

Department for Biomedical Sciences
University of Veterinary Medicine Vienna
Institute of Medical Biochemistry
(Head: Prof. Dr. Florian Grebien)

Characterization of Nucleoporin 98-Fusion Proteins

Master's thesis submitted for the fulfillment of the requirements for the degree
of

Master of Science M.Sc.

University of Veterinary Medicine Vienna

Submitted by

Theresa Humer, B.Sc.

Vienna, January 2020

SUPERVISION

Internal Supervisor (University of Veterinary Medicine Vienna)

Prof. Dr. Florian Grebien

Institute of Medical Biochemistry

SWORN DECLARATION

I hereby declare that I prepared this work independently and without help from third parties, that I did not use sources other than the ones referenced and that I have indicated passages taken from those sources.

This thesis was not previously submitted in identical or similar form to any other examination board, nor was it published.

.....

Theresa Humer

Vienna, January 2020

ACKNOWLEDGEMENT

First, I want to thank Dr. Florian Grebien for giving me the opportunity to work on this project and his constructive feedback to my work and thesis. I am very grateful for having the chance to write my thesis under his supervision. Furthermore, I would like to express my appreciation to Stefan Terlecki-Zaniewicz, who guided me through the whole project and supported me during my work. I want to thank all members of the Grebien group for their support, advice and warm atmosphere from the first day on. This research was supported using resources of the VetCore Facility (Imaging) of the University of Veterinary Medicine Vienna with great thanks to Ursula Reichart for fruitful discussions. Further, I would like to generally thank my friends and family for their steady support and especially to my parents, who encouraged me a lot during my years of study.

Table of Contents

1	Introduction	1
1.1	Hematopoiesis	1
1.2	Leukemia	2
1.2.1	Acute myeloid leukemia (AML)	3
1.2.2	Fusion proteins in leukemia and AML	5
1.2.3	Nucleoporin 98-fusion protein	6
1.3	The nuclear pore complex	8
1.3.1	Nucleoporin 98 (NUP98)	9
1.4	Biomolecular condensation	11
1.5	Aim of the study	13
2	Materials and methods.....	14
2.1	Constructs	14
2.2	Cell cultures	14
2.3	Retroviral transduction	15
2.4	Cell lysis for protein harvest	15
2.5	Biotinylated isoxazole-mediated precipitation	16
2.6	Western blot analysis	16
2.7	Immunofluorescence analysis.....	17
2.8	Live cell imaging	18

2.9	RNA isolation and preparation for RNA-seq analysis	18
2.10	Bioinformatic analysis and statistics	19
3	Results	20
3.1	Verification of exogenous expression of NUP98-fusion proteins.....	20
3.2	Biotinylated isoxazole-mediated precipitation of NUP98-fusion proteins.....	21
3.3	Optimization of staining protocols and image acquisition for confocal IF analysis of NUP98-fusion proteins	23
3.4	Visualization of NUP98-fusion proteins via confocal IF microscopy	25
3.5	Analysis of the subcellular localization of different artificial IDR-Containing KDM5A-fusion proteins by live-cell imaging.....	26
3.6	Analysis of the ability of artificial IDR-containing KDM5A-fusion proteins to drive leukemia-associated gene expression	28
4	Discussion	31
	Abstract.....	38
	Zusammenfassung.....	39
	List of figures	41
	List of tables.....	42
	References.....	43

1 INTRODUCTION

1.1 HEMATOPOIESIS

Hematopoiesis describes the continuous process of production of all cellular parts of the blood (Jagannathan-Bogdan and Zon, 2013). In a healthy person hematopoietic stem cells (HSCs) are located in the bone marrow and possess unlimited self-renewal properties. Through asymmetric cell division HSCs maintain the HSC pool and produce progenitor cells, such as multi-potent progenitor (MPP) and lineage-specific progenitors (Seita and Weissman, 2010) at the same time. The latter can be divided into myeloid and lymphoid-committed progenitors. These cells further differentiate into mature cells, including granulocytes, T-cells, erythrocytes (Figure 1) and other cells that do not have self-renewal capacity. In general, cells gradually lose their ability of self-renewal throughout the differentiation process (Chopra and Bohlander, 2019).

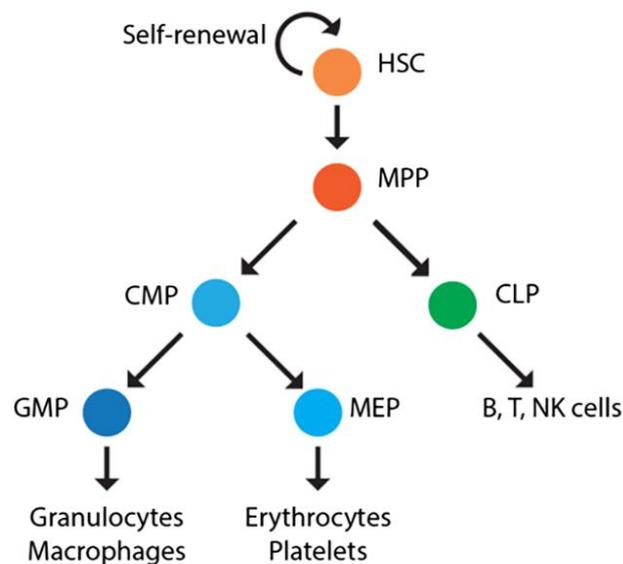


Figure 1: Schematic representation of healthy hematopoiesis, originating from the hematopoietic stem cell (HSC), which divides asymmetrically to maintain the HSC pool and produce multi-potent as well as lineage-committed progenitors, which give rise to terminally differentiated cells. CLP, common lymphoid progenitor; GMP, granulocyte macrophage progenitor; HSC, hematopoietic stem cell; CMP, common myeloid progenitor; MEP, megakaryocyte erythroid progenitor; MPP, multipotent progenitor. Reprinted with permission from Wiley for Chopra and Bohlander, 2019

1.2 LEUKEMIA

Leukemia is a malignant disease of the blood-forming cells (Zhao, Wang and Ma, 2018). Leukemic hematopoiesis partially resembles the hierarchical differentiation that is observed in normal hematopoiesis. Thus, most leukemia subtypes are maintained by leukemia stem cells (LSCs). However, LSC can be as dormant as normal HSCs or possess extended self-renewal capacity. In the blood system, accumulation of the progeny of LSC leads to various types of leukemia, depending on the cell type and differentiation status of the actual leukemia cells. In leukemias, formation of mature cells is often impaired through mutations in hematopoietic differentiation pathways (Chopra and Bohlander, 2019), such as *NOTCH* (Liu, Zhang and Ji, 2013) and *WNT* pathway (Staal *et al.*, 2016). LSCs are further characterized by their ability to initiate leukemia in transplantation experiments (Lapidot *et al.*, 1994). While leukemia is propagated by LSCs, the cell of origin of leukemia (COL) can represent a different cell in the hematopoietic hierarchy. For instance, the COL can be a HSC, which has acquired aberrant proliferation capacity (Mut1a in Figure 2) or a progenitor cell, that has gained aberrant self-renewal potential (Mut1b in Figure 2). The COL gains additional mutations (Mut2 and Mut3 in Figure 2) leading to the generation of LSCs (Chopra and Bohlander, 2019).

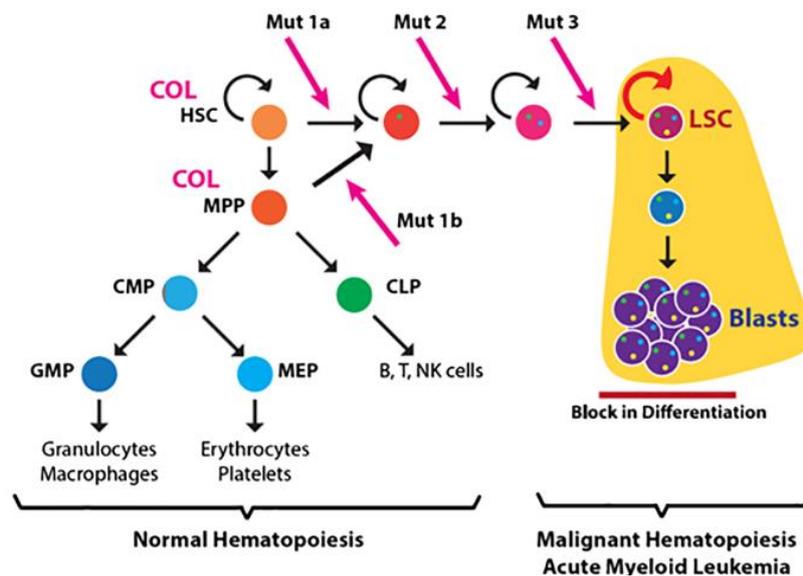


Figure 2: Hematopoiesis in a person with leukemia. The cell of origin of leukemia (COL) gains several mutations that lead to the generation of leukemic stem cells (LSC) and accumulation of blasts in the blood. Scenario 1: a HSC sequentially accumulates mutation 1a, 2 and 3 and is transformed into a LSC with unlimited self-renewal capacity and the ability to produce leukemic blasts. Scenario 2: The COL in this case is a MPP which acquires mutation 1b which allows to re-activate self-renewal pathways and afterwards follows the path as in scenario 1. Reprinted with permission from Wiley for Chopra and Bohlander, 2019.

Leukemia is divided into four major subgroups depending on their progression, COL and cell type identity:

*Chronic lymphoid leukemia (CLL)

*Acute lymphoid leukemia (ALL)

*Chronic myeloid leukemia (CML)

*Acute myeloid leukemia (AML) (Zhao, Wang and Ma, 2018)

Among acute leukemias, ~80 % of cases are acute myeloid leukemia (AML), while ~20 % represent acute lymphoblastic leukemia (ALL) (Fahrenkrog, 2014). Acute leukemia is diagnosed when more than 30 % of cells in the bone marrow are blasts. This classification is important for treatment decision, prediction of response and efficacy of therapy (Löwenberg, Downing and Burnett, 2012).

1.2.1 ACUTE MYELOID LEUKEMIA (AML)

The World Health Organization classifies AML as a tumor of the hematopoietic system. The disease is characterized by the clonal expansion of aberrantly differentiating hematopoietic cells of the myeloid compartment leading to bone marrow failure and disruption of normal hematopoiesis (Döhner, Weisdorf and Bloomfield, 2015). In AML, the progeny of LSCs is often referred to as blasts and can be either mature or immature myeloid cells, depending on the subtype of AML (Chopra and Bohlander, 2019). Myeloblasts, monoblasts, promonocytes, megakaryoblasts (but not dysplastic megakaryocytes) are regarded as blasts (Vardiman 2009). AML is diagnosed in people of every age (Bolouri *et al.*, 2018), but elderly people at a median of 69 years are predominantly affected (Juliussen *et al.*, 2009). This disease is the most frequent type of acute leukemia in adults and also very common in infants (Pavlović and Tomic, 2019). The cure rate is approximately 40 % in all patients below 60 years but only up to 15% in patients older than 60 years (Döhner, Weisdorf and Bloomfield, 2015). In pediatric AML failure rates after primary therapy are up to 15 % (McNeer *et al.*, 2019). Population studies revealed a five-year survival rate of less than 8 % in AML patients older than 60 years of age. The according rate in patients younger than 60 years is about 50 % (Oran and Weisdorf, 2012).

Prior to initiation of therapy, patient stratification is important for optimizing therapy options (Döhner, Weisdorf and Bloomfield, 2015). Classical intensive induction chemotherapy for AML is based on cytarabine and an anthracycline, commonly daunorubicin, in a 7+3 regime. In this scheme an administration of 7 days cytarabine is combined with a 3 day-administration of daunorubicin. Complete remission is morphologically defined as less than 5 % blasts in the bone marrow and recovery of bone marrow functions as reported by normalization of peripheral blood counts (Burnett, Wetzler and Löwenberg, 2014). In AML complete remission is seen in 60 % to 80 % of patients younger than 60 years and 40 % to 60 % in patients over 60 years (Döhner *et al.*, 2017). Stem cell transplantation is another way of treatment, either using allogenic – from a human leucocyte antigen-compatible relative – or autologous transplantation. It is mainly considered for patients with a high risk of relapse (Burnett, Wetzler and Löwenberg, 2014).

In the last decades, the standard therapeutic strategy for AML has not changed a lot. The improvements in survival of younger patients were mainly achieved by better supportive care and higher standards for hematopoietic stem cell transplantation (Watts and Nimer, 2018). CPX-315, a drug formulation of cytarabine and daunorubicin in a 5:1 molar ratio in liposomes, has shown positive effects in therapy-related and secondary AML (Döhner *et al.*, 2017; Watts and Nimer, 2018). Alternatively, CD33 is expressed on the surface of myeloid blasts in the majority of AML patients, exposing a possible therapeutic target by antibody-administration (De Propriis *et al.*, 2011). Gemtuzumab ozogamizine is a monoclonal antibody-toxin conjugate recognizing human CD33. In this drug, the toxin calicheamicin (ozogamizine) is bound to the antibody gemtuzumab (Baron and Wang, 2019).

Usually, AML arises from a combination of two types of mutations: type one mutations often lead to activation of signaling pathways driving proliferation, such as *FLT3*-internal tandem duplication (*FLT3-ITD*) or mutations in *NRAS*. Type two mutations interfere with transcriptional processes governing normal hematopoietic differentiation, causing a block in myeloid differentiation pathways and enhanced self-renewal potential of hematopoietic progenitor cells (Fahrenkrog, 2014) such as mutations in *RUNX1* or *CEBPA* (Gerritsen *et al.*, 2019).

The cytogenetic and molecular heterogeneity of AML represents the sequential acquisition of advantageous mutations leading to dynamic clonal evolution (Döhner, Weisdorf and Bloomfield, 2015). 55 % of all AML patients harbor clonal cytogenetic changes such as translocations, inversions, duplications, deletions and insertions. The remaining 45 % are classified as AML with normal karyotype due to non-detectable or non-existing chromosomal aberrations (Pavlović and Tosić, 2019).

