 SiRNA mediated knockdown of Nucleolin

To further define the role of NCL in ribosome biogenesis and the effect of a NCL knockdown (KD) on pre rRNA levels in HEK293 cells, RNA interference (RNAi) of the NCL mRNA was employed followed by RNA sequencing. In order to optimize the KD efficiency of NCL, three different siRNA sequences (NCL I-III, 2.1.7.2 siRNAs) against NCL were tested and transfected in HEK293 cells, either separately or in a pool of

three for a time period of 24 h up to 96 h. NCL downregulation on protein level was analyzed by western blot.

3.10.1 Depletion of Nucleolin in HEK293 cells has an impact on the expression of precursor rRNA

Western blot analysis of the three tested siRNAs revealed that siRNA I and a transfection period of 72 h showed the most efficient KD of NCL on the protein level. Almost a complete reduction of NCL within the KD sample is observed by a specific anti-NCL antibody, compared to a untransfected mock control, which is only treated with the transfection reagent. β-tubulin was used as an internal control (Figure 62).

Figure 62: Western blot analysis of Nucleolin depleted HEK293 cells by RNAi

Western blot analysis of Nucleolin depleted cells showed reduced amounts of NCL in comparison to a mock control. Protein expression was normalized to an internal β-tubulin control. HEK293 cells were transfected with siRNA I against NCL for 72 h. The mock control was only transfected with Lipofectamine 2000. A specific anti-NCL antibody (anti-C23) with a corresponding HRP conjugated secondary antibody was used to detect endogenous expression of Nucleolin.

Furthermore, to study the impact of the NCL KD on the cellular transcriptome, RNA was isolated from NCL depleted HEK293 cells and submitted for RNA sequencing. Poly (A) seq was obtained to determine the knockdown efficiency of NCL and the expression patterns of other ribosome biogenesis factors, whereas total RNA-Seq data was used to analyze the effect of NCL on rRNA and pre rRNA.

Figure 63 illustrates the RNA-Seq analysis of the NCL RNAi experiment. Expression of NCL is shown in the poly (A) RNA-Seq data set in the mock control versus the NCL depleted cells, whereas the total RNA-Seq data show the expression levels for mature rRNA and pre rRNA. NCL depleted cells are marked in black bars, while the mock

control is depicted in grey. The expression levels of the mock control sample was set to 100%. According to the expression of the RNA, NCL was down regulated approximately 90%. The expression levels of mature rRNA remains stable, whereas the pre rRNA levels are slightly decreased by 15% (Figure 63).
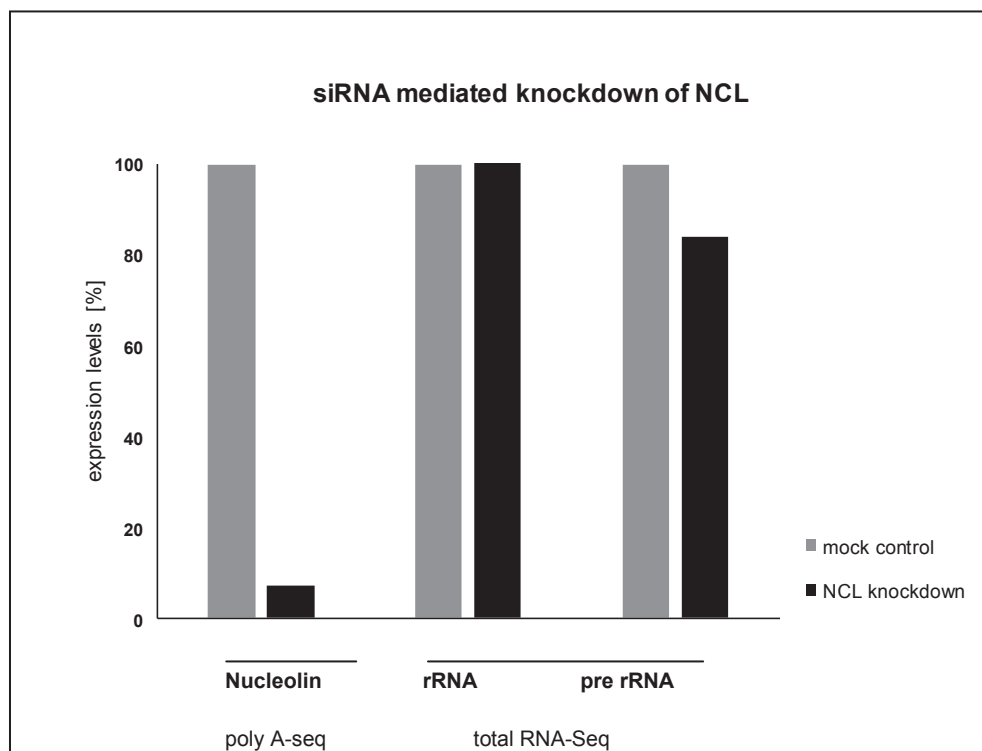
Figure 63: RNA-Sequencing of Nucleolin depleted HEK293 cells by RNAi

RNA sequencing of Nucleolin depleted cells, revealed a 90% decrease in NCL mRNA levels, stable expression of mature rRNA and a slight decrease in pre rRNA expression of about 15% in the NCL knockdown sample (I, black bars) in comparison to the mock control (untransfected) sample (I, grey bars).

To visualize the impact of NCL on the primary rRNA transcript in more detail, read coverage of the mock control is compared to the reads of the NCL KD sample (Figure 64). The figure shows log2 read coverage of the 47S rRNA derived from the NCL KD sample marked in red versus a mock control depicted in blue. The first cleavage site (01) in the 5'ETS is shown in green. The NCL KD sample shows slightly decreased expression levels within the precursor region, in the 5'ETS, ITS1 and ITS2. The diagram below shows the fold change difference of the mock control versus the NCL

depleted cells along the 47S rRNA depicted in blue. A FC above 1 is observed within the precursor region, whereas no significant changes were detected in the 18S and in the 5.8S rRNA, that are indicated by the grey boxes in Figure 64.
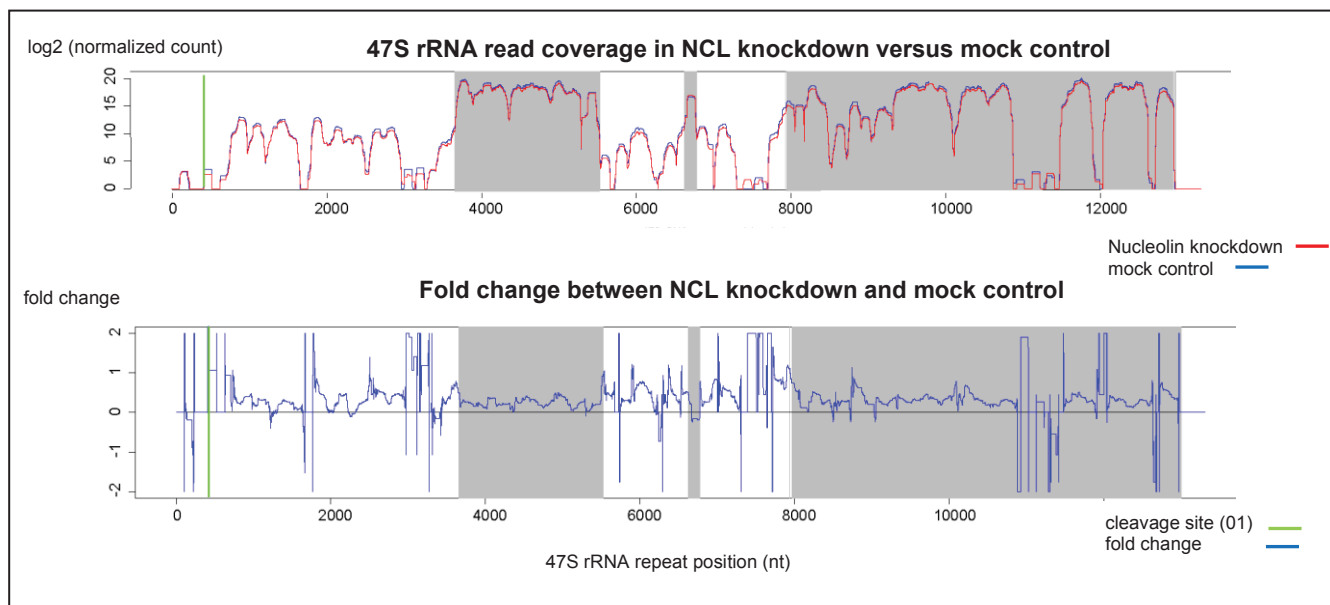
Figure 64: Read coverage of the 47S rRNA in Nucleolin depleted cells versus a mock control

The log2 coverage of reads derived from the Nucleolin knockdown sample (red) versus a mock control (blue) are illustrated on the top. The first cleavage site (01) within the 5'ETS is illustrated in green. Below, the fold change is depicted in a blue line and shows the difference of the NCL KD versus an untransfected control sample. A small decrease of the read count in the precursor rRNA region, including the 5'ETS, ITS1 and ITS2 is observed in the KD samples, whereas the expression of the mature region within the 18S rRNA and 5.8S rRNA remains stable.

In summary, KD of NCL has a slight impact on the primary rRNA transcript within the 5'ETS, ITS and ITS2, while the expression pattern of the mature rRNA remained stable. This is in accordance with the identified cellular RNA targets of NCL that were identified by RIP-Seq and PAR-CLIP. NCL interacts with the precursor rRNA and not with the mature rRNA. However, additional RNAi analysis of NCL in order to determine the impact of NCL on the pre rRNA transcription needs to be further elucidated.

4 DISCUSSION

4.1 Nucleolar localization of Nucleolin

In general, the localization of a protein provides information of the cellular functions, determines functional biological networks and defines potential interaction partners of the protein. Since altered protein localization facilitates pathogenesis of many human diseases including cancer, metabolic, cardiovascular and neurodegenerative diseases (Hung and Link, 2011). Therefore, the knowledge about the subcellular localization of a protein within a cell is highly important. Consequently, several bioinformatic prediction tools integrate the information of sequence motifs, protein domains as well as targeting sequences of protein–protein interactions in order to compute confidence scores, which predict the localization of proteins (Binder et al., 2014; Scott et al., 2010). However, it remains crucial and indispensable to test subcellular protein localization by techniques like immunohistochemistry.

Specific protein motif sequences like the NLS facilitate traffic across the nuclear membrane and mediate the localization of proteins within the nucleus. These mechanisms are well characterized, while precise mechanisms and pathways, how proteins find their target sites and accumulate in membrane-bound organelles like the nucleolus remain unknown (D. O'Day and A. Catalano, 2013, Scott et al., 2010). It is assumed that nucleolar accumulation of proteins depends on the interaction with nucleolar core components like rDNA, RNA and major protein complexes involved in ribosome biogenesis (Carmo-Fonseca et al., 2000). Although, according to Leung and Lamond the function of proteins in ribosome biogenesis alone is not sufficient for specific nucleolar localization (Leung and Lamond, 2003).

Recently, bioinformatic data analysis characterized a motif, the so called nucleolar localization signal (NoLS), a short targeting sequence like the NLS that might play a role in the accumulation of proteins within the nucleolus (Scott et al., 2010; Scott et al.,

2011). NoLS show related amino acid compositions like the NLS, which mostly contains basic residues. Leung et al. proposed that amino acids like arginine (R), lysine (K) and leucine (L) are highly enriched in proteins within the nucleolar compartments. K and R residues were found in 48% of the identifed NoLS sequences. RGG-repeats are increased up to three-fold in proteins of the human nucleolar proteome, whereas tetrapeptides like FGGR and RGGF represent a 10- and 20- fold enrichment, respectively (Leung et al., 2003). Data analysis of the human nucleolar proteome mainly identified nucleolar proteins that carry nucleic acid binding domains. A RNA-binding domain that is able to bind rRNAs or snoRNAs within the nucleolus (Carmo-Fonseca et al., 2000) was observed in 40 out of 51 nucleolar proteins (Leung et al., 2003). However, a specific target motif, present in all nucleolar proteins was not identified yet (Andersen et al., 2002; Leung et al., 2003). Within this study, a NoLS was not predicted for NCL by NoD (Figure 21), a program that analyzes and computes a confidence score to predict the presence of a NoLS within the protein (Scott et al., 2010; Scott et al., 2011). The precise mechanisms how NCL is localized within the nucleolus remain unknown.

Several studies by Creancier et al., and Schmidt-Zachmann and Nigg characterized the protein domains of NCL, which are responsible for nucleolar localization. Full length NCL and ΔN NCL that lacks the acidic stretch of the N-terminal domain are mainly present in the nucleolus. NCL variants that either lack the C-terminal domain (ΔC), which contains nine RGG domains or the RRMs (ΔRRM) that are responsible for RNA-binding, are primarily localized within the nucleoplasm (Creancier et al., 1993; Schmidt-Zachmann and Nigg, 1993). These studies support the data analysis of the nucleolar proteome (Leung et al., 2003; Leung and Lamond, 2003). Immunohistochemistry, performed within this thesis, confirms nucleolar localization of NCL and ΔN NCL (Figure 22 and Figure 23). However, a specific nucleolar signal within NCL was not identified. NCL contains several D/E repeats in the N-terminal domain, nine RGG

repeats (Lapeyre et al., 1987) and four of each of the described tetrapeptides (FGGR and RGGF) within the C terminus. It was proposed that the RGG domain of NCL binds non-specific and with low affinity to RNA (Bouvet et al., 1998; Ghisolfi-Nieto et al., 1996; Ghisolfi et al., 1992) and also interacts with RPs (Bouvet et al., 1998). The N terminus, especially the bipartite nuclear localization signal (KRKK-KKQKV) is interspersed with several K-residues, which is also characteristic for NoLS. Regarding the nucleolar localization of NCL, it was suggested that NoLS are not required for proteins that are immobilized in the nuclear matrix like NCL (Dickinson and Kohwi-Shigematsu, 1995). Consequently, a conventional NLS in combination with the described characteristic protein domains is likely sufficient for nucleolar localization of NCL. Therefore, it can be supposed that the localization of NCL is not caused by a specific sequence like the NoLS, but rather by the protein domains combined with their interactions targets that contribute to nucleolar localization. To conclude, accumulation of NCL within the nucleolus depends on the C-terminal and the central RBD, which mediates interaction with pre rRNA and snoRNA targets, that were identified by PAR-CLIP (section 3.5.2) and RIP-Seq analysis (section 3.5.5).

It was also proposed that nucleolar proteins like NPM1 are common factors for nucleolar targeting of specific proteins (Li et al., 1996b), therefore, association of NCL with NPM1 might also be responsible for nucleolar accumulation of NCL (Li et al., 1996b). Consequently, protein-protein interactions might play a crucial role in the nucleolar localization. On proteome-wide level, mass-spectrometry analysis revealed that NCL interacts significantly with several nucleolar proteins and ribosome biogenesis factors like FBL, DDX21, NOP56, NOP58, COIL, MYBBP1 and MKI67 (Table 32), which could also contribute to the nucleolar localization of NCL. In general, mechanisms and pathways that are involved in the nucleolar accumulation of proteins need to be further elucidated in the near future to understand the link between cellular functions of nucleolar proteins and the localization in a normal and a disease state.

4.1.1 Nucleolin accumulates in the nucleoplasm upon Actinomycin-D treatment

Using immunohistochemistry within this thesis, it was shown that NCL relocalizes to the nucleoplasm in the presence of Actinomycin-D, a small molecule that intercalates into G/C rich DNA duplexes and inhibits Pol I transcription (Hein et al., 2013). Treatment of Act-D leads to disruption of the nucleoli, cell cycle arrest, induction of apoptosis and results in an increase of p53. High levels of p53 are associated with nucleolar stress response and loss of integrity of the nucleolus, which is a hallmark of Pol I transcription inhibition (Colis et al., 2014). It was reported that Act-D also affects localization patterns of several other nucleolar proteins like NPM1 and FBL, these two nucleolar proteins translocate to the nucleoplasm upon Act-D treatment (Brodska et al., 2015; Chan, 1992; Chen and Jiang, 2004). Immunohistochemistry shows relocalization of NCL to the nucleoplasm in the presence of low Act-D concentration (0.4 µg/ml) and a small decrease in the size of the nucleolus (Figure 24 and Figure 25). This is in agreement with previously reported studies showing that Act-D causes a reduction in rRNA production, which leads to a loss of nucleolar integrity and reorganization of the nucleolus along with a decrease in nucleolar size (Brodska et al., 2015). However, Andersen et al. identified only small changes in the nucleolar proteome upon Act-D treatment (5 µg/mL). The nucleolar proteome data showed similar patterns in Act-D treated samples versus untreated control samples. Although, in the presence of Act-D increased abundance of certain proteins like hnRNP proteins including hnRNPs K, G, and A2/B1, DEAD box proteins including DDX9, p72, and p68, and RNA-binding proteins like PSF, PSP2/CoAA, PSP1, and p54/nrb was shown (Andersen et al., 2002).

To conclude, Act-D activates a nucleolar stress response in HEK293 cells and contributes to the relocalization of NCL to the nucleoplasm. Therefore, inhibition of Pol I along with a decrease in rRNA synthesis might play a role in NCL localization.

4.1.2 Nucleolar localization of Nucleolin is not affected by 4-thiouridine

To determine the cellular RNA targets of NCL within this thesis, PAR-CLIP analysis was performed. In PAR-CLIP experiments, cells are incubated up to 14 h with 100 µM 4SU, a photoactivatable nucleoside, which is incorporated into newly transcribed RNA and leads to T to C conversions in the corresponding cDNA. The conversion defines the crosslinking and binding site of the RNA. Previously, it was demonstrated by Burger et al. that high concentration of 4SU (> 50 µM) can prevent 47S rRNA production and can inhibit the processing of up to 75%. It was shown that 50 µM 4SU for 6 h is sufficient to damage mature 28S rRNA, whereas, long-term exposure of 100 µM 4SU, as used in PAR-CLIP, results in reduced cell proliferation. According to Burger et al., it is suggested that increased amounts of 4SU contribute to a nucleolar stress response (Burger et al., 2013). Nucleolar stress response is characterized by the loss of nucleolar integrity, high levels of p53 and relocalization of NCL. Inhibition of Pol I transcription and early rRNA processing induces nucleolar relocalization of NPM1 and disruption of the nucleolus, while inhibition of late pre rRNA processing events showed no effect on the localization of NPM1 (Burger et al., 2010). Rubbi and Milner reported that accumulation of p53 in UV or drug damaged cells occurs only if nuclear functions are affected (Rubbi and Milner, 2003). Consequently, nucleolar stress response along with nucleolar disruption is associated with increased levels of p53 (Burger et al., 2013; Rubbi and Milner, 2003). Previous studies demonstrate that localization of NCL depends on p53 activation due to stress (Daniely et al., 2002).