Recent advances in next-generation sequencing allowed a deeper understanding of mutations leading to cancer. Only somatic mutations in 10 genes (*DNMT3A*, *IDH1*, *IDH2*, *CEBPA*, *NRAS*, *FLT3*, *WT1*, *RUNX1*, *TP53* and *TET2*) have frequencies >5 % in AML. For example, *FLT3* is mutated in 25-45 % AML cases, whereas *NRAS* mutations occur in about 5-15 % of patients with AML (Pavlović and Tosić, 2019). Therefore, targeted drugs play an important role in leukemia treatment. FLT3 inhibitors such as midostaurin and gilteritinib are currently given to patients with *FLT3*-mutated AML in addition to chemotherapy. However, while most patients initially respond to targeted therapy, many of them develop resistance and relapse, which can be caused by epigenetic changes, co-occurring mutations, selecting for specific subclones, which do not respond to targeted therapy, and other factors (Watts and Nimer, 2018).

1.2.2 FUSION PROTEINS IN LEUKEMIA AND AML

The first chromosomal translocation that was detected in cancer is the Philadelphia chromosome in CML. It is the result of a translocation event between chromosomes 9 and 22, leading to fusion of the *BCR* and *ABL1* genes (Mitelman, Johansson and Mertens, 2007). The BCR-ABL fusion protein is a driver of CML development (Brehme *et al.*, 2009; Soverini *et al.*, 2018). This finding contributed to our current understanding that chromosomal changes influence the occurrence of cancer. 50 % of all fusion genes found in hematological malignancies and solid tumors involve transcription factors and tyrosine kinases. At present, 264 gene fusions were identified in hematological malignancies (Mitelman, Johansson and Mertens, 2007). In particular cases of gene fusions, one fusion partner gene recombines with numerous different partner genes (Mitelman, Johansson and Mertens, 2007). This phenomenon leads to the formation of so-called multi-partner translocation families (MPTF) (Mertens *et al.*, 2015). Examples for promiscuous gene fusions are *ETV6*, which leads to very distinct tumors

depending on its fusion partner, *EWSR1* or *MLL*. *MLL* fusions comprise the largest MPTF in leukemia, featuring over 70 fusion (Mitelman, Johansson and Mertens, 2007; Mertens *et al.*, 2015; Marschalek, 2016).

1.2.3 NUCLEOPORIN 98-FUSION PROTEIN

The N-terminal part of Nucleoporin 98 (NUP98) is rearranged with more than 25 known fusion partners in AML, and up to 2% of AML cases are caused by NUP98-fusion proteins (Gough, Slape and Aplan, 2011). All NUP98-fusion proteins emerge from translocation events between chromosome 11, where *NUP98* is located, and the chromosome of the fusion partner gene e.g. *NSD1* on chromosome 5 (Thanasopoulou, Tzankov and Schwaller, 2014) or *HOXA9* on chromosome 7 (Gough, Slape and Aplan, 2011). Not only balanced translocations but also inversions can lead to NUP98-fusion proteins, as in the case of the NUP98-DDX10 fusion (*inv(11)(p15q21–q23)*) (Fahrenkrog, 2014). Until now, more than 25 different NUP98-fusion proteins have been described in patients. In all of them the N-terminal part of endogenous NUP98, harboring phenylalanine-glycine (FG)-repeats, is fused to a C-terminal fragment of the fusion partner gene (Yung *et al.*, 2011). NUP98-fusion partners genes can be split into two groups depending on their molecular function: homeobox-domain (HD) containing proteins comprise about a dozen partner genes, while non-HD fusion proteins represent about 20 partner genes (Fahrenkrog, 2014). While HD fusion partners of NUP98 share a defined DNA binding domain, non-HD partners often feature chromatin regulatory functions, such as a plant homeodomain (PHD), which has the ability to interact with chromatin as in the NUP98-KDM5A fusions. Alternatively, the NUP98-NSD1 fusion protein comprises a Su(var)3-9 (suppressor of variegation), Enhancer-of-Zeste and Trithorax- (SET) domain, which possesses methyltransferase activity. Furthermore, the C-terminal parts of the DEAD (Asp-Glu-Ala-Asp)-box RNA helicase DDX10 or the DNA topoisomerase TOP1 can also be fused to NUP98 (Gough, Slape and Aplan, 2011). Together, around 30% of all known NUP98-fusion partners do not have any annotated DNA-, RNA- or histone binding domains but contain protein-protein interaction domains or coiled-coil structures, such as present in the NUP98-PSIP1 fusion. All NUP98-fusion proteins are expressed from the endogenous *NUP98* promoter. 9 out of 10 NUP98-related AML cases are characterized by co-occurring *FLT3-ITD*, *NRAS* or *KRAS* mutations, which supposedly enhance the malignant outgrowth of leukemia cells (Fahrenkrog,

2014). In mouse models, expression of NUP98-KDM5A (Wang *et al.*, 2009), NUP98-NSD1 (Wang *et al.*, 2007), or NUP98-HOXA9 (Kroon *et al.*, 2001) leads to AML development (Gough, Slape and Aplan, 2011).

In AML with NUP98-fusion proteins, expression of genes in the *HOXA* and *HOXB* locus is commonly changed on transcriptional levels. This gene cluster is actively transcribed during hematopoiesis and early developmental (Fahrenkrog, 2014). *HOXA* genes are expressed at very low levels in mature hematopoietic cells (Lebert-Ghali *et al.*, 2016).

Usually, HOX gene cluster activation is regulated via Trithorax and Polycomb proteins. In healthy cells, Trithorax proteins activate and Polycomb members silence *HOX* gene expression by changing the epigenetic landscape in the *HOX* gene cluster. However, leukemic cells harboring NUP98 fusions feature constitutive activation of expression of *HOX* clusters genes and this has been shown to promote leukemogenesis (Fahrenkrog, 2014; Brien, Kimberly and Armstrong, 2019). In leukemic blasts, Trithorax-associated H3K4 tri-methylation, H3K36 dimethylation and histone acetylation patterns lead to active transcription of *HOX* genes and a block of hematopoietic differentiation. For NUP98-fusion proteins containing PHD domains, such as NUP98-NSD1 and NUP98-KDM5A, it has been shown that binding of PHD fingers inhibits Polycomb binding and therefore silencing of the *HOX* locus does not occur (Fahrenkrog, 2014). In case of NUP98-NSD1, the PHD-finger reinforces the interaction of the fusion protein with chromatin at specific sites, allowing the SET domain in the NSD1-moiety to induce specific epigenetic changes (Wang *et al.*, 2007). Similarly, the PHD finger of NUP98-KDM5A has a high affinity towards H3K4me2 and H3K4me3 (Wang *et al.*, 2009). Therefore, these NUP98-fusion proteins lock the *HOX* locus in an active chromatin state (Wang *et al.*, 2007, 2009; Fahrenkrog, 2014; Brien, Kimberly and Armstrong, 2019). However, little is known about the molecular mechanism employed by other NUP98-fusion proteins to induce leukemia development.

1.3 THE NUCLEAR PORE COMPLEX

The inner and outer membranes of the nuclear envelope are connected via the nuclear pore complex (NPC). The NPC facilitates bidirectional nucleo-cytoplasmic traffic of macromolecules. Small molecules diffuse through the nuclear envelope, whereas the exchange of macromolecules is dependent on receptor proteins. In humans, the NPC has a molecular weight of approximately 120 MDa and is built in an eightfold symmetry facing the central transport channel (Ibarra and Hetzer, 2015). Each NPC contains ~30 different nuclear pore proteins called nucleoporins (NUPs) in multiple copies. These NUPs contribute to the formation of structural subunits within the NPC. The nuclear and cytoplasmic rings, the inner pore ring, the cytoplasmic filaments and the nuclear basket are the main elements of the NPC (Ibarra and Hetzer, 2015; Beck and Hurt, 2017) (Figure 3). NUPs can be subdivided into scaffold NUPs and NUPs harboring FG-repeats (Beck and Hurt, 2017). Due to the biophysical properties of the FG-repeats, FG-NUPs contain intrinsically disordered regions (IDRs). The FG-NUPs are located within the central channel of the NPC and interact transiently with nuclear transport receptors (NTR) which are required for the nucleocytoplasmic shuttling of macromolecules (Beck and Hurt, 2017). NTRs belong to the group of karyopherins, which subdivided into importins and exportins (Griffis *et al.*, 2002). FG-NUPs are not only important for transport functions but are also involved in scaffolding of the NPC (Beck and Hurt, 2017). In addition to FG-repeats FG-NUPs contain additional motifs, such as RNA recognition motifs or coiled-coiled domains, which mediate important functions (Vikal and Kaur, 2017). FG-NUPs are arranged in a sieve-like meshwork to regulate molecular transport through the NPC. Due to their FG-repeats and unstructured regions, FG-NUPs create a barrier with very distinct biophysical properties allowing for selectivity of molecular transport through the central channel (Beck and Hurt, 2017).

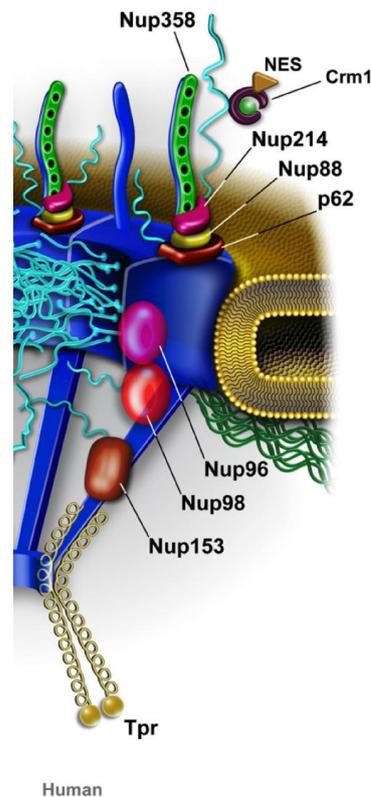


Figure 3: Schematic drawing of the NPC. The eightfold symmetrical NPC with linked filaments is embedded in the nuclear membrane. The nuclear basket is highlighted in dark blue. FG-repeat domains, depicted in turquoise, span into the inner channel of the NPC leading to selectivity of bidirectional nucleocytoplasmic transport. Reprinted with permission from Elsevier for Köhler and Hurt, 2010.

1.3.1 NUCLEOPORIN 98 (NUP98)

The NUP98 protein is an evolutionarily conserved member of the NUP family and is crucial for the activity and assembly of the NPC (Köhler and Hurt, 2010). NUP98 is located within the central transport channel (Beck and Hurt, 2017) and belongs to the group of FG-NUPs as it contains FG-repeats in its N-terminal part (Yung *et al.*, 2011). However, NUP98 differs from other nucleoporins in its high density of non-tandemly arranged glycine-leucine-phenylalanine-glycine (GLFG)-repeats. These repeats act as binding sites for karyopherins (e.g. exportin 1 protein (XPO1)) similar to the FG-repeats in other FG-NUPs. Furthermore, the histone acetyltransferase CREB-binding protein (CBP)/p300 (Kasper *et al.*, 1999) and the mRNA export factor Tip-associated protein (Tap) (Blevins *et al.*, 2003) can interact with NUP98 via binding to the FG-repeats. Furthermore, a Gle2-binding motif (GBD) in the NUP98 N-terminus provides a binding site for the RNA-export factor RAE1 (Gough, Slape and Aplan, 2011)

(Figure 4). The nuclear export of macromolecules represents the primary function of NUP98 (Griffis *et al.*, 2002). However, Griffis *et al.* (2002) demonstrated that NUP98 does not only localize to the NPC but can also reside within the nucleoplasm and assemble within punctuate structures that were termed GLFG bodies, as the GLFG-repeats of NUP98 were required to form these structures (Griffis *et al.*, 2002). Therefore, NUP98 is defined as a mobile nucleoporin (Köhler and Hurt, 2010).

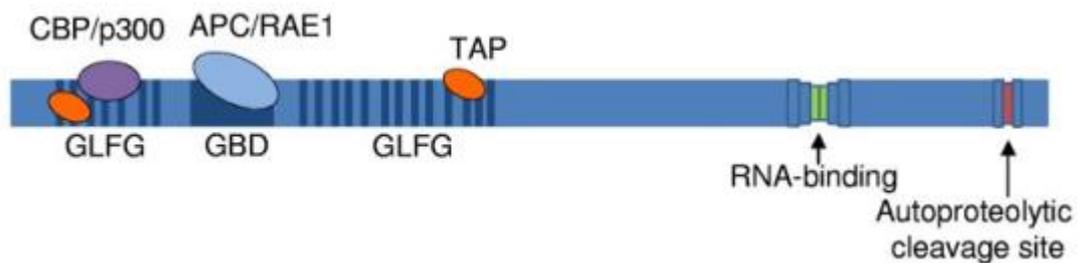


Figure 4: Schematic representation of the NUP98 protein. GLFG-repeats are indicated as blue lines, the dark blue box depicts the GLEBS binding domain (GBD). The NUP98-interactors CBP/p300 and Tap are represented in purple and orange, respectively. The GBD is bound by a complex comprised of APC and RAE1. Reprinted with permission from The American Society of Hematology for Gough, Slape and Aplan, 2011.

In addition to its role in nucleocytoplasmic transport, NUP98 is important for mitotic spindle arrangement (Funasaka *et al.*, 2011). Heterozygous knock-out of NUP98 in combination with or without RAE1 led to an early dissociation of sister chromatids (Funasaka *et al.*, 2011; Gough, Slape and Aplan, 2011). Furthermore, the anaphase promoting complex (APC) can connect with NUP98 (Baker *et al.*, 2007). Kalverda *et al.* (2010) demonstrated that NUP98 interacts with transcriptionally active genes, which are involved in the regulation of development. Genes important for cell cycle regulation (cyclin B), development and mitotic spindle formation were affected (Kalverda *et al.*, 2010; Franks and Hetzer, 2013; Liang *et al.*, 2013; Ibarra and Hetzer, 2015).

1.4 BIOMOLECULAR CONDENSATION

All eukaryotic cells feature membrane-enclosed and membraneless organelles. The formation of the latter can result from a dynamic liquid de-mixing process termed liquid-liquid phase separation (LLPS) (Altmeyer *et al.*, 2015; Gomes and Shorter, 2019). A simplified analogy of this process is the mixture of water and oil, where oil builds distinct droplets despite both phases exhibiting liquid-like behavior. Membraneless organelles are also referred to as biomolecular condensates (Gomes and Shorter, 2019). The concept of biomolecular condensates has been known for a long time, as the nucleolus was the first identified biomolecular condensate (Brangwynne, Mitchison and Hyman, 2011; Banani *et al.*, 2017). Additional biomolecular condensates can be found in the nucleus, such as Cajal bodies and paraspeckles (Mitrea and Kriwacki, 2016) and also in the cytosol, such as P-bodies or stress granules (Mitrea and Kriwacki, 2016). More recently, enhancer complexes and elongating RNA polymerase II were proposed to reside in similar structures (Sabari *et al.*, 2018). Within a cell, biomolecular condensates allow the acceleration of biochemical reactions, as molecules can reach high spatiotemporal concentrations within these compartments. A group of proteins that contain modules, that allow multiple inter- and intramolecular interaction sites, drives the formation of biomolecular condensates (Banani *et al.*, 2017). Multivalency of proteins leading to phase separation can be mediated by IDRs, as they are capable of forming weak, dynamic interactions. Alternatively, multivalent interactions can also be mediated by RNA- and DNA-binding domains, as RNA and DNA molecules themselves also can undergo phase separation (Hyman, Weber and Jülicher, 2014; Alberti, 2017; Banani *et al.*, 2017; Qamar *et al.*, 2018; Gomes and Shorter, 2019; Wheeler *et al.*, 2019). IDRs are conformationally highly flexible regions that lack defined three-dimensional structures and are therefore often referred to as low-complexity domains. Furthermore, IDRs of proteins possess specific biochemical and biophysical properties, such as a high frequency of polar and uncharged amino acids (e.g. tyrosine, serine, glycine) (Aguzzi and Altmeyer, 2016; Gomes and Shorter, 2019). As a consequence of the increased interest in biomolecular condensation, multiple tools for predicting a protein's capability of biomolecular condensation via LLPS based on its amino acid sequences have been developed, such as IUpred2A (Mészáros, Erdős and Dosztányi, 2018) and PLAAC (Lancaster *et al.*, 2014). Furthermore, various techniques to analyze and characterize biomolecular condensates have been identified, including the usage of 1,6-Hexanediol, fluorescence recovery

after photobleaching (FRAP), droplet formation assay and biotinylated isoxazole (Kato *et al.*, 2012; Boija *et al.*, 2018; Gomes and Shorter, 2019; McSwiggen *et al.*, 2019). However, all these reagents and technologies suffer from severe drawbacks due to limited specificity and/or general cytotoxicity. In 2011, Brangwynne *et al.* demonstrated the liquid-like behavior of nucleoli, as shown by their ability to fuse under mechanically induced proximity (Brangwynne, Mitchison and Hyman, 2011). To investigate biomolecular condensation, *in vitro*-studies using the recombinant proteins in combination with live cell imaging or immunofluorescence (IF) microscopy is very frequently performed (Qamar *et al.*, 2018; Sabari *et al.*, 2018; Navarro *et al.*, 2019). However, many of these studies neglect the complex molecular ratios and interactions that exist *in vivo* (Navarro *et al.*, 2019). Another approach includes the use of optogenetics. In this approach, domains from proteins that undergo light-activated dimerization are fused to the proteins of interest harboring IDRs, such as FUS, leading to light-inducible biomolecular condensates which were referred to as “optoDroplets” (Shin *et al.*, 2017).

Proteomic analysis revealed that about 40% of all human proteins harbor IDRs (Toretsky and Wright, 2014). However, the abundance of IDR-containing proteins is predicted to be higher (66%) in human cancer-associated proteins, indicating a role of IDRs and biomolecular condensation in cancer (Iakoucheva *et al.*, 2002). Indeed, several studies have shown, that aberrant biomolecular condensation and diseases are closely linked. For instance, members of the FET protein family, consisting of FUS, EWS, and TAF15, form biomolecular condensates in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (Aguzzi and Altmeyer, 2016). Within the NUP family, only the FG-NUPs contain IDRs (Xu and Powers, 2013; Beck and Hurt, 2017). The FG-domains of NUP98 are able to undergo biomolecular condensation mediated by LLPS (Schmidt and Görlich, 2015; Beck and Hurt, 2017). As the low-complexity domains are located in the NUP98 N-terminus, they are retained in oncogenic NUP98-fusion proteins found in AML and could play a role in leukemogenesis (Xu *et al.*, 2016). Already in 2002 Griffis *et al.* investigated the intra-nuclear localization of NUP98 and found that NUP98 assembles within structures that were coined “GLFG-bodies” (Griffis *et al.*, 2002). Furthermore, a speckled pattern formation was also recorded for NUP98-fusion proteins using live cell imaging and IF microscopy (Fahrenkrog *et al.*, 2016; Xu *et al.*, 2016).

1.5 AIM OF THE STUDY

The *NUP98* gene is fused to over 25 partner genes in AML. AML patients harboring NUP98-fusion proteins suffer from bad prognosis and no targeted therapies are available for this disease subtype. The molecular mechanisms underlying NUP98-fusion protein-mediated oncogenic transformation are still poorly understood. Therefore, there is a strong need to increase our understanding of NUP98-fusion protein-driven leukemogenesis to further develop tailored treatments.

As the IDR-containing part of NUP98 is maintained in all fusions and has been shown to be capable of separating by phase, we hypothesized that the FG-repeats and biomolecular condensation are implicated in NUP98-fusion protein mediated leukemogenesis. To study this potentially shared mechanism between structurally distinct NUP98-fusion proteins, we characterized the subcellular localization of five different NUP98-fusion proteins, analyzed their capacity to form biomolecular condensates and their phase separation properties in IF and live-cell imaging approaches. Furthermore, we investigated whether NUP98-fusion driven AML is dependent on FG-repeats and if biomolecular condensation via IDRs in the N-terminal part of the NUP98-fusion protein is sufficient to drive leukemia-associated gene expression.

2 MATERIALS AND METHODS

2.1 CONSTRUCTS

NUP98-fusion protein constructs were previously cloned into a retroviral vector in a Tet-On system harboring the tetracycline response element (TRE) upstream of the sequence of interest. In this system, only in the presence of tetracycline or chemical analogs, such as doxycycline, the reverse tetracycline transactivator (rtTA3) binds the TRE and induces transcription of the gene. The NUP98-fusion transgenes were tagged with a Strep-tag and an influenza virus hemagglutinin (HA) epitope on the N-terminus (pSIN-TREt-SH-gateway cloning site (gw)-IRES-GFP-PGK-BlastR). Furthermore, the expression of a blasticidin resistance (BlastR) gene was under the control of the phosphoglycerate kinase promoter (PGK). The internal ribosome entrance site (IRES) was cloned upstream of a green fluorescent protein (GFP) cassette. For constitutive gene expression a murine stem cell virus (MSCV) vector harboring mCherry as fluorescence marker was used (MSCV-SH-gw-PGK-BlastR-IRES-mCherry). pcDNA-N-emerald green fluorescent protein (EmGFP)-gw was used for generating GFP-fusion proteins. The Gateway[®] cloning strategy was applied for all cloning steps.

2.2 CELL CULTURES

HL-60 and HEK293-T cells were purchased from the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ). HL-60 cells were previously engineered to stably express the ecotropic receptor and the reverse tetracycline transactivator protein (rtTA3) (HL-60-RIEP). HL-60-RIEP cells were maintained in Roswell Park Memorial Institute 1640 (RPMI 1640, Gibco) with 10 % fetal bovine serum (FBS, Gibco), 100 U/mL Penicillin-Streptomycin (Pen/Strep) and 2 mM L-Glutamine as supplements. HEK293-T were cultivated in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) with 10 % FBS, 100 U/mL Pen/Strep and 2 mM L-Glutamine. Transduction of HL-60-RIEP cells with TREt-SH-NUP98-fusion protein-IRES-GFP-PGK-BlastR-containing retroviral particles was performed in the presence of polybrene (final concentration 10 µg/mL, Merck Millipore/TR-1003-G) and followed by selection with 2 µg/mL puromycin and 10 µg/mL blasticidin for 7 days. Cells were treated with 1 µg/mL

doxycycline (Sigma Aldrich) for 24 hours to induce transgene expression. Expression of NUP98-fusion proteins was verified by fluorescence activated cell sorting (FACS) analyses for GFP-positivity.

Murine fetal liver cells were isolated from pregnant mice (C57BL/6N background) around day E13.5 and E14.5 and maintained in DMEM/ Iscove's Modified Dulbecco's Media (IMDM), 1:1 ratio with 10 % FBS; 100 U/mL Pen/Strep, 4 mM L-Glutamine and 50 μ M beta-mercaptoethanol. 150 ng/mL Murine stem cell factor (mSCF, PreproTech). 10 ng/mL murine interleukin 3 (mIL-3, PreproTech) and 10 ng/mL murine interleukin 6 (mIL-6, PreproTech) were added to culture media after each splitting.

2.3 RETROVIRAL TRANSDUCTION

For retroviral transductions, Platinum-E (Plat-E) cells (purchased from Cell Biolabs, Inc.) were transiently transfected with pGAG-POL and retroviral expression vectors using polyethylenimine (PEI). Virus-containing supernatant was harvested and filtered (0.45 μ m). Target cells were transduced via spinoculation (900g for 45 minutes on two successive days) with retroviral supernatants supplemented with 10 μ g/mL polybrene (Merck Millipore / TR-1003-G) to enhance the transduction efficacy.

2.4 CELL LYSIS FOR PROTEIN HARVEST

For cell harvest $1-10 \times 10^7$ cells were collected, washed with phosphate buffered saline (PBS) and pellets were frozen in liquid nitrogen. For cell lysis, cell pellets were resuspended in ice-cold affinity-purification mass spectrometry (AP-MS) buffer (50 mM hydroxyethyl piperazineethanesulfonic acid-potassium hydroxide (HEPES-KOH) pH 8.0, 100 mM potassium chloride (KCl), 2 mM ethylenediamine tetraacetic acid (EDTA), 0.1 % NP40, 10 % glycerol, 1x Protease inhibitor cocktail (25x, 11697498001, Sigma), 50 mM sodium fluoride (NaF), 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM dithiothreitol (DTT), 10 μ g/mL tosyl phenylalanyl chloromethyl ketone (TPCK), ratio of frozen cell pellet to buffer 1:4). After resuspension, samples were frozen in liquid nitrogen again and thawed at 37 °C and occasionally inverted. 30 sec sonication was followed by adding 125 U benzonase (Sigma-

Aldrich) and samples were inverted a few times during a 1-hour incubation at 4 °C. Samples were centrifuged at 16,600 g for 30 minutes at 4 °C. The supernatant was moved to a new collection tube and the protein concentration was determined using the Bradford assay. Gamma-globulin was used to generate a standard curve.

2.5 BIOTINYLATED ISOXAZOLE-MEDIATED PRECIPITATION

$1-2 \times 10^7$ cells were harvested, washed in PBS and frozen in liquid nitrogen or directly used. If frozen samples were utilized, they were thawed on ice and 1 mL ice-cold EE-buffer (50 mM HEPES- sodium hydroxide (NaOH) pH7.4, 150 mM sodium chloride (NaCl), 0,1 % NP40, 1 mM EDTA, 2.5 mM EGTA, 10 % Glycerol, 1x Protease inhibitor cocktail (25x), 50 mM NaF, 1 mM PMSF, 1 mM DTT, 10 µg/mL TPCK) was applied for resuspension. Samples were incubated at 4 °C for 30 minutes with rotation and centrifuged at 12,500 rpm for 15 minutes at 4 °C. The supernatant was placed into a new collection tube on ice. 50 µl of the sample was transferred into a new tube and served as input sample. 100 µM biotinylated isoxazole was added to the remaining supernatant, tubes were rotated for 1 hour at 4 °C and then centrifuged at 12,500 rpm at 4 °C for 15 minutes. 50 µl of the supernatant were transferred to a new collection tube and used as supernatant fraction. The pellet was washed two times with EE-buffer. For western blot analysis pellets were dissolved in 60 µL Laemmli buffer containing beta-mercapoethanol. 10 µl Laemmli buffer were added to the input sample.

2.6 WESTERN BLOT ANALYSIS

For western blot analysis 10 % or 7 % sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels were prepared using the following components:

Table 1: Components for SDS-PAGE gel

Component	Running Gel (10 mL)	Stacking Gel (5 mL)
Gel Buffer 1 (4x)	2.5 mL	-
Gel Buffer 2 (1x)	-	5 mL
Acrylamide 30 %	3.33 mL (10 %)/2.33 mL (7 %)	-
ddH₂O	4.17 mL (10 %)/5.17 mL (7 %)	-

APS 10 %	100 μ L	50 μ L
TEMED	10 μ L	7.5 μ L

Samples were boiled at 95 °C for 10 minutes before they were loaded on the gel. Transfer of proteins onto a nitrocellulose membrane was achieved by using the Trans-Blot Turbo Transfer System. Blocking of membranes was performed in 5 % milk in tris-buffered saline with Tween[®]20 (TBS-T) for 30 minutes. Membranes were incubated overnight at 4 °C with the following primary antibody solutions: anti-HA.11 (BioLegend, 901513; 1:2000, RRID: AB_2565335), anti-alpha Tubulin (Abcam, ab7291; 1:5000, RRID: AB_2241126), anti-beta Actin (Cell Signaling, 4967S; 1:5000, RRID: AB_330288), anti-HSC70 (Santa Cruz, sc-7298; 1:10000, RRID: AB_627761), anti-NUP98 (Cell Signaling, #2288; 1:1000, RRID: AB_561204), anti-RAE1 (Cell Signaling, sc-374261; 1:1000, RRID: AB_11008069). After incubation, membranes were washed 3 times for 10 minutes in TBS-T followed by incubation with the secondary antibody for 1 hour at room temperature. The following secondary antibodies were used: sheep anti-mouse HRP (GE Healthcare Austria GmbH & Co OG, NA931V; 1:10000, RRID: AB_772210), goat anti-rabbit HRP (Cell Signaling, 7074; 1:10000, RRID: AB_2099233). Membranes were washed again 3 times for 10 minutes with TBS-T and incubated for 2 minutes in ECL Western Blotting Detection Reagent (Fisher scientific) (1:1 solution A:B; solution A: luminol solution, solution B: peroxide solution). A Fusion FX imaging system (Vilber) was used to detect signals.