Regarding these results, the localization of NCL and p53 levels were tested in the presence of various 4SU concentrations (50-400 µM). However, in the presence of 4SU, NCL remains localized within the nucleolus (Figure 26). Also, the expression of p53 remains stable in comparison to an internal NCL-FLAG control (Figure 27). In summary, relocalization of NCL along with nucleolar disruption and a nucleolar stress response upon 4SU treatment was not observed within this study as previously described for NPM1 by Burger et al. (Burger et al., 2013). According to the results obtained within this study, 4SU can be used in the concentration as specified in the PAR-CLIP protocol. Although, 4SU controls are highly recommended and should be tested in every initial PAR-CLIP experiment, especially when investigating RBPs that are associated with ribosome biogenesis. Consistent with Burger et al., it is crucial to consider functional ribosome biogenesis, nucleolar integrity and the cell cycle in 4SU labeling experiments (Burger et al., 2013).

4.2 Phosphorylation of Nucleolin by Casein Kinase II needs to be considered in PAR-CLIP experiments

To identify the cellular RNA targets of NCL, PAR-CLIP experiments were performed in HEK293 cells stably expressing FLAG/HA tagged NCL. Preliminary PAR-CLIP control experiments of NCL in the presence and absence of 4SU and UV-crosslinking were tested. Unexpected, the results showed a crosslinking signal in the NCL-IP sample in the absence of 4SU, UV-crosslinking and even without T4 PNK (Figure 28 and Figure 29). Usually, a crosslinking band at the size of the RBP is only expected in samples that were previously incubated with 4SU, crosslinked with UV at 365 nm and labeled with γ ^{32}P-ATP using T4 PNK. Further control experiments using ATP as well as GTP for the RNA labeling reaction (Figure 30) and specific kinase inhibitors (Figure 32), subsequently, identified CKII to be responsible for the phosphorylation reaction of NCL

that was also confirmed by co-purification in western blot analysis (Figure 31). It was shown that CKII is associated with NCL (Pfaff and Anderer, 1988; Schneider and Issinger, 1988) and is recognized through the very basic, catalytic α-subunits of CKII. The catalytic subunits show high affinity binding to the acidic N-terminal domain of NCL. This is in agreement with the western blot analysis that shows co-immunoprecipitation of the CKII α-subunit by NCL (Figure 31). It was suggested that NCL is mainly phosphorylated in actively growing cells, and phosphorylated NCL plays a major role in rRNA transcription, whereas in confluent cells NCL is not phosphorylated and rRNA synthesis is almost completely down regulated. This is also consistent with the localization and the expression pattern of CKII. The kinase is highly abundant in the nucleolus of rapidly growing cells, while CKII is decreased in confluent cells. Therefore, phosphorylation of NCL by CKII might have a crucial impact on the regulation of rRNA transcription (Li et al., 1996a; Medina et al., 2010). It was also assumed that the RRMs of NCL bind to rRNA, while the N terminus of NCL binds to the Pol I machinery and blocks transcription elongation in the presence of a protease inhibitor, leupeptine (Bouche et al., 1984). Phosphorylation by CKII might therefore be necessary for the proteolysis of NCL in order to release the transcription complex (Bouche et al., 1984; Cong et al., 2012; Ginisty et al., 1999; Storck et al., 2007). Nevertheless, additional experiments are required to confirm the function of NCL in Pol I transcription (Storck et al., 2007) and the interplay between phosphorylation and rRNA transcription.

Furthermore, it was also shown that NCL does not interact with ATP (Figure 30) as previously reported and published by Srivastava and Pollard (Srivastava and Pollard, 1999). To conclude, the crosslinking signal observed in the NCL PAR-CLIP sample using γ ^{32}P-ATP did not completely result from crosslinked RNA, but rather from phosphorylated residues of NCL by CKII. Additionally, interaction of NCL with RNA was shown by labeling either the 5' or the 3' end of the crosslinked RNA using ATP or pCp,

respectively (Figure 34). In general, negative controls are crucial and need to be considered in every PAR-CLIP experiment, especially for RBPs that either possess a kinase activity or interact with a protein kinase, like NCL. Consequently, the intensity of the phosphorylation signal is not a good indicator for crosslinked RNA quantity.

4.3 The role of Nucleolin in ribosome biogenesis

4.3.1 Nucleolin interacts with RNA targets that are directly associated with ribosome biogenesis

Nucleolin contains four RBDs, consequently, it comprises RNA-binding properties and interacts with RNA, which was previously validated using ATP and pCp in order to label either the 5' or the 3' end of the crosslinked and NCL bound RNA (Figure 34). Further PAR-CLIP (section 3.5.2) and RIP-Seq (section 3.5.5) analyses were performed in HEK293 cells that stably express FLAG/HA tagged NCL, in order to characterize and identify transcriptome-wide RNA targets. PAR-CLIP analysis of NCL was performed for the first time within this thesis, to our knowledge no NCL PAR-CLIP data set is published so far. PAR-CLIP as well as RIP-Seq analyses of NCL identified pre rRNA and snoRNAs as targets of NCL, whereas mRNAs were not identified as a specific target. However, it was previously published and reviewed by Abdelmohsen and Gorospe that NCL stabilizes several mRNAs targets and interacts with the 3' and 5' UTR of the mRNA (Abdelmohsen and Gorospe, 2012). Abdelmohsen et al. identified several mRNA targets of NCL by IP and subsequent microarray analysis, and validated these targets by RT-PCR, however, no CLIP experiment on a transcriptome-wide level of NCL was performed within this reported study in order to identified precise crosslinking targets and binding sites of NCL. In the NCL PAR-CLIP experiments within this thesis, a mean ratio of 0.7 in pre rRNA and 0.8 in snoRNAs was observed, implicating 70% and 80% of T to C conversions in all reads containing one mismatch

[d1(T→C)/d1]. As expected, the empty vector control shows very low T to C conversions in all reads with one mismatch, implicating only 30% and 10% of crosslinking sites in pre rRNA (d1(T→C)/d1=0.3) and snoRNAs (d1(T→C)/d1=0.1), respectively (Table 19). Previous studies published by Ghisolfi-Nieto et al. revealed that NCL binds only to nascent pre rRNA and is absent from mature cytoplasmic ribosomes (Ghisolfi-Nieto et al., 1996). These studies are consistent with the performed PAR-CLIP within this thesis, in mature rRNAs a mean ratio of 0.15 of T to C conversion in all reads containing one mismatch were identified for NCL, which indicates only 15% crosslinking sites. Since the empty vector, which serves as a negative control shows a similar ratio of 0.2 in rRNAs (Table 19), mature rRNAs were not determined as a true RNA target of NCL. The NCL PAR-CLIP results are in agreement with the RIP-Seq analysis. In RIP-Seq analysis, pre rRNA and snoRNA showed a three- to four- fold enrichment in the NCL IP sample versus the lysate control, while mRNA and mature rRNA were not enriched and the expression levels remained stable (Figure 38). In summary, pre rRNA and snoRNAs, two cellular RNA targets that are directly involved in ribosome biogenesis were identified as targets for NCL by two independent methods like RIP-Seq and PAR-CLIP.

To further characterize the binding properties of the N-terminal domain of NCL, PAR-CLIP was performed on full length NCL versus ΔN NCL. The N terminus of NCL is rich in aspartic and glutamic acids (Figure 10), these domains were characterized as acid blobs and are common in transcriptional activating proteins (Sigler, 1988). Previous studies by Ginisty et al. proposed that the N-terminal domain is required for the interaction with the U3 snoRNP. Northern blots revealed that U3 interacts specifically with NCL, whereas no binding was observed in NCL lacking the N terminus (Ginisty et al., 1998). In PAR-CLIP analysis, ΔN NCL showed reduced T to C conversions and therefore less crosslinking sites in pre rRNA and snoRNAs in contrast to full length

NCL. In the ΔN NCL variant only 40% of all reads containing one mismatch contained T to C conversions in pre rRNA and snoRNA in comparison to an average ratio of 70% in pre rRNA and 80% in snoRNAs in full length NCL (Table 20). Consequently, it can be considered that the N-terminal domain of NCL, without any obvious RBDs, plays a role in the recognition of snoRNAs and pre rRNA. Previous studies by Ginisty et al. support this hypothesis (Ginisty et al., 1998).

On the other hand, Abdelmohsen and Gorospe showed that NCL binds to mRNA targets, regulates mRNA stability and controls translation. They identified BCL2, IL2, TP53, HBB, MMP9 and GADD45A as mRNA targets of NCL. Interaction with BCL2 increases stability and expression of the proto-oncogene and inhibits apoptosis in cancer cells (Abdelmohsen and Gorospe, 2012; Pichiorri et al., 2013). It was demonstrated that NCL also binds to TP53 within the 5'UTR and prevents translation of p53, resulting in DNA damage (Abdelmohsen et al., 2011). Usually, these mRNA targets contain AU-rich elements (AREs), which are present in approximately 5-8% of all mRNAs (Zhang et al., 2008). Furthermore, it was described that NCL binds to G-rich sequences located in the coding region and in the 5'UTR of mRNAs, which mostly encode proteins that are involved in cancer. These studies are in contradiction to the results obtained within this thesis. According to the PAR-CLIP and RIP-Seq analyses, mRNAs were not identified as targets of NCL. In PAR-CLIP, ~40% of all reads in mRNAs carrying one mismatch represent T to C conversions, while HuR a mRBP contains up to 90% T to C conversions within all reads containing one mismatch. In RIP-Seq analysis mRNAs were not enriched and showed relatively stable expression in the IP sample in comparison to the lysate control (Figure 38).

Previous SELEX experiments of NCL performed by Ghisolfi-Nieto et al. identified two different binding motifs, the 18 nt long NCL recognition element (NRE) that shares a common pre rRNA target sequence of UCCCG and a ECM motif with the following

recognition sequence of UCGA. The NRE motif was found in over 50% of the selected sequences (Ghisolfi-Nieto et al., 1996). In PAR-CLIP analysis within this thesis, crosslinked pre rRNA targets that contained the described NRE motif (oligo 1) or the ECM (oligo 3 and 4) were also identified and further tested in gel-shift assays (Table 22 and Figure 43). From an experimental perspective, PAR-CLIP identifies transcriptome-wide binding sites *in vivo* at a very high resolution in contrast to SELEX. SELEX is a low-throughput *in vitro* method that identifies targets of RBPs and uses a randomly generated large oligonucleotide library for the identification of the RNA targets (Ellington and Szostak, 1990).

The RNA sequence containing the NRE motif was further tested in electrophoretic mobility shift assays for binding (Figure 43 and Figure 44). The NRE motif is present in the 5'ETS as well as in the 18S rRNA and 28S rRNA, however, it was demonstrated that NCL binds only to the motif within the 5'ETS. Since not all NRE sequences that are present in the entire 47S rRNA were recognized by NCL, the environment and the structure might be crucial for high affinity binding (Ghisolfi-Nieto et al., 1996). According to Ghisolfi-Nieto et al., NCL interacts transiently with pre rRNA sequences in close vicinity to the first cleavage site in the 5'ETS and plays a role in the first steps of pre rRNA processing (Ghisolfi-Nieto et al., 1996; Ugrinova et al., 2007). NCL dissociates from nascent pre rRNA as soon as the pre ribosomal particles are formed, therefore, NCL is absent from the mature cytoplasmic ribosomes (Ghisolfi-Nieto et al., 1996). To conclude, RNA-sequencing (RNA-Seq) based approaches like PAR-CLIP and RIP-Seq of NCL identified pre rRNA and snoRNAs as targets of NCL. However, NCL shows no binding and crosslinking sites in mature rRNA and mRNA.

4.3.2 Nucleolin interacts with G/C-rich target sequences within the precursor rRNA

The NRE motif, which was previously identified in SELEX by Ghisolfi-Nieto et al. was characterized as a sequence-specific target of NCL (Ghisolfi-Nieto et al., 1996). The motif was tested in biochemical and structural binding studies that confirmed the interaction with NCL. In contrast, a random unselected RNA sequence containing 36% G/C residues, that is not present in the 47S rRNA showed no binding to NCL. According to Ghisolfi-Nieto et al., mutations of any G or C residue within the NRE pre rRNA-binding motif, or any point mutation within the stem that disrupts the hairpin structure, prevents interaction or reduces binding affinity of the complex (Ghisolfi-Nieto et al., 1996). Consequently, NCL was characterized as a sequence-specific RNA-binding protein (Ghisolfi-Nieto et al., 1996) that interacts with high affinity to the NRE motif.

Within this thesis, electrophoretic mobility shift assays were performed to determine, if NCL binds sequence-specific to the RNA targets that were previously identified by PAR-CLIP. A G to C substitution within the NRE motif of an identified RNA sequence did not abolish binding of NCL (Figure 44). Deletion analysis of the sequence that contains twice the NRE motif (oligo 1), showed that binding affinity is reduced in the absence of the first motif (Figure 45). Therefore, it is assumed that the first motif in contrast to the second motif might play a role in recognition and binding. Although, no interaction was observed using a 16 nt long RNA sequence containing both NRE motifs, therefore a minimum sequence length of 18 nt is required for binding. According to these results, point mutations within the motif did not prevent interaction, whereas reduced sequence lengths of the RNA completely inhibits binding, even in the presence of both well-known NRE motifs, which was previously described as a sequence-specific recognition motif of NCL (Ghisolfi-Nieto et al., 1996). However, it

seems that NCL does not bind to a sequence specific target, but rather binds to sequences along the precursor region of the rRNA that are highly enriched for C and G residues.

Recently, it was demonstrated that the RRMs and the C terminus of NCL recognizes G-G paired DNA structures as a natural binding target (Abdelmohsen et al., 2011; Gonzalez et al., 2009; Hanakahi et al., 1999). Studies by Abdelmohsen et al. using computational tools like the entire UniGene database identified a G-rich NCL binding motif, which is abundant in the coding region, 3'UTR and 5'UTR of mRNAs (Abdelmohsen et al., 2011). Within this thesis, a G-rich binding motif (oligo 15), which was identified by DREME (Figure 42) is present in ITS1, ITS2 and also in the 28S rRNA. Interaction of the G-rich sequence with NCL was confirmed by EMSAs (Figure 46). Additional biochemical and structural binding assays including ITC and SEC support the hypothesis that NCL does not bind to a specific target sequence but preferentially interacts with G/C-rich and G-repeat sequences. These results are in accordance with the nucleotide composition of the 47S pre rRNA transcript, which contains up to 70% G and C residues (Table 21). Whereas, according to Zhang et al., NCL recognizes A/U rich elements of Bcl-X_L within the 3'UTR of the mRNA (Zhang et al., 2008), these results were not observed within this thesis by gel-shift analysis of oligo 21 (poly A) and oligo 23 (ARE) (Figure 48). Within this study, mRNAs were not identified as a target of NCL by PAR-CLIP and RIP-Seq analysis. However, ITC measurements of NCL RRM1-4 showed high binding affinity with a K_d of 118 nM to an AC-repeat sequence (oligo 31). Since NCL RRM1-2 and the AC-repeat sequence showed a high K_d of 4,500 nM, implicating low binding affinity, RRM3-4 might be responsible for the recognition and interaction (Figure 54, Table 29). So far, no specific recognition properties were described for RRM3-4 as previously published for RRM1-2 and RRM1-4 of NCL. Biochemical and structural NMR studies (Ghisolfi-Nieto et al.,

1996) revealed that the first two RRMs (RRM1-2) of NCL are responsible for the recognition of the NRE motif (Finger et al., 2003; Serin et al., 1997), whereas the ECM motif requires all four RRMs (RRM1-4) (Ginisty et al., 2001) for binding.

In general, RRM1-4 showed a lower K_d and consequently a much higher binding affinity than RRM1-2 (Table 29). This suggests that RRM3-4 of NCL plays a role in sequence recognition and interaction. According to the performed ITC measurements using the RNA sequence with the NRE binding motif (oligo 1), it is supposed that the NCL-RNA complex shows a two-site binding model (Figure 52). This is indicated by two different binding reactions, representing a two-to-one binding ratio of protein to RNA. The same effect, which was not reported until now, was also observed in the corresponding gel-shift assays (Figure 43) and in SEC measurements (Figure 50 and Figure 51). According to the fact that the NRE motif binds to the first two RRMs (RRM1-2) (Finger et al., 2003; Serin et al., 1997), each NRE motif within oligo 1 binds to the first two RRMs, requiring two recombinant NCL proteins. Consequently, interaction of the RNA with the protein results in a two-to-one binding ratio.

To further characterize the relation between structure and function of NCL within the RNA-complex, crystallography was carried out. Allain et al. proposed that interaction of NCL with the sequence upstream of the first cleavage site in the 5'ETS, during rRNA transcription, requires chaperone activity of NCL, which facilitates proper and correct folding of the pre rRNA (Allain et al., 2000a; Allain et al., 2000b). RRM1-2 and RRM-4 of NCL with the previously identified pre rRNA target containing twice the NRE motif was tested in crystallography. A crystal structure of the complex was not dissolved (Figure 55), this could be due to the described two-site binding model that was observed in previous ITC measurements (Figure 52). The two-site model implicates two binding reactions, comprising an endothermic and an exothermic reaction. This is also in accordance with the stoichiometry of the complex, which requires two

recombinant proteins for one pre rRNA sequence and was also observed in SEC analysis (Figure 50 and Figure 51). Consequently, the complex seems to be disordered and not suitable for structure determination by crystallography.

To conclude, the performed deletion and mutation analysis of the reported binding motif (NRE), tested in gel-shift assays did not completely abolish the interaction with NCL. Biochemical binding studies within this thesis propose that Nucleolin is not a sequence-specific RNA-binding protein as previously reported (Ghisolfi-Nieto et al., 1996), but preferentially binds to G/C-rich and G-repeat sequences like the G-rich DREME motif that is present in the primary precursor rRNA transcript. Another binding motif was not identified within this thesis. However, no interaction was observed in A/U-rich repeats and poly (A) sequences. Although, RRM1-4 interacts with high binding affinity to an AC-repeat sequence, whereas RRM1-2 shows low binding affinity to the same RNA sequence, this implicating that RRM3-4 could be responsible for recognition. Binding properties of the RRM3-4 of NCL need to be further tested.