2.7 IMMUNOFLUORESCENCE ANALYSIS

HL-60 cells that were treated with doxycycline to express the exogenous NUP98-fusion proteins were cytocentrifuged onto glass slides using a Shandon Cytospin[™] Centrifuge II. Spots were air-dried before storage and staining. Fixing of cells on slides was achieved by applying 4.5 % formaldehyde/histofix (Roth) for 10 minutes at 4 °C. Then cells were permeabilized using 0,2 % Triton X100 in PBS for 10 min. Afterwards spots were covered with 4 μ g/mL primary antibody (Mouse anti-HA (Santa Cruz, sc-7392, clone F-7, RRID: AB_627809)) in 2 % bovine serum albumin/0,2 % Triton X100 in PBS for 1 hour in a wet chamber. Spots were washed 3 times for 5 minutes by rocking in PBS. Subsequently, samples were incubated with the secondary antibody for 1 hour in a dark wet chamber. Alexa Fluor 568

f(ab')₂-goat anti-mouse antibody (Thermo Fisher Scientific, A-21237, RRID: AB_2535806) was applied with a final concentration of 1 µg/mL as secondary antibody. Slides were washed 3 times for 5 minutes by gentle shaking in PBS. 4,6-Diamidino-2-Phenylindole (DAPI, Dilactate, Biolegend, 522801) was applied for 2 minutes to counterstain the DNA. Afterwards spots were washed 3 times for 5 minutes with PBS and then covered with the non-aqueous mounting medium Entellan® New (Merck, 1079610100).

Image acquisition was performed using a Zeiss LSM 880 confocal microscope and Zeiss ZEN black software. Image post-processing was carried out in ImageJ and Zeiss ZEN blue software to adjust brightness, enhance contrast and scale bar insertion.

2.8 LIVE CELL IMAGING

HEK293-T cells were grown on a polystyrol 6-well plate and transfected with 250 ng of plasmid containing the EmGFP-tagged artificial fusion proteins using PEI as transfection reagent. 24-48 hours post-transfection cells were transferred to glass bottom cell culture chambers and grown for another 24 hours. 48-72 hours post-transfection nuclei were stained with 1-5 µg/mL Hoechst 33342 (Thermo Fisher Scientific, H1399) for 7 to 10 minutes and RPMI 1640 medium was applied for imaging. A Zeiss LSM880 confocal microscope (pre-heated to 37°C) with Zeiss ZEN black software was used for acquiring the live images. Z-stacks of the respective fusion proteins were recorded. Laser excitation intensity and digital gain was adjusted on each measurement. Post-processing of confocal z-slice images was accomplished using the ZEN blue software and ImageJ.

2.9 RNA ISOLATION AND PREPARATION FOR RNA-SEQ ANALYSIS

Murine fetal liver cells were retrovirally transduced to constitutively express the fusion proteins and mCherry. After FACS sorting, cells were expanded for seven days, harvested and RNA was isolated using the RNeasy Mini Kit from QIAGEN. Sequencing library preparation was performed using the QUANT-seq 3' RNA-Seq Library Prep Kit (Lexogen) and sequenced with single read 70 bp chemistry at a Illumina NextSeq550 machine.

2.10 BIOINFORMATIC ANALYSIS AND STATISTICS

Sequence quality was checked with FastQC and followed by quality trimming and filtering using PRINSEQ-lite. Remaining reads were mapped against the mouse reference genome (GRCm38) with STAR. SAMtools was used for final processing and featureCounts to acquire the counts per read. Differential gene expression and normalization was executed with DESeq2 and afterwards the resulting data was used for principal component analysis (PCA) and heatmap generation. For PCA the top 500 expressed genes with the highest variance in their normalized expression were included. The heatmap was generated with the heatmap.2 function from the R-package gplots under the usage of Pearson correlation and ward.D for clustering.

3 RESULTS

3.1 VERIFICATION OF EXOGENOUS EXPRESSION OF NUP98-FUSION PROTEINS

For our studies, five NUP98-fusion proteins were selected based on distinct function of their partner genes and their abundance in AML patients. KDM5A and NSD1 are the most frequent NUP98-fusion partners identified in patients. HOXA9 was selected as a representative NUP98-HD fusion partner and PSIP1 and DDX10 were chosen because the molecular mechanisms underlying oncogenic transformation by these fusions are unknown.

Western blot analysis of HEK293-T cells transiently expressing the oncogenic fusion proteins was performed to validate exogenous expression of all five oncogenic fusion-proteins (Figure 5). Untransfected HEK293-T cells were used as negative control (mock). N-NUP98 refers to a construct encompassing the conserved N-terminal part of NUP98-fusion proteins. The different oncogenic fusion proteins were loaded according to their size, starting with NUP98-HOXA9 which is the shortest fusion protein with 59 kDa, followed by NUP98-PSIP1 (63 kDa), NUP98-KDM5A (74 kDa), NUP98-DDX10 (122 kDa) and NUP98-NSD1 with a size of 204 kDa. As the exogenously expressed N-NUP98 and the NUP98-fusion proteins were N-terminally HA-tagged, an antibody recognizing the HA-tag was used for their detection. Actin was used as loading control, showing similar levels of lysate loading. All constructs could be detected and hence were successfully expressed in HEK293-T cells. However, different levels of expression could result from inappropriate blotting settings or could be caused by different expression levels, which could depend on the different sizes of the fusion proteins under study.

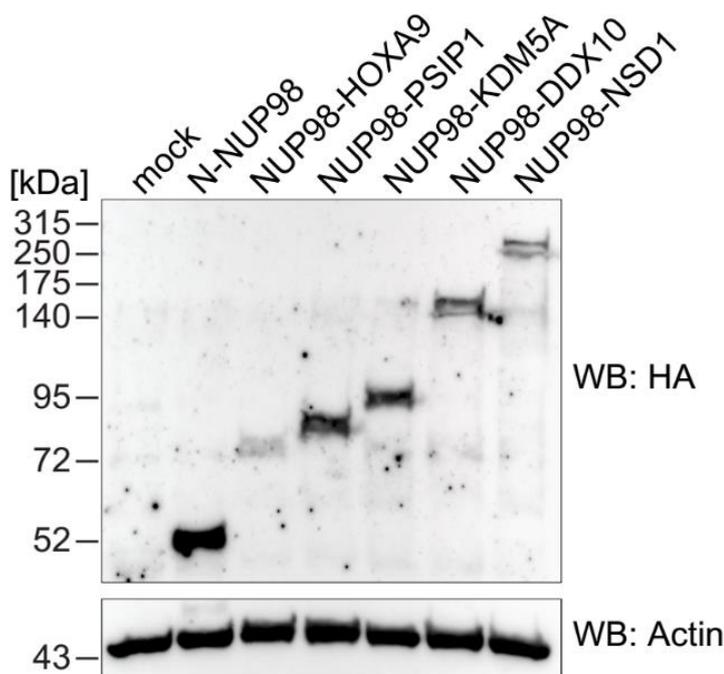


Figure 5: Different expression levels of NUP98-fusion proteins. Western blot analysis for exogenously expressed N-NUP98 and NUP98-fusion proteins using an antibody against the HA-tag (HA) and Actin in lysates of HEK293-T cells 48 hours after transient transfection.

3.2 BIOTINYLATED ISOXAZOLE-MEDIATED PRECIPITATION OF NUP98-FUSION PROTEINS

It has been previously shown that the chemical biotinylated isoxazole (b-isox) enables precipitation of IDR-containing proteins and hence proteins predicted to be involved in biomolecular condensation by the formation of crystals (Han *et al.*, 2012; Kato *et al.*, 2012) (Figure 6, A). IDR-containing proteins such as FUS and NUP98 precipitate from cell lysates following treatment with b-isox (Kato *et al.*, 2012). We used this method to further investigate the biophysical properties of NUP98-fusion proteins. HEK293-T cells were transiently transfected with cDNAs encoding NUP98-fusion proteins, 48 hours post-transfection, cells were collected and lysates were treated with b-isox and analyzed via western blot. Previous results obtained from our group showed dose-dependent precipitation behavior of NUP98-fusion proteins integrating into b-isox precipitates. We further found that known interactors of NUP98-fusion proteins, such as RAE1 could also be detected in b-isox precipitates. However, it is not clear whether RAE1 precipitates as a consequence of its interaction with NUP98-fusion proteins or if the RNA-binding domain of RAE1 causes b-isox-mediated precipitation (Kato *et al.*, 2012). Strong bands in the b-isox fractions were detected for all five NUP98-fusion

proteins using an anti-HA antibody, indicating that NUP98-fusion proteins were sensitive to b-isox precipitation (Figure 6, B). However, for NUP98-HOXA9, NUP98-DDX10 and NUP98-NSD1, a considerable amount of the NUP98-fusion protein remained in the supernatant. RAE1 was detected in all supernatant and b-isox fractions, indicating that it possibly co-precipitated with NUP98-fusion proteins in the presence of b-isox. In contrast to the IDR-containing NUP98 and NUP98-fusion proteins, highly structured proteins such as the heat shock conjugated 71 kDa protein (HSC70) and tubulin were not expected to precipitate upon b-isox treatment and were used as controls. As expected, these proteins were only detected in the supernatant fractions (Figure 6, B), supporting the hypothesis that b-isox only precipitates IDR-containing proteins and RNA-binding proteins such as RAE1 (Kato *et al.*, 2012; Gomes and Shorter, 2019).

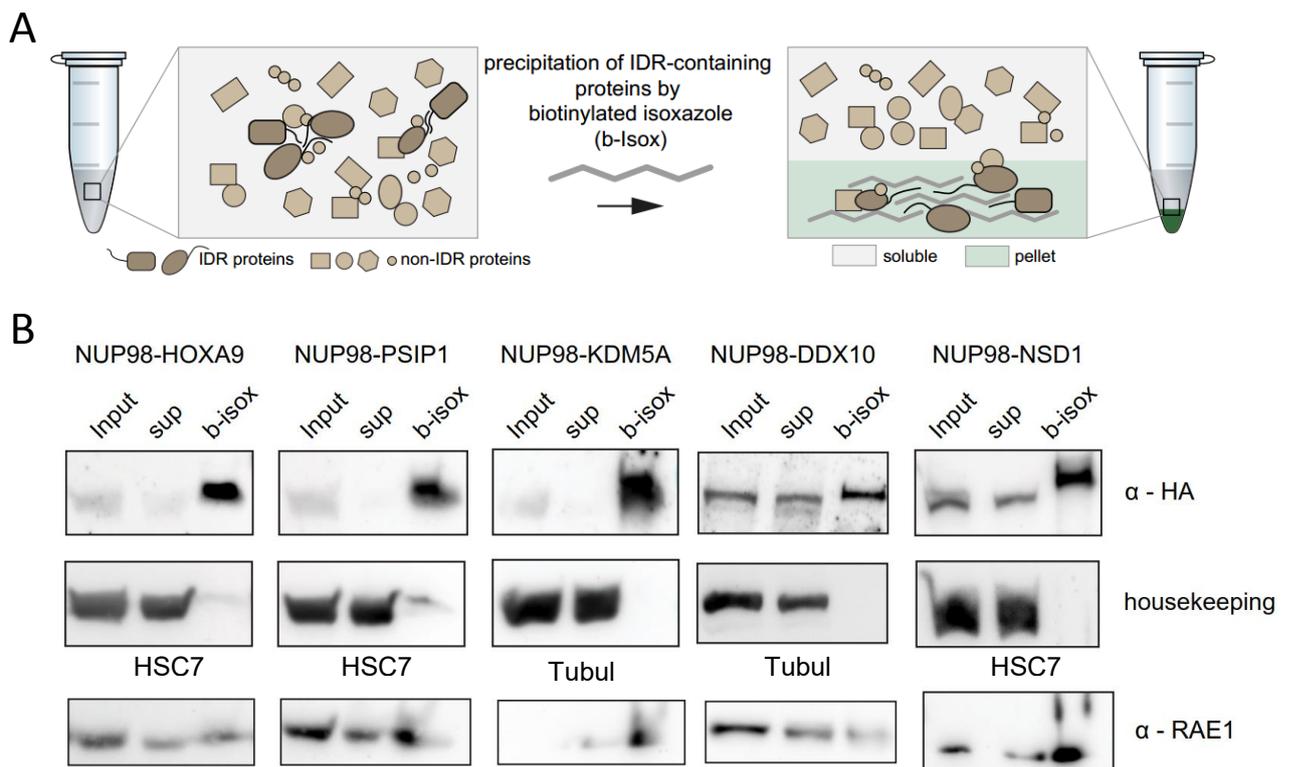


Figure 6: A) Schematic illustration of b-isox-mediated precipitation. B) Western blot analysis of the input lysate, supernatant (sup) and precipitated (b-isox) fraction of HEK293-T cells after 30 min of treatment with 100 μ M b-isox using antibodies against the HA-tag (α -HA), housekeeping genes (HSC70 or Tubulin) and RAE1 (α -RAE1)

3.3 OPTIMIZATION OF STAINING PROTOCOLS AND IMAGE ACQUISITION FOR CONFOCAL IF ANALYSIS OF NUP98-FUSION PROTEINS

To investigate a potential involvement of biomolecular condensation in the function of NUP98-fusion proteins we aimed to analyze the localization of NUP98-fusion proteins by IF microscopy. Samples for IF staining were generated by attaching cells to glass slides using a cytocentrifuge, air-dried and afterwards fixed with formaldehyde, permeabilized using TritonX100 and stained with DAPI (Figure 7). With the standard cytopspin protocol (7 min, 700 rpm) used in the laboratory, we detected leakage of DNA, as shown by DAPI staining outside the nuclear membrane (Figure 8, A) on images captured on a confocal microscope with high resolution. Cytospin settings were optimized for HL-60 cells by trying different times and varying centrifugation speed. Shortening the time of centrifugation from 7 minutes to 3 minutes, both at 700 rpm, strongly decreased the amounts of extranuclear DNA staining (Figure 8, B). Hence, for all following experiments, cytopspin conditions were changed to 3 minutes at 700 rpm. As time and speed were optimized for HL-60 cells, which have a cell diameter of 12 μm and possess a big nucleus compared to the size of the cell, settings would need to be adapted further for other cells of interest dependent on their nuclear-cytoplasm ratio.

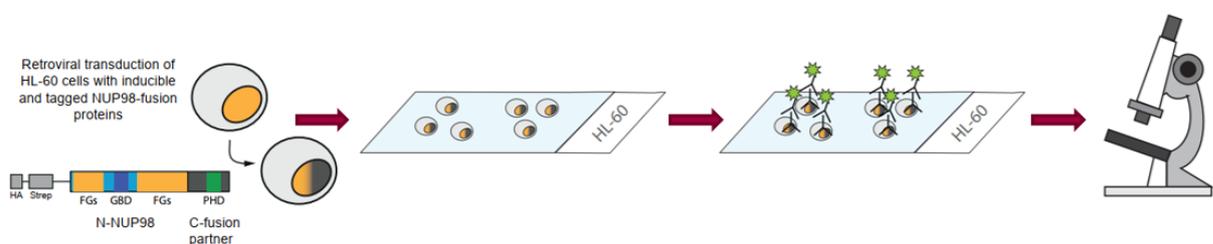


Figure 7: Schematic illustration of the IF approach. HL-60 cells expressing exogenous NUP98-fusion proteins were attached to glass slides using a cytocentrifuge and stained using an anti-HA-antibody.

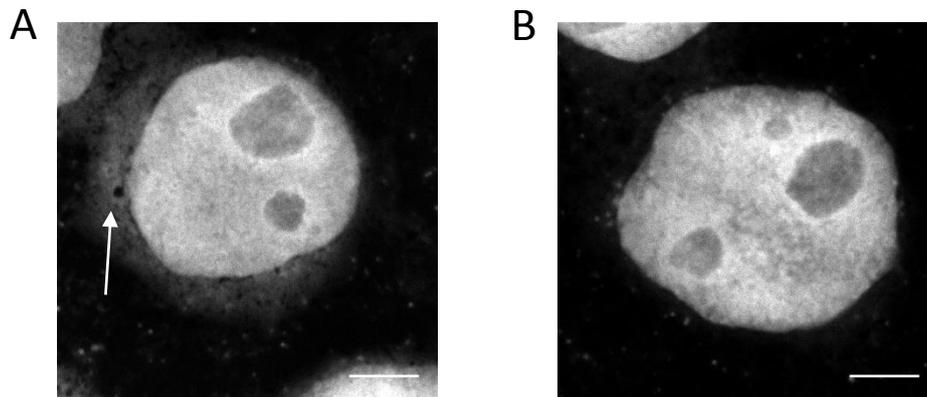


Figure 8: Cytospin setting optimization. A) DNA leakage caused by inappropriate cytospin settings (7 minutes and 700 rpm), indicated by an error; B) Optimized cytospin settings (3 minutes and 700 rpm). The scale bar accounts for 5 μm indicated by a white bar.

Another technical problem that was encountered at 60x magnification was the appearance of crystals within the fixed samples (Figure 9). The crystals appeared also while using transmitted light microscopy. This artifact was dependent on different steps of the protocol including the combination of a PBS washing step before mounting and the usage of non-aqueous mounting media and nail polish as a hydrophobic barrier preventing antibody-spillover. This problem was solved by changing from nail polish to the ImmEdge pen that can be used to establish a thin hydrophobic barrier surrounding the cytospin spots. We suspected that the crystals were composed of salt from PBS and formed through the higher distance between the cover slip and glass slide while using nail polish.

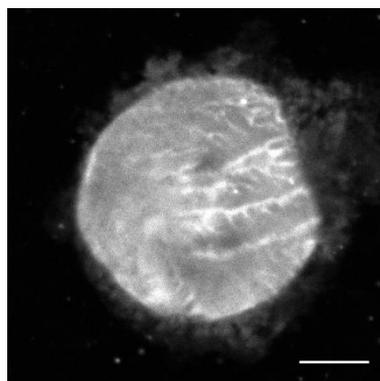


Figure 9: Crystal formation caused by the combination of PBS washing steps previous to mounting with non-aqueous mounting media and nail polish as hydrophobic barrier.

3.4 VISUALIZATION OF NUP98-FUSION PROTEINS VIA CONFOCAL IF MICROSCOPY

To determine the localization of NUP98-fusion proteins within the leukemia cell line HL-60, cells were retrovirally transduced with five different HA-tagged NUP98-fusion proteins in tetracycline-inducible constructs. 24 h after addition of doxycycline, cells were monitored for transgene expression by FACS and subsequently attached to glass slides using the optimized cytopsin protocol. Localization of endogenous NUP98 and other FG-containing NUPs was analyzed by using an antibody that is predicted to bind FXFG-repeats (Anti-FG) and DAPI was used to counterstain the nucleus (Figure 10, A). Fluorescence signal was observed in the nucleus overlapping with the DAPI signal and lining the nuclear envelope. This is in line with the observation of Griffis *et al.*, who showed that FG-containing NUPs localize at the nuclear membrane as well as in the intra-nuclear space (Griffis *et al.*, 2002). In contrast, NUP98-fusion proteins did not co-localize with the nuclear envelope, but showed speckled localization within the nucleus, which was detected using an antibody targeting the HA-tag on the N-terminal part of the fusion proteins. While all five NUP98-fusion proteins localized in characteristic nuclear speckles, their size and number was different between individual NUP98-fusion proteins (Figure 10, B). Interestingly, we also found differences in co-localization of some NUP98-fusion proteins with the nucleolus. While NUP98-PSIP1, NUP98-KDM5A, NUP98-NSD1, NUP98-HOXA9 showed extensive co-localization with DAPI-low regions within the nucleus that are characteristics of the nucleolus, this was not the case for NUP98-DDX10.

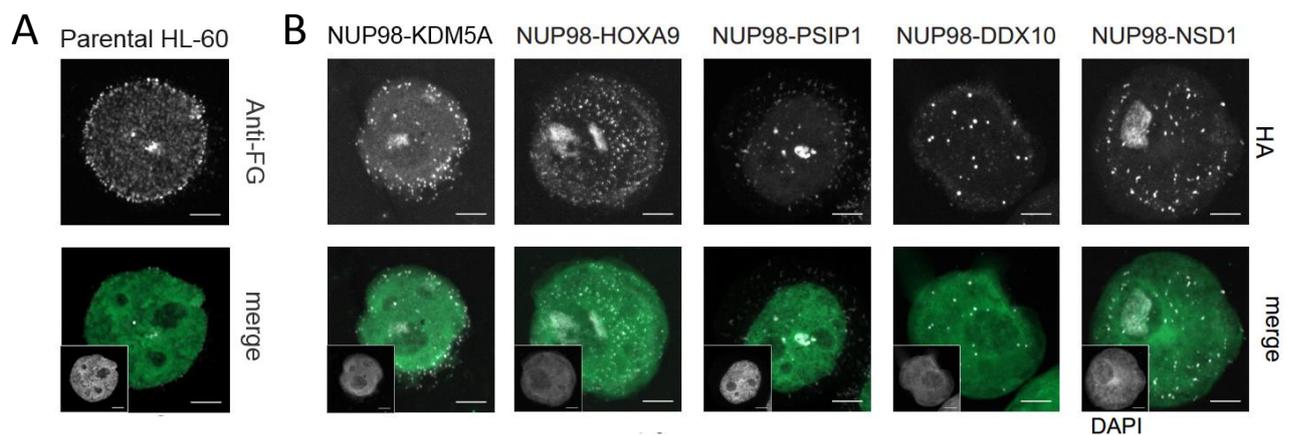


Figure 10: IF imaging of NUP98-fusion proteins expressed in HL-60 cells. A) Untransduced HL-60 cells stained with an antibody binding to FXFG-repeats. B) Localization of NUP8-fusion proteins in the upper row and merged HA/DAPI signals in the lower rows. DAPI staining is shown in the small insets in the lower left corners in the bottom row. Scale bars (5 μm) are indicated. Post-imaging processing as contrast enhancement and scale bar adjustment was performed using ImageJ software.

3.5 ANALYSIS OF THE SUBCELLULAR LOCALIZATION OF DIFFERENT ARTIFICIAL IDR-CONTAINING KDM5A-FUSION PROTEINS BY LIVE-CELL IMAGING

Using an approach combining b-isox and MS, our group has previously found that the expression of NUP98-fusion proteins changes the global composition of biomolecular condensates within the cell. As NUP98-fusion proteins are susceptible to b-isox-precipitation, we hypothesized that they are capable of forming biomolecular condensates via their IDRs. However, it is not known which biophysical properties of the FG-repeats are essential for the formation of biomolecular condensates. To investigate, we replaced the N-terminal NUP98 part of the NUP98-KDM5A fusion protein (containing the FG-repeats) with different artificial repetitive sequences. Mimicking the N-terminus of endogenous NUP98, we designed an artificial FG-KDM5A-fusion protein (artFG-KDM5A) by assembling 13 triple-FG repeats spaced by linkers with a similar amino-acid composition as found in endogenous NUP98, yielding a total of 39 FG-repeats. As controls, two additional artificial fusion proteins were constructed, in which the FG-repeats were replaced by tyrosine-glycine and alanine-alanine repeats (artYG-KDM5A and artAA-KDM5A, respectively). Repeats containing tyrosine instead of phenylalanine were expected to retain the capacity to form biomolecular condensates as the two amino acids only differ by one hydroxyl group and have similar biochemical properties. In contrast, we hypothesized that di-alanine repeats would not promote speckle formation as a consequence of lack of IDRs in the N-terminus of the artificial fusion protein. EmGFP-fusions of all three constructs were prepared to enabling their direct visualization by live cell imaging (Figure 11). Transiently transfected HEK293-T cells were stained with Hoechst 3342 to visualize the nucleus. The Hoechst staining procedure was optimized for concentration and incubation time. A final concentration of 1 μ g/mL Hoechst and 10 minutes incubation in the dark was found to provide the best signal-to-noise ratio without detaching the HEK293-T cells from the glass chamber slide. Cells were imaged 48 to 72 hours after transfection. As hypothesized, artFG-KDM5A displayed a speckled or granular localization pattern across the nucleus (Figure 12, A), supporting its ability to form biomolecular condensates in cells upon overexpression. ArtYG-KDM5A showed more defined and larger speckles (Figure 12, B) than artFG-KDM5A, supporting the hypothesis that artificial proteins containing a sequence similar to FG-repeats and possessing IDRs are able to form biomolecular condensates. In contrast, and as predicted for this artificial construct, the artAA-KDM5A

construct showed homogenous distribution across the nucleus and did not form speckles (Figure 12, C).

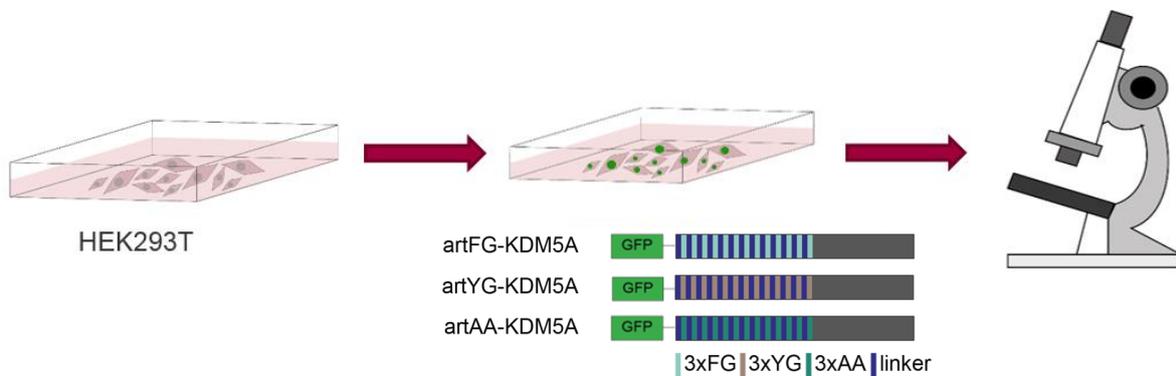


Figure 11: Schematic illustration of live cell imaging approach.

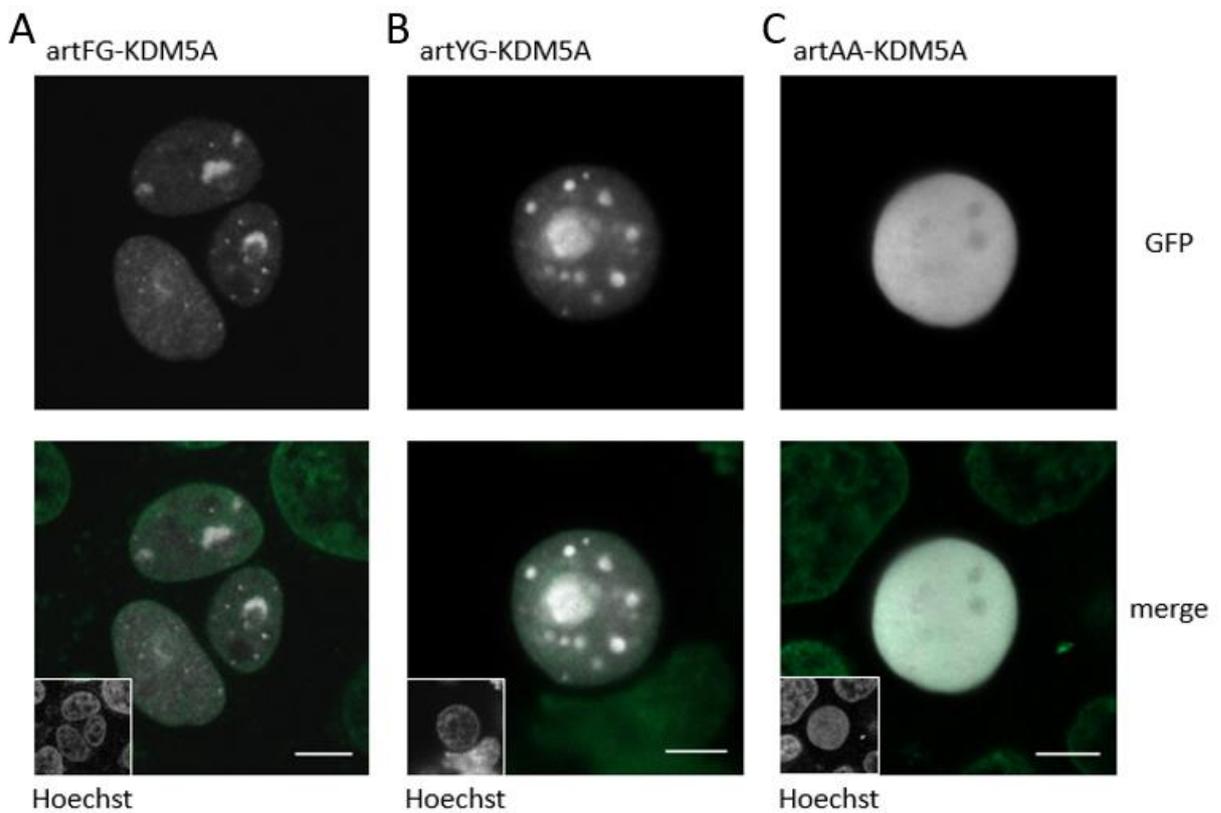


Figure 12: Live cell imaging of IDR-containing fusion proteins artFG-KDM5A, artYG-KDM5A and artAA-KDM5A 72 hours post-transfection. Lower left imaging depicts Hoechst signal, upper channel shows recorded GFP-signal of fusion proteins. Scale bar 5 μ m.