4.3.3 Nucleolin interacts with ribosome biogenesis factors that are involved in early and late processing steps of ribosome biogenesis

Mass-spectrometry analysis of Nucleolin identified 47 ribosome biogenesis factors, 39 ribosomal proteins of the small and the large subunit, 11 RNA-helicases, two subunits of the FACT complex, 3 histones, which activate rDNA transcription and only 17 other factors that are not involved in ribosome biogenesis (Figure 57). It was previously shown that NCL induces chromatin decondensation and interacts with histone H1 through the acidic stretch in the N-terminal domain, which shows a bipolar pattern of a HMG-like domain (Erard et al., 1988). In accordance with this data, three histones, which belong to the H1 family were identified by MS analysis of NCL. Previously, it was reported that NCL possesses FACT-like properties and facilitates chromatin

remodeling and histone chaperone activity (Angelov et al., 2006; Mongelard and Bouvet, 2007). Mass-spec analysis showed co-immunoprecipitated of the two subunits of the FACT complex by NCL (Figure 57, Table 32). However, it has to be considered that interaction of NCL with the FACT complex could contribute to the regulation of Pol I transcription and chromatin accessibility and could facilitate the FACT-like properties, which were previously described for NCL. Ribosomal processing factors associated with NCL like FBL, HP1BP3, NAT10, NHP2L1, NKRF, NO66, NOP56, NOP58, NSA2, PPAN, RBM34, RRS1, TOP2A and RNA-helicases such as DDX10, DDX24, DDX31, DDX51, DHX33 function in early processing and in the first cleavages sites (01, A0 and 1) (Tafforeau et al., 2013). While additional ribosome biogenesis factors that were identified by mass-spec analysis of NCL, including BRIX1, BOP1, EBNA1BP2, FTSJ3, GTPBP4, MAK16, MKI67IP, MYBBP1A, NOL9, RRP1, RSL1D1, XRN2 and the RNA-helicases DDX18, DDX27 and DDX54 are implicated in early processing steps as well as in the synthesis of the large subunit. MRTO4 exclusively plays a role in the synthesis of the large ribosomal subunit, whereas NOC2L functions in early processing steps and in the synthesis of the small ribosomal subunit, as well as COIL. COIL is also implicated in the synthesis of the 40S (Tafforeau et al., 2013). It was recently published that COIL binds to NCL (Tafforeau et al., 2013), which was also identified by mass-spectrometry within this thesis. NCL is also co-purified with the nucleolar protein FTSJ3, which is involved in early cleavage steps and associated with NIP7, a transacting factor that plays a role in the biogenesis of the 40S ribosomal subunit (Morello et al., 2011b). FBL, NOP56 and NOP58, which represent the core proteins of the snoRNP complex of C/D box snoRNAs (Dupuis-Sandoval et al., 2015; Esteller, 2011) were also co-purified by NCL. Mass-spec data analysis revealed that NCL interacts with DDX21 and NOP2, two components of the UTP-B complex (Sloan et al., 2014) and MYBBP1A that is involved in the regulation of pre rRNA processing and is associated with the Pol I complex (Hochstatter et al., 2012). NCL interacts with a wide

range of proteins that are implicated in early and late processing steps of ribosome biogenesis, therefore, NCL does not only play a crucial role in the first processing steps, as reported so far, and is rather associated with the entire process of ribosome biogenesis within the nucleolus. Consequently, NCL could be also involved in late processing steps. It is also assumed that the N terminal domain of NCL interacts with the Pol I machinery and regulates Pol I elongation (Bouche et al., 1984; Cong et al., 2012; Ginisty et al., 1999; Storck et al., 2007). However, it is still not completely understood yet, how NCL regulates RNA Pol I transcription, therefore, further experiments are essential to verify the function of NCL in rRNA transcription (Storck et al., 2007; Cong, 2012).

To further characterize the identified interaction partners of NCL, the structural domains of the co-purified proteins were compared to NCL. NCL contains 59 DE/EE/DD/ED repeats in the acidic stretch of the N-terminal domain and a score of 41 (Table 31). The score is calculated by the amount of DE/EE/DD/ED repeats in relation to the distance of ≤10 amino acids between each motif, with a minimum of three repeats in repetition. A score of two is defined by at least three DE/EE/DD/ED repeats, spaced within 10 aa or fewer apart. The higher the score, the more repeats are in close vicinity. DE/EE/DD/ED repeats are very acidic, and acid stretches were described as acid blobs or negative noodles that are common in transcriptional activating proteins (Sigler, 1988). Therefore, the identified proteins could be involved in the activation of Pol I transcription. Ribosome biogenesis factors co-purified with NCL including BOP1, CEBPZ, NOL8, MAK16, FTSJ3, DDX10, RBM28, NOM1 and the subunits of the FACT complex, SUPT16H and SSRP1 contain several stretches of DE/EE/DD/ED repeats. The amount of the repeats with the corresponding score is summarized in Table 31.

Table 31: Number of DE/EE/DD/ED repeats within ribosome biogenesis
factors that interact with Nucleolin

Protein	Number of DE/EE/DD/ED repeats	Score
BOP1	24	7
CEBPZ	43	16
DDX10	32	9
FTSJ3	33	10
MAK16	15	9
NCL	59	41
NOL8	41	8
NOM1	25	5
RBM28	29	15
SSRP1	25	9
SUPT16H	35	11

These interaction partners contain numerous acidic stretches of D/E repeats and could therefore be involved in transcription activation of Pol I.

Previous mass-spec analysis of NCL published by Yanagida et al. identified 60 proteins, comprising 40 RPs, including 13 proteins involved in the formation of the small subunit (RPS), 27 of the large subunit (RPL), and 20 non-ribosomal proteins (Takahashi et al., 2003), including poly [ADP-ribose] polymerase, transcription repressor E4BP4, NS11 associated protein, poly (A)-binding protein, NPM1, YB-1, prohibitin, hnRNPs (heterogeneous nuclear ribonucleoproteins), eIFs (eukaryotic translation initiation factors) and RNA-helicases (Takahashi et al., 2003; Yanagida et al., 2001). The overlap between the two mass-spec data sets represents several ribosomal proteins of the small (RPS6, RPS9) and the large (RPL10, RPL10a, RPL11, RPL12, RPL13, RPL13a, RPL14, RPL15, RPL17, RPL18, RPL18a, RPL21, RPL23a, RPL27, RPL27a, RPL30, RPL35, RPL36, RPLP0) subunit. The relative small overlap could be explained by the experimental approach and the data analysis of each experiment. According to the proteomic approach performed by Yanagida et al., direct protein-protein interactions were tested in the absence of RNA. The IP samples were

digested with RNase A (1 µg/ml, 37°C, 10 min) (Yanagida et al., 2001). Additionally, the flow-rate is crucial for the amount of identified proteins, the slower the flow-rate, the higher the number of eluted peptides (Takahashi et al., 2003) and vice versa. Mass-spectrometry within this thesis was performed in a single-step affinity purification of exogenously overexpressed NCL without RNAse treatment. Therefore, it has to be considered that the protein-protein interactions could be RNA-dependent. In order to reduce the background and contaminants, a more stringed two-step affinity purification and enrichment procedure using a dual affinity tag like the FLAG/HA tag can be performed. Although, weak protein-protein interactions might be lost during a more stringent two-step purification procedure. Stringent statistical data analysis and strict filtering steps ensure that true binding partners can be distinguished from contaminants and allow a single purification step, which is milder and more appropriate for transient interactors (Keilhauer et al., 2015). However, within a next step, the interaction of the identified ribosome biogenesis factors with NCL needs to be verified by co-immunoprecipitation and western blot analysis.

In summary, mass-spectrometry analyses revealed that NCL interacts with ribosomal proteins of the small, but mainly with proteins of the large subunit, it binds to ribosome biogenesis factors that are localized within the nucleolus and are involved in early and late processing events and cleavage steps. Consequently, the identified proteome wide targets of NCL, directly associate NCL with the highly coordinated process of ribosome biogenesis.

4.3.4 Knockdown of Nucleolin regulates ribosome biogenesis factors

To further determine the role of NCL in ribosome biogenesis, RNAi of NCL followed by RNA sequencing was performed in HEK293 cells. The effect of the KD was analyzed according to the expression of pre rRNA and ribosome biogenesis factors. RNAi

experiments of NCL performed within this thesis were analyzed in western blots and RNA-Seq. The RNA-Seq data showed almost 90% downregulation of NCL, stable expression of rRNA and a decrease of 15% in pre rRNA (Figure 63). To further characterize the minor down regulation of pre rRNA, read coverage of the 47S rRNA was analyzed and small fold changes were observed within the 5'ETS, ITS1 and ITS2 (Figure 64). It was shown that NCL plays a crucial role in ribosome biogenesis and it was also suggested that NCL regulates Pol I transcription, however no remarkable effect of the 47S rRNA expression was observed in the NCL KD by RNAi (Figure 64).

Recently, Durut and Saez-Vasquez reviewed a dual role for NCL involved in rDNA transcription (Durut and Saez-Vasquez, 2015). On the one hand, several groups showed that depletion of NCL mediated by RNAi leads to cell cycle arrest, nucleolar disruption, increase of apoptosis and inhibition of Pol I transcription along with decreased levels of pre rRNA (Abdelmohsen et al., 2011; Berger et al., 2015; Tajrishi et al., 2011; Ugrinova et al., 2007). Conditional KO of NCL in chicken DT40 cells showed decreased levels of NCL gene expression along with inhibition of RNA Pol I transcription and reduced accumulation of pre rRNA (Storck et al., 2009). Another model, described by Cong et al. suggests that depletion of NCL leads to accumulation of Pol I at the beginning of the transcription unit, at the 5'end of the rDNA and to reduced levels of pre rRNA. This is accompanied with decrease levels of UBTF along the entire rDNA. According to Cong et al., KD of NCL increases the interaction of TTF1 (transcription termination factor 1) with T0 (promoter-proximal terminator), which results in rRNA silencing. This leads to the recruitment of the nucleolar remodeling complex, induces a heterochromatin state and consequently inhibits Pol I transcription (Cong et al., 2012). Within this thesis, TTF1 was identified as a co-purified protein by NCL using mass-spectrometry analysis (Table 32).

In ChIP-seq, enrichment of NCL was observed in the coding as well as in the promoter region of rDNA, and showed equal distribution of UBTF and RPA116, the second

largest subunit of Pol I. While siRNA mediated KD of NCL leads to a ~80% decrease in rRNA. Although, Cong et al. proposed that depletion of NCL has an impact on rRNA production without affecting processing of the precursor rRNA. Therefore, Cong et al. claimed that reduced levels of pre rRNA does not result from inhibition of rRNA processing and maturation, but rather suggests that NCL plays a crucial and a direct role in Pol I transcription. Although, it has to be pointed out that Cong et al. normalized the relative 45S processing rate and the processing rate of the 32S rRNA to the 45S rRNA in the control sample and in the siRNA mediated NCL KD cells (Cong et al., 2012). To suggest, an independent Pol I or Pol III transcript would serve as a better control for normalization.

In summary, several studies showed that a decrease of RNA Pol I transcription correlates with NCL depletion, while overexpression of NCL is associated with increased levels of pre rRNA and rDNA transcription (Berger et al., 2015; Cong et al., 2012; Durut and Saez-Vasquez, 2015; Rickards et al., 2007; Storck et al., 2009). On the other hand, Roger et al. demonstrated that interaction of NCL with nascent pre rRNA inhibits Pol I transcription, whereas Pol II and III transcription were not affected. Injection of a two- to three- fold excess of NCL into *Xenopus laevis* stage IV oocytes showed reduced accumulation of pre rRNA, whereas injection of anti-NCL antibody into the nucleus of *C. tentans* salivary glands lead to a two- to almost four- fold increase in the transcription rate of rDNA (Durut and Saez-Vasquez, 2015; Egyhazi et al., 1988; Roger et al., 2002).

5 SUMMARY AND OUTLOOK

Nucleolin is a highly abundant RNA-binding protein, that is localized in all three compartments of the nucleolus, even though it does not contain a common nucleolar localization signal (NoLS). The precise mechanisms, how NCL localizes within the nucleolus are still not completely understood yet. Results within this thesis demonstrate that localization of NCL is independent on the N-terminal domain. Therefore, it is assumed that the RRM domains, which are required for RNA-interaction with snoRNAs and pre rRNA, contribute to the localization of NCL within the nucleolus. Treatment of 4SU does not induce a nucleolar stress response, which would be indicated by increased levels of p53. Consequently, 4SU has no impact on the localization of NCL, and NCL remains localized within the nucleolus in the presence of the photoactivatable ribonucleoside, whereas in the presence of Act-D, which inhibits Pol I transcription, NCL relocalizes to the nucleoplasm.

The three compartments of the nucleolus are associated with different functions in the process of ribosome biogenesis. According to the fact that NCL is distributed in the FC, DFC and GC of the nucleolus, interacts with pre rRNA and snoRNAs, targets that are directly associated with ribosome biogenesis and binds to several ribosome biogenesis factors, it can be concluded that NCL plays an essential role in the processing steps of ribosome biogenesis within the nucleolus. Figure 65 illustrates sketchily the localization and the interaction partners on proteome- and transcriptome- wide level that are associated with NCL and are linked to the highly coordinated process of ribosome biogenesis.
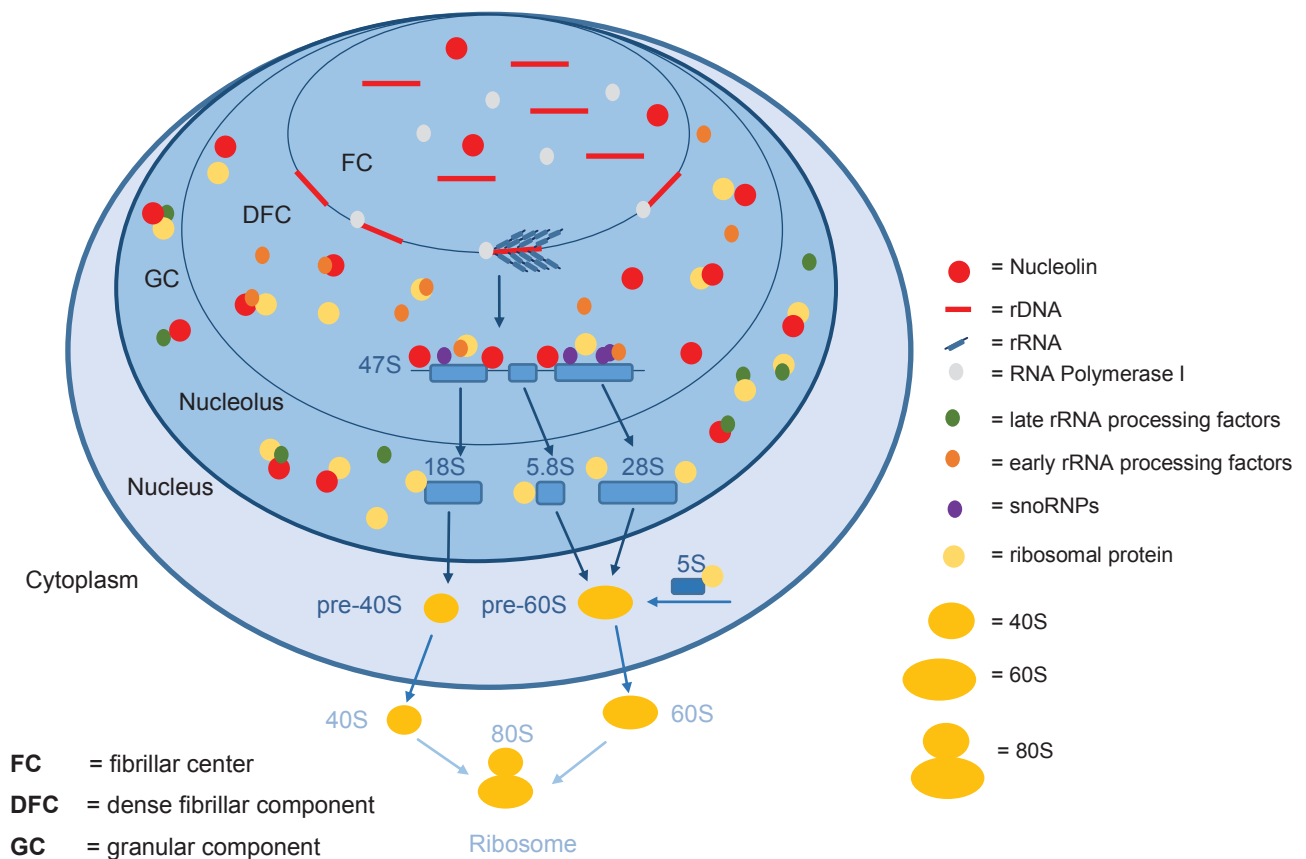
Figure 65: Localization and interactions partners of Nucleolin associated with ribosome biogenesis

Nucleolin is present in all three compartments of the nucleolus. It is mainly located in the dense fibrillar center (DFC) and less abundant in the fibrillar center (FC) and in the granular center (GC). On transcriptome-wide level, NCL binds to snoRNAs and to the precursor region of the 47S rRNA within the 5'ETS, ITS1 and ITS2. On proteome-level, NCL interacts with several ribosome biogenesis factors and ribosomal proteins of the small and the large subunit.

The thesis mainly focused on the RNA-binding properties and the protein-protein interactions of NCL. To elucidate the cellular RNA targets of NCL, PAR-CLIP and RIP-Seq analysis were performed. Pre rRNA and snoRNAs, two targets that are directly associated with ribosome biogenesis were identified as targets of NCL. To our knowledge, this thesis provides the first PAR-CLIP data sets of NCL. Biochemical and structural binding studies including EMSA, SEC and ITC showed that NCL does not bind sequence-specific to a certain target motif like the NRE, but reveals binding preferences to G and C rich sequences. This is in agreement with the nucleotide composition of the primary rRNA transcript that contains 72% G/C residues.

On proteome level, mass-spectrometry analysis of NCL identified 119 proteins, such as ribosome biogenesis factors including RBPs, ribosomal proteins, RNA-helicases and histones. RNAi experiments were conducted to elucidate the impact of NCL on ribosome biogenesis and the expression pattern of the precursor rRNA transcript. SiRNA mediated KD of NCL showed a slight decrease of 15% of precursor rRNA, stable expression of mature rRNA and a complete down regulation up to 90% of NCL on protein, as well as on RNA level. The expression levels of several ribosome biogenesis factors, especially XRN2 (Lu et al., 2010; Park et al., 2007) and DDX18 (Dubaele and Chene, 2007; Payne et al., 2011), which are known to be overexpressed in tumor cells were highly enriched in the NCL RIP-Seq sample along with increased levels of NCL, while the factors showed down regulation in NCL depleted cells along with decreased levels of NCL.