3.6 ANALYSIS OF THE ABILITY OF ARTIFICIAL IDR-CONTAINING KDM5A-FUSION PROTEINS TO DRIVE LEUKEMIA-ASSOCIATED GENE EXPRESSION

Even though we could demonstrate that FG-repeats-containing NUP98-fusion proteins localize to nuclear speckles that are characteristics of biomolecular condensates, it is not known whether the IDRs in the N-terminus of NUP98-fusion proteins and hence the formation of biomolecular condensates is sufficient for the induction of leukemia-associated gene expression. To address this question, we aimed to compare the effects of NUP98-KDM5A to the artificial constructs containing different repeat sequences in their N-terminus (artFG-KDM5A, artYG-KDM5A and artAA-KDM5A) in murine fetal liver cells. Cells were virally transduced with the artificial IDR-containing fusion proteins and NUP98-KDM5A and afterwards RNA was isolated and analyzed by Next Generation Sequencing.

Principal component analysis of RNA-sequencing data revealed that the gene expression profiles of NUP98-KDM5A and artFG-KDM5A were similar (Figure 13). This is in line with the observation that artFG-KDM5A and NUP98-KDM5A showed comparable subnuclear localization patterns (Figure 12). Furthermore, global gene expression in cells expressing artAA-KDM5A, which did not form any speckles, was substantially different from gene expression induced by FG-containing fusion proteins, but very similar to wild type fetal liver cells instead. Surprisingly, global gene expression in cells expressing the artYG-KDM5A construct, which was able to form biomolecular condensates and strongly resembles artFG-KDM5A, was highly similar to untransduced, wild-type murine fetal liver cells (Figure 13), indicating a strong influence of the phenylalanine residue in FG-repeats on the regulation of gene expression.

Similar observations were made at the level of differentially expressed genes. NUP98-KDM5A and artFG-KDM5A exhibited largely overlapping profiles of differentially expressed genes, and the other two artificial constructs and wild type fetal liver cells showed opposing patterns (Figure 14, A).

Amongst the shared genes that were overexpressed upon NUP98-KDM5A and artFG-KDM5A expression are several direct targets of the NUP98-KDM5A oncoprotein (Figure 14, B-E), as previously identified by our group (Schmöllnerl et al 2020, unpublished). Normalized expression

levels of HOXA9, HOXA19 and HOXA7 were lower in wild-type murine fetal liver cells compared to murine fetal liver cells transduced with NUP98-KDM5A or artFG-KDM5A (Figure 14, B-D). Even though levels of CDK6 were mildly increased in artYG-KDM5A and artAA-KDM5A, expression levels were strongly elevated upon NUP98-KDM5A or artFG-KDM5A expression (Figure 14, E). From these data, we conclude that expression of artFG-KDM5A is capable to mimic the leukemogenic gene expression profile induced by NUP98-KDM5A, whereas artYG-KDM5A and artAA-KDM5A are incapable of inducing similar gene expression changes. Further, these results indicate that FG-repeats spaced by NUP98-like linkers coupled to the C-terminus of KDM5A are sufficient to drive NUP98-KDM5A-like gene expression through their biophysical features and might have the capacity to induce leukemia.

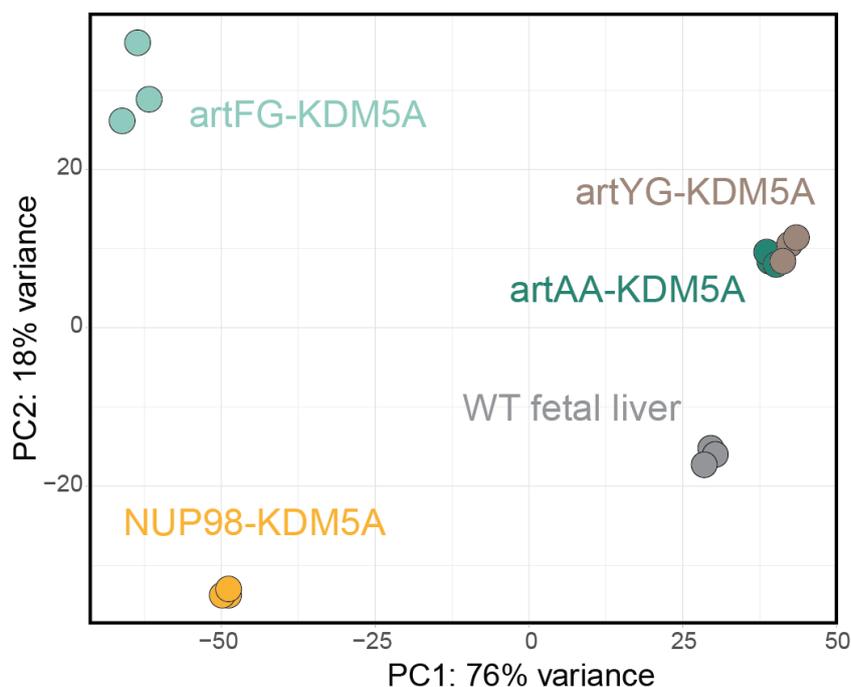


Figure 13: Principal component analysis of individual RNA-sequencing data depicting the top 500 expressed genes

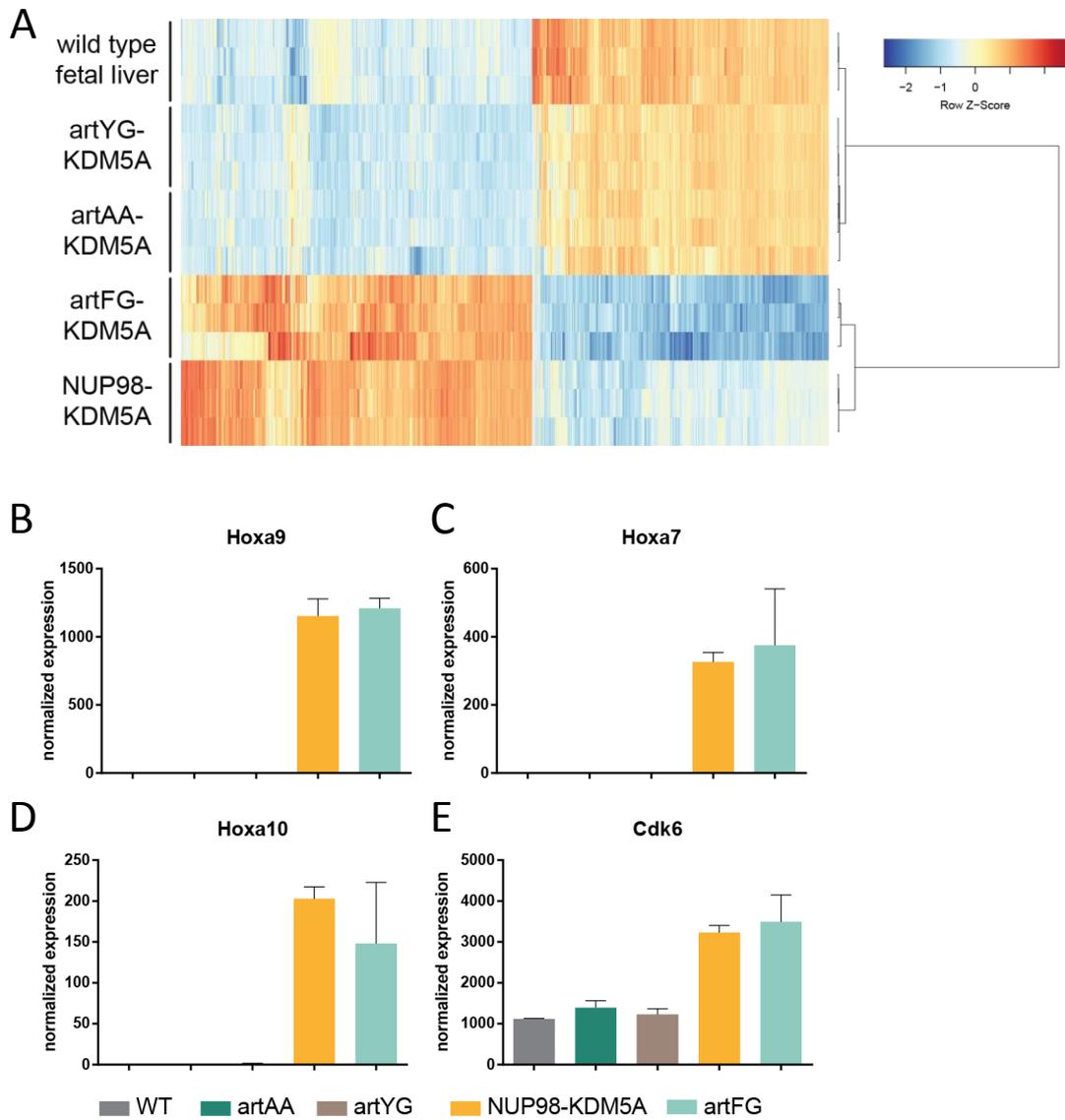


Figure 14: A) Heatmap depicting the significantly differential gene expression in NUP98-KDM5A and artFG-KDM5A-expressing cells compared to murine fetal liver cell expression obtained by RNA-sequencing. B), C), D), E) normalized expression of direct targets of NUP98-fusion proteins.

4 DISCUSSION

AML is the second most common type of leukemia and affects the myeloid compartment of the hematopoietic system. Up to 2 % of AML are caused by NUP98-fusion proteins, where the N-terminal part of NUP98 is fused to more than 25 distinct fusion partner genes (Gough, Slape and Aplan, 2011). So far, no targeted therapy has been developed for this subtype of AML that is associated with very poor prognosis. However, the biological mechanisms underlying NUP98-fusion protein-mediated oncogenic transformation are still poorly understood and it is not known how different NUP98-fusions can lead to a similar leukemia phenotypes. The aim of this work was to characterize the subcellular localization of five structurally different NUP98-fusion proteins, to analyze their capacity to form biomolecular condensates and to investigate their phase separation properties using IF and live-cell imaging approaches. Furthermore, we wanted to determine whether NUP98-fusion protein-driven AML is dependent on FG-repeats in the NUP98-N-terminus and if biomolecular condensation induced by exogenous, artificial FG-repeat sequences is able to phenocopy NUP98-fusion-induced gene expression patterns.

When we performed western blot analysis to investigate the expression of the five distinct NUP98-fusion proteins selected for our studies, we found significant differences in the expression levels of the different constructs (Figure 5). These differences in protein abundance could result from various reasons, including stability of the fusion protein and time of production from initiation of transcription over translation until the correctly folded protein exists. In the case of NUP98-NSD1 and NUP98-DDX10, the lower expression levels might result from increased time that is required for the transcription and translation of these larger fusion proteins compared to the remaining NUP98-fusion proteins, which are smaller. Furthermore, the transfer protocol was not optimized for blotting proteins with high molecular weight. This problem was addressed in the western blot experiment analyzing supernatants and pellet fractions from the b-isox precipitation experiments by applying improved blotting settings. This led to a stronger signal of proteins with higher molecular weights (Figure 6, B). The signals from NUP98-fusion proteins that were detected in the supernatant fractions could possibly result from incomplete b-isox-induced precipitation of the fusion proteins. A higher

amount of b-isox could increase the fraction of precipitated NUP98-DDX10 and NUP98-NSD1 fusion proteins. As expected, there was no signal for HSC70 and tubulin detected in the pellet fraction as these proteins are highly structured and are not expected to be susceptible to b-isox precipitation. This finding supports the theory that b-isox selectively precipitates proteins containing IDRs and proteins harboring specific domains including RNA-recognition motifs or K-homology domains (Kato *et al.*, 2012; Gomes and Shorter, 2019). Our group has previously found that the NUP98-fusion core interactome, as generated by affinity-purification of tagged fusion proteins and subsequent MS analysis, was enriched for proteins that are known to be involved in the formation of biomolecular condensates, such as FUS (Altmeyer *et al.*, 2015; Patel *et al.*, 2015) and HNRNPA1 (Molliex *et al.*, 2015). Furthermore, the RNA-export factor RAE1, which harbors a RNA-binding motif and is predicted to be involved in molecular condensation, is an interaction partner of endogenous NUP98 (Gough, Slape and Aplan, 2011) and NUP98-fusion proteins. Previous data from our laboratory revealed that RAE1 is susceptible to b-isox precipitation in a MS approach. The chemical compound b-isox captures all proteins that are prone to aggregate. Therefore, this assay does not only cause precipitation of the IDR-containing NUP98-fusion proteins, but also all other proteins that are sensitive to this compound, such as the RNA-recognition motif-harboring RAE1. Hence, it is not clear whether b-isox-mediated precipitation of RAE1 occurs from co-precipitation with the NUP98- and NUP98-fusion condensates or if RAE1 precipitates independently of the overexpressed NUP98-fusion transgenes. Additionally, it is so far not known if b-isox is able to precipitate entire protein complexes at once, or only captures proteins that harbor multivalent domains.