This study provides new information about the role of NCL in ribosome biogenesis, pre rRNA and snoRNAs were identified as targets on transcriptome-wide level, whereas ribosomal proteins and ribosome biogenesis factors including several RBPs were identified on proteome-wide level. Within a next step, the interaction of the identified ribosome biogenesis factors with NCL needs to be verified and additionally tested by co-immunoprecipitation and western blot analysis. Also, siRNA mediated KD of several other ribosome biogenesis factors along with the analysis of the expression levels of the pre rRNA transcript will provide information of their function and their role in the different processing steps of ribosome biogenesis. This could be analyzed by RNA-Seq or northern blot analysis. Additionally, several other RNA-binding proteins that are associated with ribosome biogenesis and were identified as interactors of NCL by mass-spec need to be further elucidated using methods like PAR-CLIP, RIP-Seq and RNAi to analyze and characterize their interactions partners and targets in order to reveal their function. Alternatively, proteomic based approaches like mass-spec will also contribute to the identification and characterization of new RBPs that are involved

in ribosome biogenesis. A combination of RNA-protein UV-crosslinking and a RNA pull-down assay using labeled RNA as a bait, followed by mass-spectrometry could be carried out to further determine the RNA-binding proteome.

Alterations in RNA-binding proteins, mutations in genes encoding ribosomal proteins or ribosome biogenesis factors, and defects in Pol I transcription are implicated in cancer and other genetic diseases. Consequently, it will be important in the near future to understand the molecular mechanisms and the process of Pol I transcription in order to find specific targets and to develop new therapeutic approaches to treat cancer and other genetic diseases that are associated with the highly coordinated process of ribosome biogenesis.

6 APPENDIX OF TABLES

Table 32: Significant interaction partners of Nucleolin identified by Mass-Spectrometry analysis

Protein	Gene	UniProt prey identifier	Spectral counts for the bait-prey pair	Spectral counts in the negative controls (Empty and YBX1)	FC	BFDR	SAINT probability score
Ribosome biogenesis factors							
Ribosome biogenesis protein BOP1	BOP1	Q14137	5|4|3|3|2	0|0|0|0|0|0|0|0|0	34.00	0.00	1.00
Ribosome biogenesis protein BRX1 homolog	BRIX1	Q8TDN6	34|10|33|16|13	0|0|0|0|1|0|0|0|2|9	7.07	0.00	0.95
CCAAT/enhancer-binding protein zeta	CEBPZ	Q03701	5|0|4|4|2	0|0|0|0|0|0|0|0|0	30.00	0.04	0.80
Coilin	COIL	P38432	12|3|8|11|1	0|0|0|0|0|0|0|0|0	70.00	0.03	0.80
Probable rRNA-processing protein EBP2	EBNA1BP2	Q99848	36|21|31|18|21	1|0|0|0|0|0|0|0|0	101.60	0.00	1.00
Eukaryotic translation initiation factor 6	EIF6	P56537	5|4|6|7|4	0|0|0|0|0|0|0|0|0	52.00	0.00	1.00
rRNA 2-O-methyltransferase fibrillarin	FBL	P22087	19|9|16|16|10	0|0|1|0|3|0|0|0|4|5	4.31	0.00	0.94
pre-rRNA processing protein FTSJ3	FTSJ3	Q8IY81	8|8|4|5|6	0|0|0|0|0|0|0|0|0	62.00	0.00	1.00
Guanine nucleotide-binding protein-like 3	GNL3	Q9BVP2	10|12|10|10|7	0|0|0|1|5|0|0|0|0	6.53	0.00	0.97
Glutamate-rich WD repeat-containing protein 1	GRWD1	Q9BQ67	7|4|11|3|3	0|0|0|0|0|0|0|0|0	56.00	0.00	1.00
Nucleolar GTP-binding protein 1	GTPBP4	Q9BZE4	12|15|13|10|9	1|0|0|0|2|0|0|0|0	15.73	0.00	1.00
Heterochromatin protein 1-binding protein 3	HP1BP3	Q5SSJ5	9|6|8|8|6	0|0|0|2|0|0|0|0|0	14.80	0.00	1.00
Pumilio domain-containing protein KIAA0020	KIAA0020	Q15397	25|6|19|14|8	0|0|0|0|0|0|0|0|0	144.00	0.00	1.00
Protein MAK16 homolog	MAK16	Q9BXY0	7|0|4|5|2	0|0|0|0|0|0|0|0|0	36.00	0.04	0.80
Antigen KI-67	MKI67	P46013	20|9|16|29|0	0|0|0|0|0|0|0|0|0	148.00	0.02	0.80
MKI67 FHA domain-interacting nucleolar phosphoprotein	MKI67IP	Q9BYG3	11|9|9|9|6	0|0|0|0|1|0|0|0|0	35.20	0.00	1.00
mRNA turnover protein 4 homolog	MRTO4	Q9UKD2	13|7|12|8|2	0|0|0|0|0|0|0|0|0	84.00	0.00	1.00

Myb-binding protein 1A	MYBBP1A	Q9BQG0	66|38|55|42|52	0|0|0|0|0|0|0|0|1	202.40	0.00	1.00
N-acetyltransferase 10	NAT10	Q9H0A0	4|5|11|12|3	0|0|0|0|0|0|0|0|0	70.00	0.00	1.00
Nucleolin	NCL	P19338	329|307|368|257|403	0|24|12|19|26|1|17|59|42|111	5.59	0.00	1.00
NHP2-like protein 1	NHP2L1	P55769	7|1|7|4|3	0|0|0|0|1|0|0|0|0	17.60	0.05	0.80
60S ribosome subunit biogenesis protein NIP7 homolog	NIP7	Q9Y221	16|4|12|10|7	0|0|0|0|0|0|0|0|0	98.00	0.00	1.00
NF-kappa-B-repressing factor	NKRF	O15226	13|14|10|15|16	3|1|1|3|4|0|0|0|0	4.95	0.00	1.00
Bifunctional lysine-specific demethylase and histidyl-hydroxylase NO66	NO66	Q9H6W3	5|10|4|7|5	0|0|0|0|0|0|0|0|0	62.00	0.00	1.00
Nucleolar protein 8	NOL8	Q76FK4	2|1|3|4|3	0|0|0|0|0|0|0|0|0	26.00	0.05	0.80
Polynucleotide 5-hydroxyl-kinase NOL9	NOL9	Q5SY16	3|3|2|3|2	0|0|0|0|0|0|0|0|0	26.00	0.00	0.99
Nucleolar MIF4G domain-containing protein 1	NOM1	Q5C9Z4	2|3|2|4|0	0|0|0|0|0|0|0|0|0	22.00	0.05	0.79
Putative ribosomal RNA methyltransferase NOP2	NOP2	P46087	64|39|71|62|70	2|0|0|2|1|0|0|2|0|9	16.32	0.00	1.00
Nucleolar protein 56	NOP56	O00567	46|20|52|49|31	0|1|0|2|6|0|0|0|1	15.84	0.00	1.00
Nucleolar protein 58	NOP58	Q9Y2X3	43|14|43|22|18	1|0|0|3|4|0|0|0|0	14.00	0.00	1.00
Ribosome biogenesis protein NSA2 homolog	NSA2	O95478	5|4|2|5|0	0|0|0|0|0|0|0|0|0	32.00	0.04	0.80
Suppressor of SWI4 1 homolog	PPAN	Q9NQ55	10|8|9|9|6	0|0|0|0|0|0|0|0|0	84.00	0.00	1.00
RNA-binding protein 28	RBM28	Q9NW13	72|37|61|56|71	1|0|0|0|0|1|0|0|1	79.20	0.00	1.00
RNA-binding protein 34	RBM34	P42696	39|24|43|31|34	1|2|1|0|2|0|1|2|1|3	15.20	0.00	1.00
RNA exonuclease 4	REXO4	Q9GZR2	19|7|19|13|14	0|0|0|0|0|0|0|0|0	144.00	0.00	1.00
Ribosome production factor 2 homolog	RPF2	Q9H7B2	6|2|6|4|4	0|0|0|0|0|0|0|0|0	44.00	0.00	1.00
RRP15-like protein	RRP15	Q9Y3B9	4|3|2|2|2	0|0|0|0|0|0|0|0|0	26.00	0.00	0.99
Ribosomal RNA processing protein 1 homolog B	RRP1B	Q14684	34|27|30|26|23	0|0|0|0|1|0|0|0|0	112.00	0.00	1.00
Ribosomal RNA-processing protein 8	RRP8	O43159	8|6|10|9|11	0|0|0|0|0|0|0|0|0	88.00	0.00	1.00
Ribosome biogenesis regulatory protein homolog	RRS1	Q15050	12|4|15|10|9	0|0|0|0|0|0|0|0|0	100.00	0.00	1.00

Ribosomal L1 domain-containing protein 1	RSL1D1	O76021	49|32|46|46|33	1|0|2|4|8|0|0|0|1|2	10.30	0.00	1.00
Surfeit locus protein 6	SURF6	O75683	3|6|6|8|6	0|0|0|0|0|0|0|0|0|0	58.00	0.00	1.00
Testis-expressed sequence 10 protein	TEX10	Q9NXF1	4|4|7|3|2	0|0|0|0|0|0|0|0|0|0	40.00	0.00	1.00
DNA topoisomerase 2-alpha	TOP2A	P11388	6|4|3|3|3	0|0|0|0|0|0|0|0|0|0	38.00	0.00	1.00
Transcription termination factor 1	TTF1	Q15361	2|2|1|5|4	0|0|0|0|0|0|0|0|0|0	28.00	0.05	0.79
Unhealthy ribosome biogenesis protein 2 homolog	URB2	Q14146	11|1|9|14|6	0|0|0|0|0|0|0|0|0|0	82.00	0.02	0.80
5-3 exoribonuclease 2	XRN2	Q9H0D6	13|9|13|12|9	0|0|0|0|4|0|0|0|0|0	11.20	0.00	0.99
RNA-helicases							
Probable ATP-dependent RNA helicase DDX10	DDX10	Q13206	5|4|4|6|3	0|0|0|0|0|0|0|0|0|0	44.00	0.00	1.00
ATP-dependent RNA helicase DDX18	DDX18	Q9NVP1	7|1|7|4|5	0|0|0|0|1|0|0|0|0|0	19.20	0.03	0.80
Nucleolar RNA helicase 2	DDX21	Q9NR30	118|110|116|119|101	3|0|0|5|10|0|1|3|2|11	15.56	0.00	1.00
ATP-dependent RNA helicase DDX24	DDX24	Q9GZR7	21|18|17|20|22	0|0|0|0|0|0|0|0|0|0	196.00	0.00	1.00
Probable ATP-dependent RNA helicase DDX27	DDX27	Q96GQ7	9|11|9|8|6	0|0|0|0|0|0|0|0|0|0	86.00	0.00	1.00
Probable ATP-dependent RNA helicase DDX31	DDX31	Q9H8H2	8|3|4|5|6	0|0|0|0|0|0|0|0|0|0	52.00	0.00	1.00
ATP-dependent RNA helicase DDX50	DDX50	Q9BQ39	20|21|22|17|14	0|0|0|0|1|0|0|0|0|0	75.20	0.00	1.00
ATP-dependent RNA helicase DDX51	DDX51	Q8N8A6	8|4|3|6|6	0|0|0|0|0|0|0|0|0|0	54.00	0.00	1.00
ATP-dependent RNA helicase DDX54	DDX54	Q8TDD1	18|19|18|22|22	0|0|0|0|0|0|0|0|0|1	79.20	0.00	1.00
Probable ATP-dependent RNA helicase DDX56	DDX56	Q9NY93	6|2|9|8|5	0|0|0|0|0|0|0|0|0|0	60.00	0.00	1.00
Putative ATP-dependent RNA helicase DHX33	DHX33	Q9H6R0	2|2|4|2|1	0|0|0|0|0|0|0|0|0|0	22.00	0.05	0.79
Histones							
Histone H1x	H1FX	Q92522	18|19|15|10|15	0|1|0|2|5|0|0|0|0|0	7.70	0.00	0.99
Histone H1.3	HIST1H1D	P16402	2|5|2|2|2	0|0|0|0|1|0|0|0|0|0	10.40	0.00	0.94
Histone H1.4	HIST1H1E	P10412	17|24|18|15|19	4|1|7|5|9|0|0|0|0|0	2.98	0.01	0.82

FACT complex							
FACT complex subunit SSRP1	SSRP1	Q08945	11|7|9|5|4	0|0|0|0|0|0|0|0|0	72.00	0.00	1.00
FACT complex subunit SPT16	SUPT16H	Q9Y5B9	8|3|8|6|3	0|0|1|0|0|0|0|0|0	22.40	0.00	0.99
others							
ATPase family AAA domain-containing protein 3A	ATAD3A	Q9NVI7	4|8|4|8|2	0|0|0|0|5|0|0|0|0	4.16	0.04	0.80
Uncharacterized protein C7orf50	C7orf50	Q9BRJ6	10|8|11|11|5	0|0|0|0|0|0|0|0|0	90.00	0.00	1.00
Coiled-coil domain-containing protein 137	CCDC137	Q6PK04	5|4|5|3|3	0|0|0|0|0|0|0|0|0	40.00	0.00	1.00
Centromere protein V	CENPV	Q7Z7K6	13|12|11|16|11	0|0|0|0|0|0|0|0|0	126.00	0.00	1.00
Putative oxidoreductase GLYR1	GLYR1	Q49A26	12|4|4|5|7	0|0|0|0|0|0|0|0|0	64.00	0.00	1.00
G patch domain-containing protein 4	GPATCH4	Q5T3I0	6|4|7|5|3	0|0|0|0|0|0|0|0|0	50.00	0.00	1.00
Lysine-rich nucleolar protein 1	KNOP1	Q1ED39	10|6|15|10|8	0|0|0|0|0|0|0|0|0	98.00	0.00	1.00
Cell growth-regulating nucleolar protein	LYAR	Q9NX58	31|18|38|22|18	0|0|1|0|0|0|0|0|0	101.60	0.00	1.00
Nuclear valosin-containing protein-like	NVL	O15381	11|21|13|6|7	0|0|0|0|0|0|0|0|0	116.00	0.00	1.00
Very-long-chain (3R)-3-hydroxyacyl-[acyl-carrier protein] dehydratase 3	PTPLAD1	Q9P035	3|5|4|3|0	0|0|0|0|1|0|0|0|0	12.00	0.05	0.79
Periodic tryptophan protein 1 homolog	PWP1	Q13610	23|12|24|11|16	2|0|1|1|0|0|0|0|2	11.47	0.00	1.00
RNA-binding protein 39	RBM39	Q14498	2|5|2|4|3	0|0|0|0|2|0|0|0|0	6.40	0.01	0.90
Serine/arginine repetitive matrix protein 1	SRRM1	Q8IYB3	6|14|6|6|1	0|0|0|1|1|0|0|0|0	13.20	0.03	0.80
Tubulin beta chain;Tubulin beta-2A chain	TUBB	P07437	20|5|16|23|12	0|0|2|1|7|0|0|0|0	6.08	0.01	0.84
Histone-lysine N-methyltransferase NSD2	WHSC1	O96028	6|4|3|9|4	0|0|0|0|0|0|0|0|0	52.00	0.00	1.00
Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	Q7Z2W4	6|5|6|6|4	0|0|0|0|3|0|0|0|0	7.20	0.00	0.96
Zinc finger protein 512	ZNF512	Q96ME7	5|3|4|5|4	0|0|0|0|0|0|0|0|0	42.00	0.00	1.00

Mass-spectrometry analysis of Nucleolin identified 47 ribosome biogenesis factors, 11 RNA-helicases, 3 histones, the two subunits of the FACT complex and 17 other proteins that are not involved in ribosome biogenesis. BFDR refer to the Bayesian false discovery rate, FC to fold change.

Table 33: Significant ribosomal proteins identified by Mass-Spectrometry analysis of Nucleolin

Protein	Gene	UniProt prey identifier	Spectral counts for the bait-prey pair	Spectral counts in the negative controls (Empty and YBX1)	FC	BFDR	SAINT probability score
Ribosomal proteins of the small subunit (40S)							
40S ribosomal protein S24	RPS24	P62847	16|8|7|8|5	1|1|1|2|3|0|0|0|0|0	5.03	0.00	0.94
40S ribosomal protein S6	RPS6	P62753	26|16|20|19|14	1|0|1|2|6|0|0|0|0|2	6.91	0.00	1.00
40S ribosomal protein S9	RPS9	P46781	13|5|10|15|13	2|1|2|3|6|0|0|0|0|0	3.45	0.05	0.79
Ribosomal proteins of the large subunit (60S)							
60S ribosomal protein L10	RPL10	P27635	29|27|20|12|11	0|0|0|0|3|0|0|0|0|0	26.40	0.00	1.00
60S ribosomal protein L10a	RPL10A	P62906	31|21|30|26|20	2|1|0|0|2|0|0|0|2|1	14.63	0.00	1.00
60S ribosomal protein L11	RPL11	P62913	14|13|12|7|6	1|0|2|0|1|0|0|0|0|1	8.32	0.00	0.99
60S ribosomal protein L12	RPL12	P30050	36|22|39|30|23	1|1|2|2|2|0|0|0|0|3	13.33	0.00	1.00
60S ribosomal protein L13	RPL13	P26373	23|22|21|21|20	2|0|0|1|4|0|0|0|0|0	12.23	0.00	1.00
60S ribosomal protein L13a	RPL13A	P40429	24|25|22|15|18	1|0|0|0|1|0|0|0|0|0	41.60	0.00	1.00
60S ribosomal protein L14	RPL14	P50914	26|26|27|23|23	1|1|1|1|2|1|0|0|0|6	10.00	0.00	1.00
60S ribosomal protein L15	RPL15	P61313	22|15|20|17|7	1|0|0|0|0|0|1|0|0|0	32.40	0.00	1.00
60S ribosomal protein L17	RPL17	P18621	9|11|10|11|12	1|0|0|1|1|0|0|0|0|0	14.13	0.00	1.00
60S ribosomal protein L18	RPL18	Q07020	36|37|40|46|29	0|2|2|4|3|0|2|3|10|10	5.57	0.00	1.00
60S ribosomal protein L18a	RPL18A	Q02543	31|34|31|31|22	2|1|1|0|2|0|0|0|0|1	19.87	0.00	1.00
60S ribosomal protein L19	RPL19	P84098	9|16|16|10|15	1|0|2|2|6|1|0|0|0|0	4.80	0.00	0.96
60S ribosomal protein L21	RPL21	P46778	26|20|18|17|13	0|0|1|0|1|0|0|0|0|3	15.04	0.00	1.00
60S ribosomal protein L23	RPL23	P62829	5|3|3|5|6	0|0|0|0|0|0|0|0|0|0	44	0.00	1.00
60S ribosomal protein L23a	RPL23A	P62750	22|16|20|21|29	1|0|3|2|8|0|0|0|0|5	4.80	0.00	1.00
60S ribosomal protein L24	RPL24	P83731	17|16|11|11|8	2|0|3|3|4|0|0|0|0|0	4.20	0.01	0.94
60S ribosomal protein L27	RPL27	P61353	21|13|21|18|10	1|0|0|0|2|0|0|0|0|0	22.13	0.00	1.00

60S ribosomal protein L27a	RPL27A	P46776	14|15|16|11|8	0|0|0|1|4|0|0|0|0|0	10.24	0.00	0.99
60S ribosomal protein L28	RPL28	P46779	19|15|10|14|16	3|0|0|0|0|0|0|0|0|0	19.73	0.00	1.00
60S ribosomal protein L29	RPL29	P47914	3|5|4|3|3	1|0|1|1|0|1|0|0|0|0	3.60	0.02	0.81
60S ribosomal protein L3	RPL3	P39023	50|40|37|30|26	3|0|1|2|5|0|0|1|0|0	13.31	0.00	1.00
60S ribosomal protein L30	RPL30	P62888	37|29|28|34|18	1|2|2|1|3|0|0|0|1|5	9.73	0.00	1.00
60S ribosomal protein L32	RPL32	P62910	18|12|11|9|6	0|0|0|0|0|0|0|0|0|0	112.00	0.00	1.00
60S ribosomal protein L34	RPL34	P49207	3|4|3|3|7	0|0|0|0|0|0|0|0|0|0	40.00	0.00	1.00
60S ribosomal protein L35	RPL35	P42766	8|7|5|3|7	1|0|1|1|1|0|0|0|0|0	6.00	0.01	0.94
60S ribosomal protein L35a	RPL35A	P18077	9|7|7|8|6	0|0|0|0|2|0|0|0|0|0	14.80	0.00	1.00
60S ribosomal protein L36	RPL36	Q9Y3U8	10|8|6|4|6	0|0|0|0|0|0|0|0|0|0	68.00	0.00	1.00
60S ribosomal protein L37a	RPL37A	P61513	11|8|11|9|8	0|0|0|1|1|0|1|0|4|2	4.70	0.00	0.99
60S ribosomal protein L4	RPL4	P36578	72|74|71|63|64	1|4|4|1|6|1|0|1|0|1	18.35	0.00	1.00
60S ribosomal protein L5	RPL5	P46777	20|17|17|11|12	0|0|0|0|2|0|0|0|0|3	12.32	0.00	1.00
60S ribosomal protein L6	RPL6	Q02878	35|35|37|29|33	1|0|1|2|2|0|0|0|0|1	22.53	0.00	1.00
60S ribosomal protein L7	RPL7	P18124	44|38|44|44|33	2|1|2|3|6|0|0|1|3|5	9.55	0.00	1.00
60S ribosomal protein L7a	RPL7A	P62424	47|43|48|43|34	4|2|3|4|4|0|0|1|2|5	10.12	0.00	1.00
60S ribosomal protein L8	RPL8	P62917	27|19|21|16|17	4|2|5|2|3|0|0|1|2|8	4.00	0.00	1.00
60S ribosomal protein L0	RPLP0	P05388	64|51|62|45|40	1|5|5|4|10|0|0|5|6|16	5.66	0.00	1.00
60S ribosomal protein L2	RPLP2	P05387	16|13|20|17|6	0|0|0|0|0|0|0|0|0|0	144.00	0.00	1.00

Mass-spectrometry analysis of Nucleolin identified 3 ribosomal proteins, which belong to the small ribosomal subunit and 36 ribosomal proteins of the large subunit. BFDR refer to the Bayesian false discovery rate, FC to fold change.

Table 34: Nucleolin co-purified RNA-binding proteins identified by Mass-Spectrometry

RNA-binding protein class	Gene name	Implicated in disease (Reference)
Ribosomal proteins	RPL10	Autism (Klauck, 2006; Klauck et al., 2006)
	RPL10A	Autism (Klauck, 2006; Klauck et al., 2006)
	RPL11	Diamond-Blackfan anemia (Gazda et al., 2008; Narla and Ebert, 2010)
	RPL12	Hepatocellular cancer (van Riggelen et al., 2010)
	RPL13	Gastrointestinal cancer (van Riggelen et al., 2010)
	RPL13A	Hepatocellular cancer (van Riggelen et al., 2010), Colorectal cancer (Shenoy et al., 2012)
	RPL14	
	RPL15	Oesophageal cancer (van Riggelen et al., 2010), Shwachman-Diamond syndrome (Narla and Ebert, 2010)
	RPL17	
	RPL18	Colorectal cancer (Shenoy et al., 2012)
	RPL18A	
	RPL19	Colorectal cancer (van Riggelen et al., 2010), prostate cancer (Shenoy et al., 2012)
	RPL21	
	RPL23	Shwachman-Diamond syndrome (Narla and Ebert, 2010)
	RPL23A	Hepatocellular cancer (van Riggelen et al., 2010)
	RPL24	
	RPL27	Hepatocellular cancer (van Riggelen et al., 2010)
	RPL27A	
	RPL28	Colorectal cancer (Shenoy et al., 2012)
	RPL29	Shwachman-Diamond syndrome (Narla and Ebert, 2010)
	RPL3	
	RPL30	Hepatocellular cancer (van Riggelen et al., 2010)
	RPL32	Colorectal cancer (Shenoy et al., 2012)
	RPL34	
	RPL35	Colorectal cancer (van Riggelen et al., 2010)
	RPL35A	Diamond-Blackfan anemia (Gazda et al., 2008; Narla and Ebert, 2010), Colorectal cancer (Shenoy et al., 2012)
	RPL36	Diamond-Blackfan anemia (Narla and Ebert, 2010), hepatocellular cancer (Shenoy et al., 2012)
	RPL37A	Prostate Cancer (van Riggelen et al., 2010)
	RPL4	
	RPL5	Colorectal cancer (van Riggelen et al., 2010), Diamond-Blackfan anemia (Gazda et al., 2008) , Colorectal cancer (Shenoy et al., 2012)
	RPL6	Shwachman-Diamond syndrome (Narla and Ebert, 2010)
	RPL7	Colorectal cancer (Shenoy et al., 2012)
	RPL7A	

	RPL8	
	RPLP0	
	RPLP2	
	RPS24	Diamond-Blackfan anemia (Gazda et al., 2006; Narla and Ebert, 2010)
	RPS6	Colorectal cancer (Shenoy et al., 2012; van Riggelen et al., 2010), Non-Hodgkin lymphoma (Shenoy et al., 2012)
	RPS9	Shwachman-Diamond syndrome (Shenoy et al., 2012)
mRNA-binding	EIF6	
	KIAA0020	
	MRTO4	
	PPAN	
	RBM28	Neurologic defects, endocrinopathy syndrome (Nousbeck et al., 2008)
	RBM34	
	RBM39	
	SRRM1	
rRNA and pre-rRNA-binding	BOP1	Overexpressed in cancer (Chung et al., 2011; Killian et al., 2006)
	BRIX1	
	CEBPZ	Down-regulated in chronic myeloid leukemia (Qian et al., 2005)
	DDX10	Mutated in cancer, fusion with NUP98 causes tumor formation (Arai et al., 1997)
	DDX18	Mutated in acute myeloid leukemia, overexpressed in tumor cells (Dubaele and Chene, 2007; Payne et al., 2011)
	DDX21	
	DDX24	
	DDX27	
	DDX31	
	DDX50	
	DDX51	
	DDX54	
	DDX56	
	EBNA1BP2	Overexpressed in carcinoma (Pilarsky et al., 2004)
	FTSJ3	
	GNL3	
	GRWD1	
	GTPBP4	Down-regulated in cancer (Lee et al., 2007)
	MAK16	
	NAT10	
	NIFK	
	NIP7	
	NOL8	

	NOL9	
	NOM1	
	NOP2	
	NSA2	
	REXO4	
	RPF2	
	RRP15	
	RRP1B	
	RRP8	
	RRS1	
	SURF6	
	URB2	
	NCL	Overexpressed in cancer (Abdelmohsen and Gorospe, 2012), associated with human breast cancer, lung cancer, gastric cancer and leukemia (Hammoudi et al., 2013; Otake et al., 2007; Pichiorri et al., 2013; Qiu et al., 2013; Shen et al., 2014; Soundararajan et al., 2009; Zhao et al., 2013)
snRNA-binding	NHP2L1	
	RBM28	Neurologic defects, endocrinopathy syndrome (Nousbeck et al., 2008)
snoRNA-binding	FBL	
	NCL	
	NOP56	
	NOP58	
ncRNA-binding	ZC3HAV1	
Diverse	XRN2	Overexpressed in cancer (Lu et al., 2010; Park et al., 2007)
Unknown	GPATCH4	
	NKRF	
	RSL1D1	

Table 35: Specific interactors of Nucleolin and Nucleophosmin identified by Mass-Spectrometry

Protein	Gene	Mean NCL Intensity (mean log2)	Mean NPM1 Intensity (mean log2)	Ratio of NCL/NPM1	d-value	p-value	q-value
Ribosome biogenesis factors							
Ribosome biogenesis protein BRX1 homolog	BRIX1	248,768 (18)	62,975 (16)	3.95	-2.75	0.011	0.013
Probable rRNA-processing protein EBP2	EBNA1BP2	457,417 (19)	9,4911 (17)	4.82	-3.52	0.004	0.010
rRNA 2-O-methyltransferase fibrillarin	FBL	168,971 (17)	39,749 (15)	4.25	-2.91	0.009	0.013
Glioma tumor suppressor candidate region gene 2 protein	GLTSCR2	6,078 (13)	43,093 (16)	0.14	6.47	0.001	0.006
Guanine nucleotide-binding protein-like 3	GNL3	37,436 (15)	83,118 (16)	0.45	3.74	0.003	0.008
Nucleolin	NCL	24,058,905 (25)	375,039 (19)	64.15	-14.79	0.0002	0.006
Nucleophosmin	NPM1	37,537 (14)	8,031,078 (23)	0.005	10.47	0.0004	0.006
Proline-, glutamic acid- and leucine-rich protein 1	PELP1	10,567 (13)	30,500 (15)	0.35	3.53	0.003	0.010
RNA-binding protein 28	RBM28	695,703 (20)	180,187 (17)	3.86	-2.70	0.012	0.013
RNA-binding protein 34	RBM34	536,418 (19)	140,565 (17)	3.82	-3.84	0.002	0.008
RNA exonuclease 4	REXO4	99,391 (17)	2,1076 (14)	4.72	-3.27	0.005	0.012
RRP15-like protein	RRP15	15,315 (14)	6,739 (13)	2.27	-3.20	0.006	0.013
Ribosomal RNA processing protein 1 homolog B	RRP1B	178,909 (18)	100,686 (17)	1.78	-3.26	0.005	0.012
Ribosomal L1 domain-containing protein 1	RSL1D1	417,786 (19)	144,604 (17)	2.89	-2.88	0.010	0.013
WD repeat-containing protein 18	WDR18	6,909 (12)	81,215 (17)	0.09	5.70	0.001	0.006
Ribosomal protein							
60S ribosomal protein L32	RPL32	347,326 (18)	142,183 (17)	2.44	-3.17	0.007	0.013
60S ribosomal protein L6	RPL6	1,577,157 (21)	807,605 (20)	1.95	-3.16	0.007	0.013

60S ribosomal protein L7	RPL7	1,513,703 (21)	528,342 (19)	2.87	-2.84	0.010	0.013
60S ribosomal protein L0	RPLP0	109,6207 (20)	451,983 (19)	2.43	-4.24	0.001	0.006
60S ribosomal protein L2	RPLP2	90,115 (16)	20,306 (14)	4.44	-3.03	0.008	0.013
RNA-helicases							
Nucleolar RNA helicase 2	DDX21	2,574,778 (21)	608,845 (19)	4.23	-4.60	0.001	0.006
ATP-dependent RNA helicase DDX50	DDX50	88,435 (16)	48,209 (16)	1.83	-2.80	0.010	0.013
ATP-dependent RNA helicase DDX24	DDX24	71,977 (16)	38,023 (15)	1.89	-2.55	0.016	0.016
Histones							
Histone H1x	H1FX	108,797 (17)	36,960 (15)	2.94	-2.68	0.013	0.014
Histone H1.4	HIST1H1E	964,141 (20)	441,546 (19)	2.18	-3.18	0.007	0.013
Others							
Cell growth-regulating nucleolar protein	LYAR	288,384 (18)	60900 (16)	4.74	-4.13	0.002	0.007
Periodic tryptophan protein 1 homolog	PWP1	203,234 (18)	80,632 (16)	2.52	-2.90	0.010	0.013
Zinc finger CCCH-type antiviral protein 1	ZC3HAV1	5,731 (13)	13,421 (14)	0.43	5.66	0.001	0.006

7 LIST OF REFERENCES

Abdelmohsen, K., and M. Gorospe. 2012. RNA-binding protein nucleolin in disease. *RNA biology*. 9:799-808.

Abdelmohsen, K., K. Tominaga, E.K. Lee, S. Srikantan, M.J. Kang, M.M. Kim, R. Selimyan, J.L. Martindale, X. Yang, F. Carrier, M. Zhan, K.G. Becker, and M. Gorospe. 2011. Enhanced translation by Nucleolin via G-rich elements in coding and non-coding regions of target mRNAs. *Nucleic acids research*. 39:8513-8530.

Ahmad, Y., F.M. Boisvert, P. Gregor, A. Cobley, and A.I. Lamond. 2009. NOPdb: Nucleolar Proteome Database--2008 update. *Nucleic acids research*. 37:D181-184.

Allain, F.H., P. Bouvet, T. Dieckmann, and J. Feigon. 2000a. Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *The EMBO journal*. 19:6870-6881.

Allain, F.H., D.E. Gilbert, P. Bouvet, and J. Feigon. 2000b. Solution structure of the two N-terminal RNA-binding domains of nucleolin and NMR study of the interaction with its RNA target. *Journal of molecular biology*. 303:227-241.

Andersen, J.S., Y.W. Lam, A.K. Leung, S.E. Ong, C.E. Lyon, A.I. Lamond, and M. Mann. 2005. Nucleolar proteome dynamics. *Nature*. 433:77-83.

Andersen, J.S., C.E. Lyon, A.H. Fox, A.K. Leung, Y.W. Lam, H. Steen, M. Mann, and A.I. Lamond. 2002. Directed proteomic analysis of the human nucleolus. *Current biology : CB*. 12:1-11.

Angelov, D., V.A. Bondarenko, S. Almagro, H. Menoni, F. Mongelard, F. Hans, F. Mietton, V.M. Studitsky, A. Hamiche, S. Dimitrov, and P. Bouvet. 2006. Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes. *The EMBO journal*. 25:1669-1679.

Anger, A.M., J.P. Armache, O. Berninghausen, M. Habeck, M. Subklewe, D.N. Wilson, and R. Beckmann. 2013. Structures of the human and Drosophila 80S ribosome. *Nature*. 497:80-85.

Arai, Y., F. Hosoda, H. Kobayashi, K. Arai, Y. Hayashi, N. Kamada, Y. Kaneko, and M. Ohki. 1997. The inv(11)(p15q22) chromosome translocation of de novo and therapy-related myeloid malignancies results in fusion of the nucleoporin gene, NUP98, with the putative RNA helicase gene, DDX10. *Blood*. 89:3936-3944.

Armistead, J., R. Hemming, N. Patel, and B. Triggs-Raine. 2014. Mutation of EMG1 causing Bowen-Conradi syndrome results in reduced cell proliferation rates concomitant with G2/M arrest and 18S rRNA processing delay. *BBA clinical*. 1:33-43.

Ascano, M., S. Gerstberger, and T. Tuschl. 2013. Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. *Current opinion in genetics & development*. 23:20-28.

Ascano, M., M. Hafner, P. Cekan, S. Gerstberger, and T. Tuschl. 2012. Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley interdisciplinary reviews. RNA*. 3:159-177.

Bailey, T.L. 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics (Oxford, England)*. 27:1653-1659.

Bartel, D.P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell*. 136:215-233.

Berger, C.M., X. Gaume, and P. Bouvet. 2015. The roles of nucleolin subcellular localization in cancer. *Biochimie*. 113:78-85.

Binder, J.X., S. Pletscher-Frankild, K. Tsafou, C. Stolte, S.I. O'Donoghue, R. Schneider, and L.J. Jensen. 2014. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database : the journal of biological databases and curation*. 2014:bau012.

Boisvert, F.M., Y. Ahmad, M. Gierlinski, F. Charriere, D. Lamont, M. Scott, G. Barton, and A.I. Lamond. 2012. A quantitative spatial proteomics analysis of proteome turnover in human cells. *Molecular & cellular proteomics : MCP*. 11:M111 011429.

Boisvert, F.M., S. van Koningsbruggen, J. Navascues, and A.I. Lamond. 2007. The multifunctional nucleolus. *Nature reviews. Molecular cell biology*. 8:574-585.

Borer, R.A., C.F. Lehner, H.M. Eppenberger, and E.A. Nigg. 1989. Major nucleolar proteins shuttle between nucleus and cytoplasm. *Cell*. 56:379-390.

Bouche, G., M. Caizergues-Ferrer, B. Bugler, and F. Amalric. 1984. Interrelations between the maturation of a 100 kDa nucleolar protein and pre rRNA synthesis in CHO cells. *Nucleic acids research*. 12:3025-3035.

Bouvet, P., J.J. Diaz, K. Kindbeiter, J.J. Madjar, and F. Amalric. 1998. Nucleolin interacts with several ribosomal proteins through its RGG domain. *The Journal of biological chemistry*. 273:19025-19029.

Brina, D., A. Miluzio, S. Ricciardi, and S. Biffo. 2015. eIF6 anti-association activity is required for ribosome biogenesis, translational control and tumor progression. *Biochimica et biophysica acta*. 1849:830-835.

Brodska, B., A. Holoubek, P. Otevrelova, and K. Kuzelova. 2015. Low-Dose Actinomycin-D Induces Redistribution of Wild-Type and Mutated Nucleophosmin Followed by Cell Death in Leukemic Cells. *Journal of cellular biochemistry*.

Burger, K., B. Muhl, T. Harasim, M. Rohrmoser, A. Malamoussi, M. Orban, M. Kellner, A. Gruber-Eber, E. Kremmer, M. Holzel, and D. Eick. 2010. Chemotherapeutic drugs inhibit ribosome biogenesis at various levels. *The Journal of biological chemistry*. 285:12416-12425.