As b-isox might not only precipitate biomolecular condensates, additional approaches are required to show that NUP98-fusions proteins are capable of undergoing biomolecular condensation. Confocal microscopy is commonly used to directly visualize biomolecular condensates either through IF imaging or live cell microscopy. In 2012, Fahrenkrog *et al.* and in the following year Xu and Powers found that various NUP98-fusion proteins show speckled patterns of localization in live cell imaging or in IF imaging studies (Xu and Powers, 2013; Fahrenkrog *et al.*, 2016). Based on these data, we aimed to investigate whether all five NUP98-fusion protein (NUP98-HOXA9, NUP98-PSIP1, NUP98-KDM5A, NUP98-DDX10 and NUP98-NSD1) are able to localize to similar nuclear structures despite their distinct protein

structures. Additionally, by using the human leukemia cell line HL-60 we wanted to investigate the subcellular localization of NUP98-fusion protein in a physiologically relevant cell context, which is in contrast to previous reports that used HeLa cells as biological models (Xu and Powers, 2013; Fahrenkrog *et al.*, 2016). Optimization of the staining protocol to detect HA-tagged NUP98-fusion proteins was required, as the first images acquired at a Zeiss LSM 880 confocal microscope with a 60x immersion oil objective revealed extra-nuclear DAPI staining resulting from suboptimal cytospin settings. For visualization of endogenous NUP98, an antibody predicted to detect FXFG-repeats, which are also found in other FG-containing NUPs, was used. This antibody revealed a signal at the nuclear envelope as well as in the nucleoplasm. This localization behavior was already shown for GFP-tagged NUP98 in HeLa cells, resulting in the term “GLFG-bodies” for the specific localization pattern of intra-nuclear NUP98 (Griffis *et al.*, 2002).

As formation of these speckled structures or GLFG bodies was regarded an artifact from high overexpression of proteins, we reduce the amount of transfected DNA to 250 ng per six-well to limit the possibility of artificial accumulation of aggregates. Similar amounts of DNA for the transfection of NUP-fusion proteins were used in Fahrenkrog *et al.* (2016) and in Xu and Powers (2013). Another approach to prove that biomolecular condensation is not an artifact of protein overexpression would be the analysis of primary patient samples harboring NUP98 translocations. Unfortunately however, patient samples do not allow unambiguous visualization of NUP98-fusion proteins. For our experiments, we used an anti-HA antibody to specifically visualize different NUP98-fusion proteins. When using the anti-FXFG-repeat antibody on cells expressing exogenous NUP98-fusion proteins, we observed a speckled pattern similar to HL-60 cells that do not express NUP98-fusion proteins (data not shown). This similarity might result from the binding of the monoclonal antibody the N-terminus of NUP98, which is retained in NUP98-fusion proteins, as well as to several other FG-containing NUPs. Therefore, in this setting, it is not possible to distinguish between endogenous NUP98 and exogenous NUP98-fusion proteins by imaging approaches, as no NUP98-fusion protein-specific antibody has been developed so far.

A different method to visualize the exact localization and behavior of exogenously expressed transgenes is live cell imaging. This approach requires the protein of interest to be coupled to a

fluorescent reporter protein. However, this technique cannot be applied to patient samples, as patients do not express fluorescently tagged NUP98-fusion proteins. Therefore, a possible way to visualize NUP98-fusion proteins within primary patient samples would be by tagging the endogenous fusion proteins with a fluorescent protein using CRISPR/Cas9-mediated genome engineering. This approach would enable co-localization studies involving other proteins harboring a fluorescence tag or stained with a specific antibody.

Our group and other groups found that RAE1 interacts with endogenous NUP98. Furthermore, MS data from previous experiments in the laboratory revealed that RAE1 interacts with different NUP98-fusion proteins. When visualizing RAE1 by IF microscopy, we would expect to detect signals at the nuclear envelope and in NUP98-fusion protein-containing speckles. Another interactor that shows specific binding to NUP98-fusion proteins, such as the transcriptional regulator FUS, would be predicted to co-localize with the oncogenic fusion proteins in biomolecular condensates. Likewise, more detailed investigation of other NUP98-fusion protein-specific interaction partners involved in the regulation of gene expression and associated with biomolecular condensation could deepen our knowledge about the functions of NUP98-fusion proteins.

Another question we addressed in our study was whether NUP98-fusion driven AML is dependent on FG-repeats or if biomolecular condensation via different IDRs in the N-terminal part of the NUP98-fusion protein would be sufficient to drive leukemic gene expression. In 2013, Xu and Powers showed that the number of FG-repeats in NUP-fusion proteins influences their intra-nuclear localization. Reduced numbers of FG-repeats led to loss of speckled localization and more homogenous intra-nuclear distribution of the fusion proteins (Xu and Powers, 2013). However, both the N-terminal part of a NUP98 fusion as well as expression of a C-terminal fusion partner alone was insufficient to drive AML (Yung *et al.*, 2011). To examine the importance of the FG-repeats in NUP98-fusion proteins, artificial IDRs consisting of 39 di-amino acid-repeats (FG, YG or AA), were fused to the C-terminus of KDM5A. As expected from their biochemical properties, artFG-KDM5A and artYG-KDM5A fusion proteins exhibited a speckled nuclear localization pattern, whereas the artAA-KDM5A fusion showed a homogenous distribution across the nucleus. Cells expressing the artFG-KDM5A fusion displayed a strongly structured localization pattern of the fusion and the observed

structures were generally smaller than speckles detected in artYG-KDM5A expressing cells. Potential reasons for these differences could include the time of measurement after transfection and cell cycle stage of the recorded cell. The latter could be prevented by synchronizing cells by treatment with thymidine, which blocks cell cycle at the G1/S transition, as was done by Fahrenkrog *et al.* (Fahrenkrog *et al.*, 2016). Another explanation for the different behavior of phenylalanine-containing and tyrosine-containing repeats might be altered post-translational modification patterns and/or distinct protein conformations that are influenced by the different properties of the amino acid repeats.

Further studies of the behavior of the observed NUP98-fusion-containing speckles needs to be performed. To support the theory that these structures are indeed biomolecular condensates, FRAP experiments could be performed. The FRAP technique requires photobleaching of regions of interest with a high-intensity laser. The time required for the recovery of the fluorescence signal within this region is recorded and provides further information about the dynamic behavior of interactions and liquidity of the structures (Gray and Price, 2018).

Higher levels of roundness (as a representation of surface tension) and dynamic fusion of the condensates would indicate the fluid-like nature of the observed structures in a qualitative way. Quantitative results such as fast fluorescence recovery may reinforce LLPS as the driving mechanism. However, the potential caveats of these techniques have to be considered, as not only liquid-liquid phase transition can occur but also gel-like and liquid-crystalline phase transitions were described (McSwiggen *et al.*, 2019). As stated in 2019 by McSwiggen *et al.*, the diffraction limit, which leads to the appearance of round structures if their size is below or close to the resolution of the microscope, needs to be considered a false-positive qualitative result. Alternatively, several reports have shown that biomolecular condensates can be disrupted by treating cells with 1,6-hexanediol (McSwiggen *et al.*, 2019) or lipoic acid (Wheeler *et al.*, 2019). However, 1,6-hexanediol is not specific to the disruption of biomolecular condensates formed by LLPS, as it induces enhanced membrane permeability and disrupts all compartments that are formed by weak interactions. In fact, it was recently proposed that the term “hub” would be more appropriate than “biomolecular condensates”, which are suspected to be formed by LLPS, as there is a lack of evidence and well-established quantitative methods for making explicit conclusions about their biophysical and biochemical nature. Suitable

quantitative approaches for analyzing phase separation mechanisms are super-resolution microscopy (McSwiggen *et al.*, 2019) and fluorescence time-lapse microscopy, which allows live recording of this mechanism (Alberti, Gladfelter and Mittag, 2019).

Many proteins have been shown to undergo phase separation *in vitro*. These experiments are usually performed under idealized conditions regarding temperature, pH and concentration of co-solutes, which can be very different from the physiological state within a cell. In overexpression experiments the same proteins might form condensates *in vivo*. However, the separation of these proteins by phase at high concentrations within a cell does not allow to draw any conclusion about the functional relevance of this mechanism within the cell (Alberti, Gladfelter and Mittag, 2019). In general, however, biomolecular condensation represents a biochemical mechanism that can lead to higher spatiotemporal concentration of molecules within the membraneless organelle and to sequestration of molecules from the surrounding environment.

The artFG-KDM5A transgene was able to mimic NUP98-KDM5A leukemia-specific gene expression pattern, including the expression of direct targets of NUP98-fusion proteins (Figure 14). The gene expression pattern induced by artAA-KDM5A resembled the wild type fetal liver cells. These results indicate a dependency of NUP98-fusion proteins on the FG-repeats in their N-terminal part. While artYG-KDM5A was able to form condensates, it was incapable of evoking the NUP98-KDM5A-leukemia-specific gene expression changes. The dependency on FG-repeats within the fusion proteins might elucidate the biological mechanism for the development of similar disease phenotypes caused by the NUP98-MPTF, which features more than 25 structurally diverse NUP98-fusion proteins. Therefore, the unique combination of the N-terminal part of NUP98, which contains the FG-repeats that lead to biomolecular condensation, and a C-terminal partner gene that has gene regulatory functions might be required for the leukemogenicity of NUP98-fusion proteins.

In summary, this thesis shows that five structurally distinct NUP98-fusion proteins show molecular features that are compatible with biomolecular condensation. Further steps in this project could be a global analysis of the proteins in biomolecular condensates composed of the artificial IDR-containing fusion proteins and the determination of specific interactors for artFG-

KDM5A that differ from artYG-KDM5A. Additionally, the biophysical properties of the condensates need to be investigated using more specific imaging approaches, such as FRAP. To substantiate the hypothesis that artFG-KDM5A mimics the leukemogenic behavior of NUP98-KDM5A, *in vivo* experiments should be performed. Thus, this study extends our understanding of NUP98-fusion proteins regarding their oncogenic mechanisms and the involvement of phase separation in NUP98-fusion protein-driven AML.

ABSTRACT

Leukemia is a heterogeneous hematopoietic disorder. Acute myeloid leukemia (AML) is a particularly aggressive and common type of this cancer that affects the myeloid compartment of the hematopoietic system. Up to 2 % of AML cases are caused by Nucleoporin 98 (NUP98)-fusion proteins and affected cases are associated with particularly poor prognosis. More than 25 distinct chromosomal translocations involving the *NUP98* gene have been described. All NUP98-fusion proteins share the N-terminal part of endogenous NUP98, which consists of two intrinsically disordered regions harboring phenylalanine-glycine (FG)-repeats. Intrinsically disordered regions are highly flexible protein modules that lack a defined secondary structure. Furthermore, intrinsically disordered region-containing proteins have a higher tendency to form membraneless organelles through biomolecular condensation. The N-terminal part of *NUP98* is linked to the C-terminal part of several fusion partner genes, of which most have functions in epigenetic regulation and gene expression. The aim of this thesis was the investigation of the subcellular localization of NUP98-fusion proteins and their potential involvement in biomolecular condensates. Using IF and confocal microscopy we found that five structurally distinct NUP98-fusion proteins do not co-localize with the nuclear envelope but showed their capability of forming biomolecular condensates within the nucleus. In further experiments, three different GFP-tagged artificial IDR-containing fusion proteins were designed on the basis of NUP98-KDM5A. This artificial N-terminal part contained either FG, tyrosine-glycine (YG) or alanine-alanine (AA)-repeats. The influence of the N-terminal part and their ability to form biomolecular condensates was analyzed via live cell imaging. The artificial AA-KDM5A fusion protein showed a homogeneous signal within the nucleus, whereas artificial YG-KDM5A and artificial FG-KDM5A fusions were expressed in speckled patterns consistent with biomolecular condensation. RNA sequencing of fetal liver cells expressing the artificial fusion proteins showed overlapping gene expression profiles induced by NUP98-KDM5A and the artificial FG-KDM5A fusion, which were not shared by cells expressing the artificial AA-KDM5A, or artificial YG-KDM5A fusions, supporting the conclusion that artificial FG-KDM5A is capable of inducing the leukemia-specific gene expression pattern of NUP98-KDM5A. This work deepens our knowledge of NUP98-fusion proteins and their involvement in phase separation and the dependency of FG-repeats in NUP98-fusion protein-driven AML.

ZUSAMMENFASSUNG

Akute myeloische Leukämie (AML) ist eine besonders aggressive und häufige Krebsart, die das myeloische Kompartiment des hämatopoetischen Systems betrifft. Bis zu 2% der AML-Fälle werden durch Nucleoporin 98 (NUP98)-Fusionsproteine verursacht. Betroffene Fälle gehen mit einer schlechten Prognose einher und enden unbehandelt tödlich. Es wurden mehr als 25 verschiedene chromosomale Translokationen mit Beteiligung des NUP98-Gens beschrieben. Alle NUP98-Fusionsproteine beinhalten den N-terminalen Teil des endogenen NUP98, der aus zwei ungeordneten Regionen (IDRs) besteht, die Phenylalanin-Glyzin (FG)-Aminosäurewiederholungen enthalten. IDRs sind hochflexible Proteinmodule, denen eine definierte Sekundärstruktur fehlt. Darüber hinaus neigen IDR-enthaltende Proteine stärker dazu, durch biomolekulare Kondensation membranlose Organellen zu bilden. Der N-terminale Teil von NUP98 ist mit dem C-terminalen Teil mehrerer Fusionspartnergenen verbunden, von denen die meisten Funktionen in der epigenetischen Regulation und der Genexpression übernehmen. Ziel dieser Arbeit war die Untersuchung der subzellulären Lokalisation von NUP98-Fusionsproteinen und ihrer möglichen Beteiligung an biomolekularen Kondensaten. Unter Verwendung von Immunfluoreszenz und konfokaler Mikroskopie fanden wir, dass fünf strukturell unterschiedliche NUP98-Fusionsproteine nicht mit der Kernhülle kolokalisieren, sondern ihre Fähigkeit zur Bildung von biomolekularen Kondensaten im Kern zeigen. In weiteren Experimenten wurden drei verschiedene GFP-markierte künstliche IDR-enthaltenden Fusionsproteine auf der Basis von NUP98-KDM5A entworfen. Dieser künstliche N-terminale Teil enthielt entweder FG-, Tyrosin-Glyzin (YG)- oder Alanin-Alanin (AA)-Aminosäurewiederholungen. Der Einfluss des N-terminalen Teils und dessen Fähigkeit zur Bildung von biomolekularen Kondensaten wurde mittels Lebendzellmikroskopie analysiert. Das AA-KDM5A-Fusionsprotein zeigte ein homogenes Signal im Kern, während YG-KDM5A- und FG-KDM5A-Fusionen gesprenkeltes Muster aufwiesen, die charakteristisch für biomolekulare Kondensation sind. RNA-Sequenzierungsexperimente von fötalen Leberzellen, welche die künstliche Fusionsproteine exprimierten, zeigten überlappende Genexpressionsprofile, die durch NUP98-KDM5A und die FG-KDM5A-Fusion induziert wurden. Jedoch wurde dieses Genexpressionsmuster nicht von Zellen geteilt, die AA-KDM5A oder YG-KDM5A-Fusionen exprimierten. Dies lässt die

Schlussfolgerung zu, dass FG-KDM5A in der Lage ist, Leukämie-spezifische Genexpressionsmuster von NUP98-KDM5A hervorzurufen. Diese Arbeit vertieft unser Wissen über NUP98-Fusionsproteine hinsichtlich der Beteiligung der biomolekularen Kondensatbildung und der Abhängigkeit von FG-Wiederholungen bei NUP98-Fusionsprotein-gesteuerter AML.