Burger, K., B. Muhl, M. Kellner, M. Rohrmoser, A. Gruber-Eber, L. Windhager, C.C. Friedel, L. Dolken, and D. Eick. 2013. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA biology*. 10:1623-1630.

Burgess, R.J., and Z. Zhang. 2013. Histone chaperones in nucleosome assembly and human disease. *Nature structural & molecular biology*. 20:14-22.

Calo, E., R.A. Flynn, L. Martin, R.C. Spitale, H.Y. Chang, and J. Wysocka. 2015. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature*. 518:249-253.

Carmo-Fonseca, M., L. Mendes-Soares, and I. Campos. 2000. To be or not to be in the nucleolus. *Nature cell biology*. 2:E107-112.

Cech, T.R., and J.A. Steitz. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*. 157:77-94.

Chan, P.K. 1992. Characterization and cellular localization of nucleophosmin/B23 in HeLa cells treated with selected cytotoxic agents (studies of B23-translocation mechanism). *Experimental cell research*. 203:174-181.

Chartier-Harlin, M.-C., Justus C. Dachsel, C. Vilariño-Güell, Sarah J. Lincoln, F. Leprêtre, Mary M. Hulihan, J. Kachergus, Austen J. Milnerwood, L. Tapia, M.-S. Song, E. Le Rhun, E. Mutez, L. Larvor, A. Duflot, C. Vanbesien-Mailliot, A.

Kreisler, Owen A. Ross, K. Nishioka, Alexandra I. Soto-Ortolaza, Stephanie A. Cobb, Heather L. Melrose, B. Behrouz, Brett H. Keeling, Justin A. Bacon, E. Hentati, L. Williams, A. Yanagiya, N. Sonenberg, Paul J. Lockhart, Abba C. Zubair, Ryan J. Uitti, Jan O. Aasly, A. Krygowska-Wajs, G. Opala, Zbigniew K. Wszolek, R. Frigerio, Demetrius M. Maraganore, D. Gosal, T. Lynch, M. Hutchinson, Anna R. Bentivoglio, Enza M. Valente, William C. Nichols, N. Pankratz, T. Foroud, Rachel A. Gibson, F. Hentati, Dennis W. Dickson, A. Destée, and Matthew J. Farrer. 2011. Translation Initiator EIF4G1 Mutations in Familial Parkinson Disease. *American Journal of Human Genetics*. 89:398-406.

Chen, M., and P. Jiang. 2004. Altered subcellular distribution of nucleolar protein fibrillarin by actinomycin D in HEp-2 cells. *Acta pharmacologica Sinica*. 25:902-906.

Choi, H., T. Glatter, M. Gstaiger, and A.I. Nesvizhskii. 2012a. SAINT-MS1: protein-protein interaction scoring using label-free intensity data in affinity purification-mass spectrometry experiments. *Journal of proteome research*. 11:2619-2624.

Choi, H., B. Larsen, Z.Y. Lin, A. Breitkreutz, D. Mellacheruvu, D. Fermin, Z.S. Qin, M. Tyers, A.C. Gingras, and A.I. Nesvizhskii. 2011. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature methods*. 8:70-73.

Choi, H., G. Liu, D. Mellacheruvu, M. Tyers, A.C. Gingras, and A.I. Nesvizhskii. 2012b. Analyzing protein-protein interactions from affinity purification-mass spectrometry data with SAINT. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*. Chapter 8:Unit8 15.

Chung, K.Y., I.K. Cheng, A.K. Ching, J.H. Chu, P.B. Lai, and N. Wong. 2011. Block of proliferation 1 (BOP1) plays an oncogenic role in hepatocellular carcinoma by promoting epithelial-to-mesenchymal transition. *Hepatology (Baltimore, Md.)*. 54:307-318.

Coady, T.H., and J.L. Manley. 2015. ALS mutations in TLS/FUS disrupt target gene expression. *Genes & development*. 29:1696-1706.

Colis, L., G. Ernst, S. Sanders, H. Liu, P. Sirajuddin, K. Peltonen, M. DePasquale, J.C. Barrow, and M. Laiho. 2014. Design, synthesis, and structure-activity relationships of pyridoquinazolinecarboxamides as RNA polymerase I inhibitors. *Journal of medicinal chemistry*. 57:4950-4961.

Cong, R., S. Das, I. Ugrinova, S. Kumar, F. Mongelard, J. Wong, and P. Bouvet. 2012. Interaction of nucleolin with ribosomal RNA genes and its role in RNA polymerase I transcription. *Nucleic acids research*. 40:9441-9454.

Cooper, G., 2000. The Cell, A Molecular Approach. 2nd edition, *Sunderland (MA), Sinauer Associates*, ISBN-10 0-87893-106-6

Corcoran, D.L., S. Georgiev, N. Mukherjee, E. Gottwein, R.L. Skalsky, J.D. Keene, and U. Ohler. 2011. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome biology*. 12:R79.

Cox, J., M.Y. Hein, C.A. Luber, I. Paron, N. Nagaraj, and M. Mann. 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics : MCP*. 13:2513-2526.

Creancier, L., H. Prats, C. Zanibellato, F. Amalric, and B. Bugler. 1993. Determination of the functional domains involved in nucleolar targeting of nucleolin. *Molecular biology of the cell*. 4:1239-1250.

Damian, L. 2013. Isothermal titration calorimetry for studying protein-ligand interactions. *Methods in molecular biology (Clifton, N.J.)*. 1008:103-118.

Daniely, Y., D.D. Dimitrova, and J.A. Borowiec. 2002. Stress-dependent nucleolin mobilization mediated by p53-nucleolin complex formation. *Molecular and cellular biology*. 22:6014-6022.

Darnell, R.B. 2010. HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley interdisciplinary reviews. RNA*. 1:266-286.

De Keersmaecker, K., S.O. Sulima, and J.D. Dinman. 2015. Ribosomopathies and the paradox of cellular hypo- to hyperproliferation. *Blood*. 125:1377-1382.

de la Cruz, J., K. Karbstein, and J.L. Woolford. 2015. Functions of Ribosomal Proteins in Assembly of Eukaryotic Ribosomes In Vivo. *Annual review of biochemistry*. 84:93-129.

Dickinson, L.A., and T. Kohwi-Shigematsu. 1995. Nucleolin is a matrix attachment region DNA-binding protein that specifically recognizes a region with high base-unpairing potential. *Molecular and cellular biology*. 15:456-465.

Dong, X.Y., P. Guo, J. Boyd, X. Sun, Q. Li, W. Zhou, and J.T. Dong. 2009. Implication of snoRNA U50 in human breast cancer. *Journal of genetics and genomics = Yi chuan xue bao*. 36:447-454.

Dreyfuss, G., V.N. Kim, and N. Kataoka. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nature reviews. Molecular cell biology*. 3:195-205.

Drygin, D., W.G. Rice, and I. Grummt. 2010. The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annual review of pharmacology and toxicology*. 50:131-156.

Drygin, D., A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C.B. Ho, C. Proffitt, K. Trent, J.P. Whitten, J.K. Lim, D. Von Hoff, K. Anderes, and W.G. Rice. 2009. Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer research*. 69:7653-7661.

Dubaele, S., and P. Chene. 2007. Cellular studies of MrDb (DDX18). *Oncology research*. 16:549-556.

Duncan, J.S., and D.W. Litchfield. 2008. Too much of a good thing: the role of protein kinase CK2 in tumorigenesis and prospects for therapeutic inhibition of CK2. *Biochimica et biophysica acta*. 1784:33-47.

Dupuis-Sandoval, F., M. Poirier, and M.S. Scott. 2015. The emerging landscape of small nucleolar RNAs in cell biology. *Wiley interdisciplinary reviews. RNA*. 6:381-397.

Durut, N., and J. Saez-Vasquez. 2015. Nucleolin: dual roles in rDNA chromatin transcription. *Gene*. 556:7-12.

Eberhardt, W., A. Doller, and J. Pfeilschifter. 2012. Regulation of the mRNA-binding protein HuR by posttranslational modification: spotlight on phosphorylation. *Current protein & peptide science*. 13:380-390.

Egyhazi, E., A. Pigon, J.H. Chang, S.H. Ghaffari, T.D. Dreesen, S.E. Wellman, S.T. Case, and M.O. Olson. 1988. Effects of anti-C23 (nucleolin) antibody on transcription of ribosomal DNA in Chironomus salivary gland cells. *Experimental cell research*. 178:264-272.

Ellington, A.D., and J.W. Szostak. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature*. 346:818-822.

Erard, M.S., P. Belenguer, M. Caizergues-Ferrer, A. Pantaloni, and F. Amalric. 1988. A major nucleolar protein, nucleolin, induces chromatin decondensation by binding to histone H1. *European journal of biochemistry / FEBS*. 175:525-530.

Escande, M.L., N. Gas, and B.J. Stevens. 1985. Immunolocalization of the 100 K nucleolar protein in CHO cells. *Biology of the cell / under the auspices of the European Cell Biology Organization*. 53:99-109.

Esteller, M. 2011. Non-coding RNAs in human disease. *Nature reviews. Genetics*. 12:861-874.

Feig, A.L. 2007. Applications of isothermal titration calorimetry in RNA biochemistry and biophysics. *Biopolymers*. 87:293-301.

Finger, L.D., L. Trantirek, C. Johansson, and J. Feigon. 2003. Solution structures of stem-loop RNAs that bind to the two N-terminal RNA-binding domains of nucleolin. *Nucleic acids research*. 31:6461-6472.

Finkbeiner, E., M. Haindl, and S. Muller. 2011a. The SUMO system controls nucleolar partitioning of a novel mammalian ribosome biogenesis complex. *The EMBO journal*. 30:1067-1078.

Finkbeiner, E., M. Haindl, N. Raman, and S. Muller. 2011b. SUMO routes ribosome maturation. *Nucleus (Austin, Tex.)*. 2:527-532.

Foulkes, W.D., J.R. Priest, and T.F. Duchaine. 2014. DICER1: mutations, microRNAs and mechanisms. *Nature reviews. Cancer*. 14:662-672.

Fritz, G., O.G. Issinger, and B.B. Olsen. 2009. Selectivity analysis of protein kinase CK2 inhibitors DMAT, TBB and resorufin in cisplatin-induced stress responses. *International journal of oncology*. 35:1151-1157.

Gardner, P.P., A. Bateman, and A.M. Poole. 2010. SnoPatrol: how many snoRNA genes are there? *Journal of biology*. 9:4.

Gasser, T. 2004. Genetics of Parkinson's disease. *Dialogues in Clinical Neuroscience*. 6:295-301.

Gazda, H.T., A. Grabowska, L.B. Merida-Long, E. Latawiec, H.E. Schneider, J.M. Lipton, A. Vlachos, E. Atsidaftos, S.E. Ball, K.A. Orfali, E. Niewiadomska, L. Da Costa, G. Tchernia, C. Niemeyer, J.J. Meerpohl, J. Stahl, G. Schratt, B. Glader, K. Backer, C. Wong, D.G. Nathan, A.H. Beggs, and C.A. Sieff. 2006. Ribosomal protein S24 gene is mutated in Diamond-Blackfan anemia. *Am J Hum Genet*. 79:1110-1118.

Gazda, H.T., M.R. Sheen, A. Vlachos, V. Choesmel, M.F. O'Donohue, H. Schneider, N. Darras, C. Hasman, C.A. Sieff, P.E. Newburger, S.E. Ball, E. Niewiadomska, M. Matysiak, J.M. Zaucha, B. Glader, C. Niemeyer, J.J. Meerpohl, E. Atsidaftos, J.M. Lipton, P.E. Gleizes, and A.H. Beggs. 2008. Ribosomal protein L5 and L11 mutations are associated with cleft palate and abnormal thumbs in Diamond-Blackfan anemia patients. *Am J Hum Genet*. 83:769-780.

Gerstberger, S., M. Hafner, M. Ascano, and T. Tuschl. 2014a. Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. *Advances in experimental medicine and biology*. 825:1-55.

Gerstberger, S., M. Hafner, and T. Tuschl. 2014b. A census of human RNA-binding proteins. *Nature reviews. Genetics*. 15:829-845.

Ghisolfi-Nieto, L., G. Joseph, F. Puvion-Dutilleul, F. Amalric, and P. Bouvet. 1996. Nucleolin is a sequence-specific RNA-binding protein: characterization of targets on pre-ribosomal RNA. *Journal of molecular biology*. 260:34-53.

Ghisolfi, L., G. Joseph, F. Amalric, and M. Erard. 1992. The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties. *The Journal of biological chemistry*. 267:2955-2959.

Ginisty, H., F. Amalric, and P. Bouvet. 1998. Nucleolin functions in the first step of ribosomal RNA processing. *The EMBO journal*. 17:1476-1486.

Ginisty, H., F. Amalric, and P. Bouvet. 2001. Two different combinations of RNA-binding domains determine the RNA binding specificity of nucleolin. *The Journal of biological chemistry*. 276:14338-14343.

Ginisty, H., H. Sicard, B. Roger, and P. Bouvet. 1999. Structure and functions of nucleolin. *Journal of cell science*. 112 (Pt 6):761-772.

Glisovic, T., J.L. Bachorik, J. Yong, and G. Dreyfuss. 2008. RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters*. 582:1977-1986.

Gonzalez, V., K. Guo, L. Hurley, and D. Sun. 2009. Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *The Journal of biological chemistry*. 284:23622-23635.

Goodfellow, S.J., and J.C. Zomerdijk. 2013. Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes. *Sub-cellular biochemistry*. 61:211-236.

Granneman, S., and S.J. Baserga. 2004. Ribosome biogenesis: of knobs and RNA processing. *Experimental cell research*. 296:43-50.

Greasley, P.J., C. Bonnard, and B. Amati. 2000. Myc induces the nucleolin and BN51 genes: possible implications in ribosome biogenesis. *Nucleic acids research*. 28:446-453.

Hafner, M., P. Landgraf, J. Ludwig, A. Rice, T. Ojo, C. Lin, D. Holoch, C. Lim, and T. Tuschl. 2008. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods (San Diego, Calif.)*. 44:3-12.

Hafner, M., M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, Jr., A.C. Jungkamp, M. Munschauer, A. Ulrich, G.S. Wardle, S. Dewell, M. Zavolan, and T. Tuschl. 2010a. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 141:129-141.

Hafner, M., M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, A.C. Jungkamp, M. Munschauer, A. Ulrich, G.S. Wardle, S. Dewell, M. Zavolan, and T. Tuschl. 2010b. PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *Journal of visualized experiments : JoVE*.

Hafner, M., N. Renwick, M. Brown, A. Mihailovic, D. Holoch, C. Lin, J.T. Pena, J.D. Nusbaum, P. Morozov, J. Ludwig, T. Ojo, S. Luo, G. Schroth, and T. Tuschl. 2011. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA (New York, N.Y.)*. 17:1697-1712.

Hammoudi, A., F. Song, K.R. Reed, R.E. Jenkins, V.S. Meniel, A.J. Watson, D.M. Pritchard, A.R. Clarke, and J.R. Jenkins. 2013. Proteomic profiling of a mouse model of acute intestinal Apc deletion leads to identification of potential novel biomarkers of human colorectal cancer (CRC). *Biochemical and biophysical research communications*. 440:364-370.

Hanakahi, L.A., L.A. Dempsey, M.J. Li, and N. Maizels. 1997. Nucleolin is one component of the B cell-specific transcription factor and switch region binding protein, LR1. *Proceedings of the National Academy of Sciences of the United States of America*. 94:3605-3610.

Hanakahi, L.A., H. Sun, and N. Maizels. 1999. High affinity interactions of nucleolin with G-G-paired rDNA. *The Journal of biological chemistry*. 274:15908-15912.

Hannan, K.M., E. Sanij, L.I. Rothblum, R.D. Hannan, and R.B. Pearson. 2013. Dysregulation of RNA polymerase I transcription during disease. *Biochimica et biophysica acta*. 1829:342-360.

Hathaway, G.M., T.H. Lubben, and J.A. Traugh. 1980. Inhibition of casein kinase II by heparin. *The Journal of biological chemistry*. 255:8038-8041.

Hein, N., K.M. Hannan, A.J. George, E. Sanij, and R.D. Hannan. 2013. The nucleolus: an emerging target for cancer therapy. *Trends in molecular medicine*. 19:643-654.

Hellman, L.M., and M.G. Fried. 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature protocols*. 2:1849-1861.

Henras, A.K., J. Soudet, M. Gerus, S. Lebaron, M. Caizergues-Ferrer, A. Mougin, and Y. Henry. 2008. The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cellular and molecular life sciences : CMLS*. 65:2334-2359.

Hernandez-Verdun, D., P. Roussel, and J. Gebrane-Younes. 2002. Emerging concepts of nucleolar assembly. *Journal of cell science*. 115:2265-2270.

Hochstatter, J., M. Holzel, M. Rohrmoser, L. Schermelleh, H. Leonhardt, R. Keough, T.J. Gonda, A. Imhof, D. Eick, G. Langst, and A. Nemeth. 2012. Myb-binding protein 1a (Mybbp1a) regulates levels and processing of pre-ribosomal RNA. *The Journal of biological chemistry*. 287:24365-24377.

Hong, P., S. Koza, and E.S. Bouvier. 2012. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *Journal of liquid chromatography & related technologies*. 35:2923-2950.

Hovanessian, A.G., C. Soundaramourty, D. El Khoury, I. Nondier, J. Svab, and B. Krust. 2010. Surface expressed nucleolin is constantly induced in tumor cells to mediate calcium-dependent ligand internalization. *PloS one*. 5:e15787.

Hung, M.C., and W. Link. 2011. Protein localization in disease and therapy. *Journal of cell science*. 124:3381-3392.

Iggo, R.D., D.J. Jamieson, S.A. MacNeill, J. Southgate, J. McPheat, and D.P. Lane. 1991. p68 RNA helicase: identification of a nucleolar form and cloning of related genes containing a conserved intron in yeasts. *Molecular and cellular biology*. 11:1326-1333.

Jakobi, R., and J.A. Traugh. 1995. Analysis of the ATP/GTP binding site of casein kinase II by site-directed mutagenesis. *Physiological chemistry and physics and medical NMR*. 27:293-301.

Jalal, C., H. Uhlmann-Schiffler, and H. Stahl. 2007. Redundant role of DEAD box proteins p68 (Ddx5) and p72/p82 (Ddx17) in ribosome biogenesis and cell proliferation. *Nucleic acids research*. 35:3590-3601.

James, A., Y. Wang, H. Raje, R. Rosby, and P. DiMario. 2014. Nucleolar stress with and without p53. *Nucleus (Austin, Tex.)*. 5:402-426.

Kass, S., and B. Sollner-Webb. 1990. The first pre-rRNA-processing event occurs in a large complex: analysis by gel retardation, sedimentation, and UV cross-linking. *Molecular and cellular biology*. 10:4920-4931.

Keene, J.D. 2007. RNA regulons: coordination of post-transcriptional events. *Nature reviews. Genetics*. 8:533-543.

Keene, J.D., J.M. Komisarow, and M.B. Friedersdorf. 2006. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature protocols*. 1:302-307.

Keilhauer, E.C., M.Y. Hein, and M. Mann. 2015. Accurate Protein Complex Retrieval by Affinity Enrichment Mass Spectrometry (AE-MS) Rather than Affinity Purification

Mass Spectrometry (AP-MS). *Molecular & cellular proteomics : MCP*. 14:120-135.

Khandjian, E.W. 1999. Biology of the fragile X mental retardation protein, an RNA-binding protein. *Biochemistry and cell biology = Biochimie et biologie cellulaire*. 77:331-342.

Khatter, H., A.G. Myasnikov, S.K. Natchiar, and B.P. Klaholz. 2015. Structure of the human 80S ribosome. *Nature*. 520:640-645.

Killian, A., N. Sarafan-Vasseur, R. Sesboue, F. Le Pessot, F. Blanchard, A. Lamy, M. Laurent, J.M. Flaman, and T. Frebourg. 2006. Contribution of the BOP1 gene, located on 8q24, to colorectal tumorigenesis. *Genes, chromosomes & cancer*. 45:874-881.

Kishore, S., and S. Stamm. 2006. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science (New York, N.Y.)*. 311:230-232.

Kiss, T. 2004. Biogenesis of small nuclear RNPs. *Journal of cell science*. 117:5949-5951.

Klauck, S.M. 2006. Genetics of autism spectrum disorder. *European journal of human genetics : EJHG*. 14:714-720.

Klauck, S.M., B. Felder, A. Kolb-Kokocinski, C. Schuster, A. Chiocchetti, I. Schupp, R. Wellenreuther, G. Schmotzer, F. Poustka, L. Breitenbach-Koller, and A. Poustka. 2006. Mutations in the ribosomal protein gene RPL10 suggest a novel modulating disease mechanism for autism. *Molecular psychiatry*. 11:1073-1084.

Kobayashi, H., K. Abe, T. Matsuura, Y. Ikeda, T. Hitomi, Y. Akechi, T. Habu, W. Liu, H. Okuda, and A. Koizumi. 2011. Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement. *Am J Hum Genet*. 89:121-130.

Konig, J., K. Zarnack, N.M. Luscombe, and J. Ule. 2011. Protein-RNA interactions: new genomic technologies and perspectives. *Nature reviews. Genetics*. 13:77-83.

Korobeinikova, A.V., M.B. Garber, and G.M. Gongadze. 2012. Ribosomal proteins: structure, function, and evolution. *Biochemistry. Biokhimiia*. 77:562-574.

Krishna, S.S., I. Majumdar, and N.V. Grishin. 2003. Structural classification of zinc fingers: survey and summary. *Nucleic acids research*. 31:532-550.

Lafontaine, D.L. 2015. Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nature structural & molecular biology*. 22:11-19.

Lafontaine, D.L., and D. Tollervey. 2001. The function and synthesis of ribosomes. *Nature reviews. Molecular cell biology*. 2:514-520.

Lam, Y.W., and L. Trinkle-Mulcahy. 2015. New insights into nucleolar structure and function. *F1000prime reports*. 7:48.

Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*. 10:R25.

Lapeyre, B., H. Bourbon, and F. Amalric. 1987. Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. *Proceedings of the National Academy of Sciences of the United States of America*. 84:1472-1476.

Larsson, O., C. Wahlestedt, and J.A. Timmons. 2005. Considerations when using the significance analysis of microarrays (SAM) algorithm. *BMC bioinformatics*. 6:129.

Leary, D.J., and S. Huang. 2001. Regulation of ribosome biogenesis within the nucleolus. *FEBS letters*. 509:145-150.

Lebedeva, S., M. Jens, K. Theil, B. Schwanhausser, M. Selbach, M. Landthaler, and N. Rajewsky. 2011. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Molecular cell*. 43:340-352.

Lee, H., D. Kim, H.C. Dan, E.L. Wu, T.M. Gritsko, C. Cao, S.V. Nicosia, E.A. Golemis, W. Liu, D. Coppola, S.S. Brem, J.R. Testa, and J.Q. Cheng. 2007. Identification and characterization of putative tumor suppressor NGB, a GTP-binding protein that interacts with the neurofibromatosis 2 protein. *Molecular and cellular biology*. 27:2103-2119.

Leiva, L., D. Carrasco, A. Taylor, M. Veliz, C. Gonzalez, C.C. Allende, and J.E. Allende. 1987. Casein kinase II is a major protein phosphorylating activity in the nuclei of Xenopus laevis oocytes. *Biochemistry international*. 14:707-717.

Leung, A.K., J.S. Andersen, M. Mann, and A.I. Lamond. 2003. Bioinformatic analysis of the nucleolus. *The Biochemical journal*. 376:553-569.

Leung, A.K., and A.I. Lamond. 2003. The dynamics of the nucleolus. *Critical reviews in eukaryotic gene expression*. 13:39-54.

Leung, A.K., L. Trinkle-Mulcahy, Y.W. Lam, J.S. Andersen, M. Mann, and A.I. Lamond. 2006. NOPdb: Nucleolar Proteome Database. *Nucleic acids research*. 34:D218-220.

Li, D., G. Dobrowolska, and E.G. Krebs. 1996a. The physical association of casein kinase 2 with nucleolin. *The Journal of biological chemistry*. 271:15662-15668.

Li, X., H. Kazan, H.D. Lipshitz, and Q.D. Morris. 2014. Finding the target sites of RNA-binding proteins. *Wiley interdisciplinary reviews. RNA*. 5:111-130.

Li, Y.P., R.K. Busch, B.C. Valdez, and H. Busch. 1996b. C23 interacts with B23, a putative nucleolar-localization-signal-binding protein. *European journal of biochemistry / FEBS*. 237:153-158.

Liao, J., L. Yu, Y. Mei, M. Guarnera, J. Shen, R. Li, Z. Liu, and F. Jiang. 2010. Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. *Molecular cancer*. 9:198.

Lin, D., Z. Shkedy, T. Burzykowski, R. Ion, H.W. Gohlmann, A.D. Bondt, T. Perer, T. Geerts, I. Van den Wyngaert, and L. Bijnens. 2008. An investigation on performance of Significance Analysis of Microarray (SAM) for the comparisons of several treatments with one control in the presence of small-variance genes. *Biometrical journal. Biometrische Zeitschrift*. 50:801-823.

Lu, Y., P. Liu, M. James, H.G. Vikis, H. Liu, W. Wen, A. Franklin, and M. You. 2010. Genetic variants cis-regulating Xrn2 expression contribute to the risk of spontaneous lung tumor. *Oncogene*. 29:1041-1049.

Lukong, K.E., K.W. Chang, E.W. Khandjian, and S. Richard. 2008. RNA-binding proteins in human genetic disease. *Trends in genetics : TIG*. 24:416-425.

Lunde, B.M., C. Moore, and G. Varani. 2007. RNA-binding proteins: modular design for efficient function. *Nature reviews. Molecular cell biology*. 8:479-490.

Lundgren, D.H., S.I. Hwang, L. Wu, and D.K. Han. 2010. Role of spectral counting in quantitative proteomics. *Expert review of proteomics*. 7:39-53.

Mannoor, K., J. Liao, and F. Jiang. 2012. Small nucleolar RNAs in cancer. *Biochimica et biophysica acta*. 1826:121-128.

Martin, R., A.U. Straub, C. Doebele, and M.T. Bohnsack. 2013. DExD/H-box RNA helicases in ribosome biogenesis. *RNA biology*. 10:4-18.

Marcel, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 17, 10–12 10.14806/ej.17.1.200

Mayer, C., and I. Grummt. 2006. Ribosome biogenesis and cell growth: mTOR coordinates transcription by all three classes of nuclear RNA polymerases. *Oncogene*. 25:6384-6391.

Medina, F.J., F. González-Camacho, A.I. Manzano, A. Manrique, and R. Herranz. 2010. Nucleolin, a major conserved multifunctional nucleolar phosphoprotein of proliferating cells. *Journal of Applied Biomedicine*. 8:141-150.

Meister, G. 2011. RNA Biology, An Introduction. *Wiley-VCH*, ISBN 978-3-527-32278-7

Meister, G. 2013. Argonaute proteins: functional insights and emerging roles. *Nature reviews. Genetics*. 14:447-459.

Miller, O.L., Jr., and B.R. Beatty. 1969. Visualization of nucleolar genes. *Science (New York, N.Y.)*. 164:955-957.

Miluzio, A., A. Beugnet, V. Volta, and S. Biffo. 2009. Eukaryotic initiation factor 6 mediates a continuum between 60S ribosome biogenesis and translation. *EMBO reports*. 10:459-465.

Mongelard, F., and P. Bouvet. 2007. Nucleolin: a multiFACeTed protein. *Trends in cell biology*. 17:80-86.

Mongelard, F., and P. Bouvet. 2010. AS-1411, a guanosine-rich oligonucleotide aptamer targeting nucleolin for the potential treatment of cancer, including acute myeloid leukemia. *Current opinion in molecular therapeutics*. 12:107-114.

Morello, L.G., P.P. Coltri, A.J. Quaresma, F.M. Simabuco, T.C. Silva, G. Singh, J.A. Nickerson, C.C. Oliveira, M.J. Moore, and N.I. Zanchin. 2011a. The human nucleolar protein FTSJ3 associates with NIP7 and functions in pre-rRNA processing. *PloS one*. 6:e29174.

Morello, L.G., P.P. Coltri, A.J.C. Quaresma, F.M. Simabuco, T.C.L. Silva, G. Singh, J.A. Nickerson, C.C. Oliveira, M.J. Moore, and N.I.T. Zanchin. 2011b. The Human Nucleolar Protein FTSJ3 Associates with NIP7 and Functions in Pre-rRNA Processing. *PloS one*. 6:e29174.

Morris, A.R., N. Mukherjee, and J.D. Keene. 2010. Systematic analysis of posttranscriptional gene expression. *Wiley interdisciplinary reviews. Systems biology and medicine*. 2:162-180.

Moss, T., F. Langlois, T. Gagnon-Kugler, and V. Stefanovsky. 2007. A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis. *Cellular and molecular life sciences : CMLS*. 64:29-49.

Mougey, E.B., M. O'Reilly, Y. Osheim, O.L. Miller, Jr., A. Beyer, and B. Sollner-Webb. 1993. The terminal balls characteristic of eukaryotic rRNA transcription units in chromatin spreads are rRNA processing complexes. *Genes & development*. 7:1609-1619.

Mukherjee, N., D.L. Corcoran, J.D. Nusbaum, D.W. Reid, S. Georgiev, M. Hafner, M. Ascano, Jr., T. Tuschl, U. Ohler, and J.D. Keene. 2011. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Molecular cell*. 43:327-339.

Mullineux, S.T., and D.L. Lafontaine. 2012. Mapping the cleavage sites on mammalian pre-rRNAs: where do we stand? *Biochimie*. 94:1521-1532.

Murano, K., M. Okuwaki, M. Hisaoka, and K. Nagata. 2008. Transcription regulation of the rRNA gene by a multifunctional nucleolar protein, B23/nucleophosmin,

through its histone chaperone activity. *Molecular and cellular biology*. 28:3114-3126.

Narla, A., and B.L. Ebert. 2010. Ribosomopathies: human disorders of ribosome dysfunction. *Blood*. 115:3196-3205.

Nesvizhskii, A.I. 2012. Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. *Proteomics*. 12:1639-1655.

Nicoloso, M., M. Caizergues-Ferrer, B. Michot, M.C. Azum, and J.P. Bachellerie. 1994. U20, a novel small nucleolar RNA, is encoded in an intron of the nucleolin gene in mammals. *Molecular and cellular biology*. 14:5766-5776.

Nikolay, R., D. van den Bruck, J. Achenbach, and K.H. Nierhaus. 2001. Ribosomal Proteins: Role in Ribosomal Functions. *In* eLS. John Wiley & Sons, Ltd.

Nousbeck, J., R. Spiegel, A. Ishida-Yamamoto, M. Indelman, A. Shani-Adir, N. Adir, E. Lipkin, S. Bercovici, D. Geiger, M.A. van Steensel, P.M. Steijlen, R. Bergman, A. Bindereif, M. Choder, S. Shalev, and E. Sprecher. 2008. Alopecia, neurological defects, and endocrinopathy syndrome caused by decreased expression of RBM28, a nucleolar protein associated with ribosome biogenesis. *Am J Hum Genet*. 82:1114-1121.

O'Day, D. H., Catalano, A. 2013. Proteins of the Nucleolus: Regulation, *Translocation, & Biomedical Functions*. ISBN 978-94-007-5818-6

Olsen, J.V., S.E. Ong, and M. Mann. 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular & cellular proteomics : MCP*. 3:608-614.

Olson, M.O.J., 2011. The Nucleolus. *Springer,* ISBN 978-1-4614-0514-6

Orrick, L.R., M.O. Olson, and H. Busch. 1973. Comparison of nucleolar proteins of normal rat liver and Novikoff hepatoma ascites cells by two-dimensional polyacrylamide gel electrophoresis. *Proceedings of the National Academy of Sciences of the United States of America*. 70:1316-1320.

Otake, Y., S. Soundararajan, T.K. Sengupta, E.A. Kio, J.C. Smith, M. Pineda-Roman, R.K. Stuart, E.K. Spicer, and D.J. Fernandes. 2007. Overexpression of nucleolin in chronic lymphocytic leukemia cells induces stabilization of bcl2 mRNA. *Blood*. 109:3069-3075.

Pagano, M.A., G. Poletto, G. Di Maira, G. Cozza, M. Ruzzene, S. Sarno, J. Bain, M. Elliott, S. Moro, G. Zagotto, F. Meggio, and L.A. Pinna. 2007. Tetrabromocinnamic acid (TBCA) and related compounds represent a new class of specific protein kinase CK2 inhibitors. *Chembiochem : a European journal of chemical biology*. 8:129-139.

Panov, K.I., J.K. Friedrich, J. Russell, and J.C. Zomerdijk. 2006. UBF activates RNA polymerase I transcription by stimulating promoter escape. *The EMBO journal*. 25:3310-3322.

Panova, T.B., K.I. Panov, J. Russell, and J.C. Zomerdijk. 2006. Casein kinase 2 associates with initiation-competent RNA polymerase I and has multiple roles in ribosomal DNA transcription. *Molecular and cellular biology*. 26:5957-5968.

Park, M.H., S.A. Cho, K.H. Yoo, M.H. Yang, J.Y. Ahn, H.S. Lee, K.E. Lee, Y.C. Mun, D.H. Cho, C.M. Seong, and J.H. Park. 2007. Gene expression profile related to prognosis of acute myeloid leukemia. *Oncology reports*. 18:1395-1402.

Payne, E.M., N. Bolli, J. Rhodes, O.I. Abdel-Wahab, R. Levine, C.V. Hedvat, R. Stone, A. Khanna-Gupta, H. Sun, J.P. Kanki, H.T. Gazda, A.H. Beggs, F.E. Cotter, and

A.T. Look. 2011. Ddx18 is essential for cell-cycle progression in zebrafish hematopoietic cells and is mutated in human AML. *Blood*. 118:903-915.

Pederson, T. 2002. Proteomics of the nucleolus: more proteins, more functions? *Trends in biochemical sciences*. 27:111-112.

Pfaff, M., and F.A. Anderer. 1988. Casein kinase II accumulation in the nucleolus and its role in nucleolar phosphorylation. *Biochimica et biophysica acta*. 969:100-109.

Phipps, K.R., J. Charette, and S.J. Baserga. 2011a. The small subunit processome in ribosome biogenesis-progress and prospects. *Wiley interdisciplinary reviews. RNA*. 2:1-21.

Phipps, K.R., J.M. Charette, and S.J. Baserga. 2011b. The SSU Processome in Ribosome Biogenesis – Progress and Prospects. *Wiley interdisciplinary reviews. RNA*. 2:1-21.

Pichiorri, F., D. Palmieri, L. De Luca, J. Consiglio, J. You, A. Rocci, T. Talabere, C. Piovan, A. Lagana, L. Cascione, J. Guan, P. Gasparini, V. Balatti, G. Nuovo, V. Coppola, C.C. Hofmeister, G. Marcucci, J.C. Byrd, S. Volinia, C.L. Shapiro, M.A. Freitas, and C.M. Croce. 2013. In vivo NCL targeting affects breast cancer aggressiveness through miRNA regulation. *The Journal of experimental medicine*. 210:951-968.

Pickard, A.J., and U. Bierbach. 2013. The cell's nucleolus: an emerging target for chemotherapeutic intervention. *ChemMedChem*. 8:1441-1449.

Pilarsky, C., M. Wenzig, T. Specht, H.D. Saeger, and R. Grutzmann. 2004. Identification and validation of commonly overexpressed genes in solid tumors by comparison of microarray data. *Neoplasia (New York, N.Y.)*. 6:744-750.

Popov, A., E. Smirnov, L. Kovacik, O. Raska, G. Hagen, L. Stixova, and I. Raska. 2013. Duration of the first steps of the human rRNA processing. *Nucleus (Austin, Tex.)*. 4:134-141.

Qian, J., Z.X. Chen, J. Lin, S.Y. Hu, J.N. Cen, and W. Wang. 2005. [Quantitative study on C/EBP zeta gene transcripts in patients with chronic myeloid leukemia using real-time quantitative PCR]. *Zhonghua yi xue yi chuan xue za zhi = Zhonghua yixue yichuanxue zazhi = Chinese journal of medical genetics*. 22:628-631.

Qiu, W., F. Zhou, Q. Zhang, X. Sun, X. Shi, Y. Liang, X. Wang, and L. Yue. 2013. Overexpression of nucleolin and different expression sites both related to the prognosis of gastric cancer. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*. 121:919-925.

Quin, J.E., J.R. Devlin, D. Cameron, K.M. Hannan, R.B. Pearson, and R.D. Hannan. 2014. Targeting the nucleolus for cancer intervention. *Biochimica et biophysica acta*. 1842:802-816.

R Core Team (2014). R: A language and environment for statistical computing.

R Foundation for Statistical Computing, Vienna, Austria.

Rebane, A., and A. Metspalu. 1999. U82, a novel snoRNA identified from the fifth intron of human and mouse nucleolin gene. *Biochimica et biophysica acta*. 1446:426-430.

Reyes-Reyes, E.M., Y. Teng, and P.J. Bates. 2010. A new paradigm for aptamer therapeutic AS1411 action: Uptake by macropinocytosis and its stimulation by a nucleolin-dependent mechanism. *Cancer research*. 70:8617-8629.

Rickards, B., S.J. Flint, M.D. Cole, and G. LeRoy. 2007. Nucleolin is required for RNA polymerase I transcription in vivo. *Molecular and cellular biology*. 27:937-948.

Robledo, S., R.A. Idol, D.L. Crimmins, J.H. Ladenson, P.J. Mason, and M. Bessler. 2008. The role of human ribosomal proteins in the maturation of rRNA and ribosome production. *RNA (New York, N.Y.)*. 14:1918-1929.

Rocak, S., and P. Linder. 2004. DEAD-box proteins: the driving forces behind RNA metabolism. *Nature reviews. Molecular cell biology*. 5:232-241.

Roger, B., A. Moisand, F. Amalric, and P. Bouvet. 2002. Repression of RNA polymerase I transcription by nucleolin is independent of the RNA sequence that is transcribed. *The Journal of biological chemistry*. 277:10209-10219.

Roger, B., A. Moisand, F. Amalric, and P. Bouvet. 2003. Nucleolin provides a link between RNA polymerase I transcription and pre-ribosome assembly. *Chromosoma*. 111:399-407.

Ronnau, C.G., G.W. Verhaegh, M.V. Luna-Velez, and J.A. Schalken. 2014. Noncoding RNAs as novel biomarkers in prostate cancer. *BioMed research international*. 2014:591703.

Rubbi, C.P., and J. Milner. 2003. Disruption of the nucleolus mediates stabilization of p53 in response to DNA damage and other stresses. *The EMBO journal*. 22:6068-6077.

Ruggero, D., and P.P. Pandolfi. 2003. Does the ribosome translate cancer? *Nature reviews. Cancer*. 3:179-192.

Sanij, E., G. Poortinga, K. Sharkey, S. Hung, T.P. Holloway, J. Quin, E. Robb, L.H. Wong, W.G. Thomas, V. Stefanovsky, T. Moss, L. Rothblum, K.M. Hannan, G.A. McArthur, R.B. Pearson, and R.D. Hannan. 2008. UBF levels determine the number of active ribosomal RNA genes in mammals. *The Journal of cell biology*. 183:1259-1274.

Saporita, A.J., H.C. Chang, C.L. Winkeler, A.J. Apicelli, R.D. Kladney, J. Wang, R.R. Townsend, L.S. Michel, and J.D. Weber. 2011. RNA helicase DDX5 is a p53-independent target of ARF that participates in ribosome biogenesis. *Cancer research*. 71:6708-6717.

Sarno, S., H. Reddy, F. Meggio, M. Ruzzene, S.P. Davies, A. Donella-Deana, D. Shugar, and L.A. Pinna. 2001. Selectivity of 4,5,6,7-tetrabromobenzotriazole, an ATP site-directed inhibitor of protein kinase CK2 ('casein kinase-2'). *FEBS letters*. 496:44-48.

Scherl, A., Y. Coute, C. Deon, A. Calle, K. Kindbeiter, J.C. Sanchez, A. Greco, D. Hochstrasser, and J.J. Diaz. 2002. Functional proteomic analysis of human nucleolus. *Molecular biology of the cell*. 13:4100-4109.

Schlosser, I., M. Hölzel, M. Mürnseer, H. Burtscher, U.H. Weidle, and D. Eick. 2003. A role for c-Myc in the regulation of ribosomal RNA processing. *Nucleic acids research*. 31:6148-6156.

Schmidt-Zachmann, M.S., and E.A. Nigg. 1993. Protein localization to the nucleolus: a search for targeting domains in nucleolin. *Journal of cell science*. 105 (Pt 3):799-806.

Schneider, H.R., and O.G. Issinger. 1988. Nucleolin (C23), a physiological substrate for casein kinase II. *Biochemical and biophysical research communications*. 156:1390-1397.

Schwender H (2012). siggenes: Multiple testing using SAM and Efron's empirical

    Bayes approaches. *R package version 1.44.0.*

Scott, M.S., F.M. Boisvert, M.D. McDowall, A.I. Lamond, and G.J. Barton. 2010. Characterization and prediction of protein nucleolar localization sequences. *Nucleic acids research*. 38:7388-7399.

Scott, M.S., P.V. Troshin, and G.J. Barton. 2011. NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins. *BMC bioinformatics*. 12:317.

Semenkovich, C.F., R.E. Ostlund, Jr., M.O. Olson, and J.W. Yang. 1990. A protein partially expressed on the surface of HepG2 cells that binds lipoproteins specifically is nucleolin. *Biochemistry*. 29:9708-9713.

Serin, G., G. Joseph, L. Ghisolfi, M. Bauzan, M. Erard, F. Amalric, and P. Bouvet. 1997. Two RNA-binding domains determine the RNA-binding specificity of nucleolin. *The Journal of biological chemistry*. 272:13109-13116.

Sharma, S., and D.L. Lafontaine. 2015. 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. *Trends in biochemical sciences*. 40:560-575.

Shen, N., F. Yan, J. Pang, L.C. Wu, A. Al-Kali, M.R. Litzow, and S. Liu. 2014. A nucleolin-DNMT1 regulatory axis in acute myeloid leukemogenesis. *Oncotarget*. 5:5494-5509.

Shenoy, N., R. Kessel, T.D. Bhagat, S. Bhattacharyya, Y. Yu, C. McMahon, and A. Verma. 2012. Alterations in the ribosomal machinery in cancer and hematologic disorders. *Journal of hematology & oncology*. 5:32.

Sigler, P.B. 1988. Transcriptional activation. Acid blobs and negative noodles. *Nature*. 333:210-212.

Simos, G., and E. Hurt. 1999. Transfer RNA biogenesis: A visa to leave the nucleus. *Current biology : CB*. 9:R238-241.

Sloan, K.E., M.S. Leisegang, C. Doebele, A.S. Ramirez, S. Simm, C. Safferthal, J. Kretschmer, T. Schorge, S. Markoutsa, S. Haag, M. Karas, I. Ebersberger, E. Schleiff, N.J. Watkins, and M.T. Bohnsack. 2015. The association of late-acting snoRNPs with human pre-ribosomal complexes requires the RNA helicase DDX21. *Nucleic acids research*. 43:553-564.

Slomovic, S., D. Laufer, D. Geiger, and G. Schuster. 2006. Polyadenylation of ribosomal RNA in human cells. *Nucleic acids research*. 34:2966-2975.

Soeno, Y., Y. Taya, T. Stasyk, L.A. Huber, T. Aoba, and A. Huttenhofer. 2010. Identification of novel ribonucleo-protein complexes from the brain-specific snoRNA MBII-52. *RNA (New York, N.Y.)*. 16:1293-1300.

Sontheimer, E.J. 1994. Site-specific RNA crosslinking with 4-thiouridine. *Molecular biology reports*. 20:35-44.

Soundararajan, S., L. Wang, V. Sridharan, W. Chen, N. Courtenay-Luck, D. Jones, E.K. Spicer, and D.J. Fernandes. 2009. Plasma membrane nucleolin is a receptor for the anticancer aptamer AS1411 in MV4-11 leukemia cells. *Molecular pharmacology*. 76:984-991.

Spitzer, J., M. Hafner, M. Landthaler, M. Ascano, T. Farazi, G. Wardle, J. Nusbaum, M. Khorshid, L. Burger, M. Zavolan, and T. Tuschl. 2014. PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a step-by-step protocol to the transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods in enzymology*. 539:113-161.

Spitzer, J., M. Landthaler, and T. Tuschl. 2013. Rapid creation of stable mammalian cell lines for regulated expression of proteins using the Gateway(R)

recombination cloning technology and Flp-In T-REx(R) lines. *Methods in enzymology*. 529:99-124.

Srivastava, M., and H.B. Pollard. 1999. Molecular dissection of nucleolin's role in growth and cell proliferation: new insights. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 13:1911-1922.

Stefanovsky, V., F. Langlois, T. Gagnon-Kugler, L.I. Rothblum, and T. Moss. 2006. Growth factor signaling regulates elongation of RNA polymerase I transcription in mammals via UBF phosphorylation and r-chromatin remodeling. *Molecular cell*. 21:629-639.

Storck, S., M. Shukla, S. Dimitrov, and P. Bouvet. 2007. Functions of the histone chaperone nucleolin in diseases. *Sub-cellular biochemistry*. 41:125-144.

Storck, S., M. Thiry, and P. Bouvet. 2009. Conditional knockout of nucleolin in DT40 cells reveals the functional redundancy of its RNA-binding domains. *Biology of the cell / under the auspices of the European Cell Biology Organization*. 101:153-167.

Su, H., T. Xu, S. Ganapathy, M. Shadfan, M. Long, T.H. Huang, I. Thompson, and Z.M. Yuan. 2014. Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene*. 33:1348-1358.

Sutandy, F.X., A. Hildebrandt, and J. Konig. 2016. Profiling the Binding Sites of RNA-Binding Proteins with Nucleotide Resolution Using iCLIP. *Methods in molecular biology (Clifton, N.J.)*. 1358:175-195.

Tafforeau, L., C. Zorbas, J.L. Langhendries, S.T. Mullineux, V. Stamatopoulou, R. Mullier, L. Wacheul, and D.L. Lafontaine. 2013. The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. *Molecular cell*. 51:539-551.

Tajrishi, M.M., R. Tuteja, and N. Tuteja. 2011. Nucleolin: The most abundant multifunctional phosphoprotein of nucleolus. *Communicative & integrative biology*. 4:267-275.

Takahashi, N., M. Yanagida, S. Fujiyama, T. Hayano, and T. Isobe. 2003. Proteomic snapshot analyses of preribosomal ribonucleoprotein complexes formed at various stages of ribosome biogenesis in yeast and mammalian cells. *Mass spectrometry reviews*. 22:287-317.

Teng, B., C. Zhao, X. Liu, and Z. He. 2015. Network inference from AP-MS data: computational challenges and solutions. *Briefings in bioinformatics*. 16:658-674.

Teng, Y., A.C. Girvan, L.K. Casson, W.M. Pierce, Jr., M. Qian, S.D. Thomas, and P.J. Bates. 2007. AS1411 alters the localization of a complex containing protein arginine methyltransferase 5 and nucleolin. *Cancer research*. 67:10491-10500.

Teo, G., G. Liu, J. Zhang, A.I. Nesvizhskii, A.C. Gingras, and H. Choi. 2014. SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *Journal of proteomics*. 100:37-43.

Thomson, E., S. Ferreira-Cerca, and E. Hurt. 2013. Eukaryotic ribosome biogenesis at a glance. *Journal of cell science*. 126:4815-4821.

Townley-Tilson, W.H., S.A. Pendergrass, W.F. Marzluff, and M.L. Whitfield. 2006. Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein. *RNA (New York, N.Y.)*. 12:1853-1867.

Turner, A.J., A.A. Knox, J.L. Prieto, B. McStay, and N.J. Watkins. 2009. A novel small-subunit processome assembly intermediate that contains the U3 snoRNP, nucleolin, RRP5, and DBP4. *Molecular and cellular biology*. 29:3007-3017.

Turowski, T.W., and D. Tollervey. 2015. Cotranscriptional events in eukaryotic ribosome synthesis. *Wiley interdisciplinary reviews. RNA*. 6:129-139.

Tusher, V.G., R. Tibshirani, and G. Chu. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*. 98:5116-5121.

Tuteja, R., and N. Tuteja. 1998. Nucleolin: a multifunctional major nucleolar phosphoprotein. *Critical reviews in biochemistry and molecular biology*. 33:407-436.

Ugrinova, I., K. Monier, C. Ivaldi, M. Thiry, S. Storck, F. Mongelard, and P. Bouvet. 2007. Inactivation of nucleolin leads to nucleolar disruption, cell cycle arrest and defects in centrosome duplication. *BMC molecular biology*. 8:66.

Ule, J., K. Jensen, A. Mele, and R.B. Darnell. 2005. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods (San Diego, Calif.)*. 37:376-386.

van Riggelen, J., A. Yetil, and D.W. Felsher. 2010. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nature reviews. Cancer*. 10:301-309.

Wang, T., S.M. Bray, and S.T. Warren. 2012. New perspectives on the biology of fragile X syndrome. *Current opinion in genetics & development*. 22:256-263.

Wild, T., P. Horvath, E. Wyler, B. Widmann, L. Badertscher, I. Zemp, K. Kozak, G. Csucs, E. Lund, and U. Kutay. 2010. A protein inventory of human ribosome biogenesis reveals an essential function of exportin 5 in 60S subunit export. *PLoS biology*. 8:e1000522.

Williamson, D., Y.J. Lu, C. Fang, K. Pritchard-Jones, and J. Shipley. 2006. Nascent pre-rRNA overexpression correlates with an adverse prognosis in alveolar rhabdomyosarcoma. *Genes, chromosomes & cancer*. 45:839-845.

Wiseman, T., S. Williston, J.F. Brandts, and L.N. Lin. 1989. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical biochemistry*. 179:131-137.

Woolford, J.L., Jr., and S.J. Baserga. 2013. Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics*. 195:643-681.

Yanagida, M., A. Shimamoto, K. Nishikawa, Y. Furuichi, T. Isobe, and N. Takahashi. 2001. Isolation and proteomic characterization of the major proteins of the nucleolin-binding ribonucleoprotein complexes. *Proteomics*. 1:1390-1404.

Yedavalli, V.S., C. Neuveut, Y.H. Chi, L. Kleiman, and K.T. Jeang. 2004. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell*. 119:381-392.

Zamore, P.D., J.R. Williamson, and R. Lehmann. 1997. The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *RNA (New York, N.Y.)*. 3:1421-1433.

Zhang, J., G. Tsaprailis, and G.T. Bowden. 2008. Nucleolin stabilizes Bcl-X L messenger RNA in response to UVA irradiation. *Cancer research*. 68:1046-1054.

Zhang, Y., J.T. Forys, A.P. Miceli, A.S. Gwinn, and J.D. Weber. 2011. Identification of DHX33 as a mediator of rRNA synthesis and cell growth. *Molecular and cellular biology*. 31:4676-4691.

Zhao, H., Y. Huang, C. Xue, Y. Chen, X. Hou, Y. Guo, L. Zhao, Z. Hu, Y. Huang, Y. Luo, and L. Zhang. 2013. Prognostic significance of the combined score of endothelial expression of nucleolin and CD31 in surgically resected non-small cell lung cancer. *PloS one*. 8:e54674.

Zhao, J., T.K. Ohsumi, J.T. Kung, Y. Ogawa, D.J. Grau, K. Sarma, J.J. Song, R.E. Kingston, M. Borowsky, and J.T. Lee. 2010. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Molecular cell*. 40:939-953.

Zhou, H., M. Mangelsdorf, J. Liu, L. Zhu, and J.Y. Wu. 2014. RNA-binding proteins in neurological diseases. *Science China. Life sciences*. 57:432-444.

Zien, P., J.S. Duncan, J. Skierski, M. Bretner, D.W. Litchfield, and D. Shugar. 2005. Tetrabromobenzotriazole (TBBt) and tetrabromobenzimidazole (TBBz) as selective inhibitors of protein kinase CK2: evaluation of their effects on cells and different molecular forms of human CK2. *Biochimica et biophysica acta*. 1754:271-280.

STATEMENT OF AFFIRMATION

I declare on oath that I composed the dissertation 'Nucleolin: a nucleolar RNA-binding protein involved in ribosome biogenesis'' independently by myself. I only used the references and resources indicated and quoted within this thesis. The thesis was not submitted or presented to another institution or examination board. I have not taken a doctoral examination without success so far.

_____

Düsseldorf, 25th of April 2016   Julia Fremerey

ACKNOWLEDGMENT

CURRICULUM VITAE

Personal details

Name:	Julia Fremerey
Nationality:	German
Date of birth:	26. November 1983
City of birth:	Hamburg

Education

PhD

07/2012 - 03/2016	'Nucleolin: a nucleolar RNA-binding protein involved in ribosome biogenesis'
	Department of Pediatric Oncology, Hematology and Immunology, Prof. Dr. A. Borkhardt,
	Heinrich-Heine University, Düsseldorf
03/2013 - 12/2013	Laboratory of RNA Molecular Biology
06/2014 - 08/2014	Prof. Dr. T. Tuschl, Rockefeller University,
03/2015 - 05/2015	New York, USA

Master

10/2009 - 04/2012	Master of Science in Biology
	Westphalian Wilhelms-University of Münster
04/2010 - 09/2010	'Patch-clamp analysis of the ε-subunit of ENaC in oocytes from *Xenopus laevis'*,
	Department of Electrophysiology and Biophysics,
	Prof. Dr. M. Awayda, State University of New York at Buffalo, USA
Master Thesis	
05/2011 - 04/2012	'Antibody and peptide mediated delivery of siRNA into tumor cells',
	Institute of molecular Hematology and Oncology,
	Prof. Dr. C. Müller-Tidow, University Hospital Münster

Bachelor

10/2006 - 09/2009	Bachelor of Science in Biology, Westphalian Wilhelms-University of Münster
08/2008 - 12/2008	Norwegian University of Science and Technology, Trondheim, Norway

Bachelor Thesis

03/2009 - 08/2009	'Pharmacokinetic analysis of ENaC-specific antisense oligonculeotides for inhibition of sodium absorption in primary epithelial cells', Institute of animal physiology, Prof. Dr. M. Weber, Westphalian Wilhelms-University of Münster

Work experience and Internship

01/2011 - 03/2011	Internship at Miltenyi Biotec, Research and Development, Bergisch Gladbach
08/2005 - 10/2005	'Microparticle enzyme immunoassay for the quantitative determination of BNP in EDTA plasma', Internship at Axis-Shield Diagnostics, Research and Development, Dundee, Scotland
06/2011 - 12/2011	Student research assistant at the Max-Planck Institute of Molecular Biomedicine, Münster

Apprenticeship

08/2004 - 06/2006	Biological technical assistant (BTA), Hilden

High School

08/1994 - 05/2004	Abitur, Städtisches Gymnasium Haan
08/2000 - 07/2001	Lafayette-County High-School, Higginsville, USA

Language skills

German (mother tongue)

English (fluent)

French (basic)