LIST OF FIGURES

Figure 1: Schematic representation of healthy hematopoiesis	1
Figure 2: Hematopoiesis in a person with leukemia.....	2
Figure 3: Schematic drawing of the NPC.....	9
Figure 4: Schematic representation of the NUP98 proteina	10
Figure 5: Different expression levels of NUP98-fusion proteins	21
Figure 6: NUP98-fusion proteins precipitate upon b-isox treatment	22
Figure 7: Schematic illustration of the IF approach	23
Figure 8: DNA leakage caused by inappropriate cytopsin settings.....	24
Figure 9: Crystal formation	24
Figure 10: IF imaging of NUP98-fusion proteins	25
Figure 11: Schematic illustration of live cell imaging approach.....	27
Figure 12: Live cell imaging of IDR-containing fusion proteins	27
Figure 13: Principal component analysis of individual RNA-sequencing data.....	29
Figure 14:Heatmap and normalized expression of direct targets of NUP98-fusion proteins...	30

LIST OF TABLES

Table 1: Components for SDS-PAGE gel 16

REFERENCES

- Aguzzi, A. and Altmeyer, M. (2016) 'Phase Separation: Linking Cellular Compartmentalization to Disease', *Trends in Cell Biology*. Elsevier Ltd, 26(7), pp. 547–558. doi: 10.1016/j.tcb.2016.03.004.
- Alberti, S. (2017) 'Phase separation in biology', *Current Biology*. Elsevier, 27(20), pp. R1097–R1102. doi: 10.1016/j.cub.2017.08.069.
- Alberti, S., Gladfelter, A. and Mittag, T. (2019) 'Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates', *Cell*. Elsevier Inc., 176(3), pp. 419–434. doi: 10.1016/j.cell.2018.12.035.
- Altmeyer, M. *et al.* (2015) 'Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose)', *Nature Communications*, 6(8088). doi: 10.1038/ncomms9088.
- Baker, D. J. *et al.* (2007) 'Mitotic regulation of the anaphase-promoting complex', *Cellular and Molecular Life Science*, 64, pp. 589–600. doi: 10.1007/s00018-007-6443-1.
- Banani, S. F. *et al.* (2017) 'Biomolecular condensates: Organizers of cellular biochemistry', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 18(5), pp. 285–298. doi: 10.1038/nrm.2017.7.
- Baron, J. and Wang, E. S. (2019) 'Gemtuzumab Ozogamicin for Treatment of Acute Myeloid Leukemia', *Expert Review of Clinical Pharmacology*, 11(6), pp. 549–559. doi: 10.1080/17512433.2018.1478725.Gemtuzumab.
- Beck, M. and Hurt, E. (2017) 'The nuclear pore complex: Understanding its function through structural insight', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 18(2), pp. 73–89. doi: 10.1038/nrm.2016.147.
- Blevins, M. B. *et al.* (2003) 'Complex formation among the RNA export proteins Nup98, Rae1/Gle2, and TAP', *Journal of Biological Chemistry*, 278(23), pp. 20979–20988. doi: 10.1074/jbc.M302061200.

- Boija, A. *et al.* (2018) ‘Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains’, *Cell*. Elsevier Inc., 175(7), pp. 1842–1855. doi: 10.1016/j.cell.2018.10.042.
- Bolouri, H. *et al.* (2018) ‘The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions’, *Nature Medicine*, 24(1), pp. 103–112. doi: 10.1016/j.physbeh.2017.03.040.
- Brangwynne, C. P., Mitchison, T. J. and Hyman, A. A. (2011) ‘Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes’, *PNAS*, 108(11), pp. 4334–4339. doi: 10.1073/pnas.1017150108.
- Brehme, M. *et al.* (2009) ‘Charting the molecular network of the drug target Bcr-Abl’, *PNAS*, 106(18).
- Brien, G. L., Kimberly, S. and Armstrong, S. A. (2019) ‘Targeting chromatin complexes in fusion protein-driven malignancies’, *Nature Reviews Cancer*. Springer US, 19, pp. 255–269. doi: 10.1038/s41568-019-0132-x.
- Burnett, A., Wetzler, M. and Löwenberg, B. (2014) ‘Therapeutic Advances in Acute Myeloid Leukemia’, *JOURNAL OF CLINICAL ONCOLOGY REVIEW*, 29(5), pp. 487–494. doi: 10.1200/JCO.2010.30.1820.
- Chopra, M. and Bohlander, S. K. (2019) ‘The cell of origin and the leukemia stem cell in acute myeloid leukemia’, *Genes Chromosomes and Cancer*, 58(12), pp. 850–858. doi: 10.1002/gcc.22805.
- Döhner, H. *et al.* (2017) ‘Global Acute Myeloid Leukemia Epidemiology and Patient Flow Analysis 2016’, *Blood*, 129(4), pp. 424–448. doi: 10.1182/blood-2016-08-733196.424.
- Döhner, H., Weisdorf, D. and Bloomfield, C. (2015) ‘Acute Myeloid Leukemia’, *New England Journal of Medicine*, pp. 1136–1152. doi: 10.1056/NEJMra1406184.
- Fahrenkrog, B. (2014) ‘Nucleoporin Gene Fusions and Hematopoietic Malignancies’, *New Journal of Science*, 2014, pp. 1–18. doi: 10.1155/2014/468306.

- Fahrenkrog, B. *et al.* (2016) 'Expression of leukemia-associated Nup98 fusion proteins generates an aberrant nuclear envelope phenotype', *PLoS ONE*, 11(3), pp. 1–24. doi: 10.1371/journal.pone.0152321.
- Franks, T. M. and Hetzer, M. W. (2013) 'The role of Nup98 in transcription regulation in healthy and diseased cells', *Trends Cell Biology*, 23(3), pp. 112–117. doi: 10.1038/jid.2014.371.
- Funasaka, T. *et al.* (2011) 'RNA export factor RAE1 contributes to NUP98-HOXA9-mediated leukemogenesis', *Cell Cycle*, 10(9), pp. 1456–1467. doi: 10.4161/cc.10.9.15494.
- Gerritsen, M. *et al.* (2019) 'RUNX1 mutations enhance self-renewal and block granulocytic differentiation in human in vitro models and primary AMLs', *Blood Advances*, 3(3), pp. 320–332. doi: 10.1182/bloodadvances.2018024422.
- Gomes, E. and Shorter, J. (2019) 'The molecular language of membraneless organelles', *Journal of Biological Chemistry*, 294(18), pp. 7115–7127. doi: 10.1074/jbc.TM118.001192.
- Gough, S. M., Slape, C. I. and Aplan, P. D. (2011) 'NUP98 gene fusions and hematopoietic malignancies: Common themes and new biologic insights', *Blood*, 118(24), pp. 6247–6257. doi: 10.1182/blood-2011-07-328880.
- Gray, (Jay) Jerome W. and Price, R. L. (2018) *More Basic Confocal Microscopy: A Tutorial*, Springer Nature Switzerland AG. doi: 10.1017/s1431927603447843.
- Griffis, Eric R. *et al.* (2002) 'Nup98 Is a Mobile Nucleoporin with Transcription- dependent Dynamics', *Molecular Biology of the Cell*, 13. doi: 10.1091/mbc.01.
- Griffis, Eric R *et al.* (2002) 'Nup98 is a mobile nucleoporin with transcription-dependent dynamics.', *Molecular biology of the cell*, 13(4), pp. 1282–97. doi: 10.1091/mbc.01-11-0538.
- Han, T. W. *et al.* (2012) 'Cell-free Formation of RNA Granules: Bound RNAs Identify Features and Components of Cellular Assemblies', *Cell*. Cell Press, 149(4), pp. 768–779. doi: 10.1016/J.CELL.2012.04.016.

- Hyman, A. A., Weber, C. A. and Jülicher, F. (2014) 'Liquid-Liquid Phase Separation in Biology', *Annual Review of Cell and Developmental Biology*, 30(1), pp. 39–58. doi: 10.1146/annurev-cellbio-100913-013325.
- Iakoucheva, L. M. *et al.* (2002) 'Intrinsic disorder in cell-signaling and cancer-associated proteins', *Journal of Molecular Biology*, 323(3), pp. 573–584. doi: 10.1016/S0022-2836(02)00969-5.
- Ibarra, A. and Hetzer, M. W. (2015) 'Nuclear pore proteins and the control of genome functions', *Genes & Developments*, 29, pp. 337–349. doi: 10.1101/gad.256495.114.Cytoplasmic.
- Jagannathan-Bogdan, M. and Zon, L. I. (2013) 'Hematopoiesis', *Development*, 140(12), pp. 2463–2467. doi: 10.1242/dev.083147.
- Juliusson, G. *et al.* (2009) 'Age and acute myeloid leukemia : real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry', *Clinical Trials and Observations*, 113(18), pp. 4179–4187. doi: 10.1182/blood-2008-07-172007.An.
- Kalverda, B. *et al.* (2010) 'Nucleoporins Directly Stimulate Expression of Developmental and Cell-Cycle Genes Inside the Nucleoplasm', *Cell*. Elsevier Ltd, 140(3), pp. 360–371. doi: 10.1016/j.cell.2010.01.011.
- Kasper, L. H. *et al.* (1999) 'CREB Binding Protein Interacts with Nucleoporin-Specific FG Repeats That Activate Transcription and Mediate NUP98-HOXA9 Oncogenicity', *Molecular and Cellular Biology*, 19(1), pp. 764–776. doi: 10.1128/mcb.19.1.764.
- Kato, M. *et al.* (2012) 'Cell-Free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers Within Hydrogels', *Cell*, 149(4), pp. 753–767. doi: 10.1016/j.cell.2012.04.017.Cell-Free.
- Köhler, A. and Hurt, E. (2010) 'Gene Regulation by Nucleoporins and Links to Cancer', *Molecular Cell*, 38(1), pp. 6–15. doi: 10.1016/j.molcel.2010.01.040.
- Kroon, E. *et al.* (2001) 'NUP98-HOXA9 expression in hemopoietic stem cells induces chronic

and acute myeloid leukemias in mice’, *EMBO Journal*, 20(3), pp. 350–361. doi: 10.1093/emboj/20.3.350.

Lancaster, A. K. *et al.* (2014) ‘PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition’, *Bioinformatics*, 30(17), pp. 2501–2502. doi: 10.1093/bioinformatics/btu310.

Lapidot, T. *et al.* (1994) ‘A cell initiating human acute myeloid leukaemia after transplantation into SCID mice’, *Nature*, 367.

Lebert-Ghali, C. E. *et al.* (2016) ‘Hoxa cluster genes determine the proliferative activity of adult mouse hematopoietic stem and progenitor cells’, *Blood*, 127(1), pp. 87–90. doi: 10.1182/blood-2015-02-626390.

Liang, Y. *et al.* (2013) ‘Dynamic Association of NUP98 with the Human Genome’, *PLOS Genetics*, 9(2). doi: 10.1371/journal.pgen.1003308.

Liu, N., Zhang, J. and Ji, C. (2013) ‘The emerging roles of Notch signaling in leukemia and stem cells’, *Biomarker Research*, 1(23). doi: 10.1186/2050-7771-1-23.

Löwenberg, B., Downing, J. and Burnett, A. (2012) ‘Acute Myeloid Leukemia’, *The New England Journal of Medicine*, pp. 1051–1062. doi: 10.1007/978-3-319-45343-9_5.

Marschalek, R. (2016) ‘Classification of mixed-lineage leukemia fusion partners predicts additional cancer pathways’, *Annals of Laboratory Medicine*, 36(2), pp. 85–100. doi: 10.3343/alm.2016.36.2.85.

McNeer, N. A. *et al.* (2019) ‘Genetic mechanisms of primary chemotherapy resistance in pediatric acute myeloid leukemia’, *Leukemia*. Springer US, 33, pp. 1934–1943. doi: 10.1038/s41375-019-0402-3.

McSwiggen, D. T. *et al.* (2019) ‘Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences’, *Genes & Development*, 33, pp. 1–16. doi: 10.1101/gad.331520.119.

Mertens, F. *et al.* (2015) 'The emerging complexity of gene fusions in cancer', *Nature Reviews Cancer*. Nature Publishing Group, 15(6), pp. 371–381. doi: 10.1038/nrc3947.

Mészáros, B., Erdős, G. and Dosztányi, Z. (2018) 'IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding', *Nucleic Acids Research*, 46, pp. 329–337. doi: 10.1093/nar/gky384.

Mitelman, F., Johansson, B. and Mertens, F. (2007) 'The impact of translocations and gene fusions on cancer causation', *Nature Reviews Cancer*, 7(4), pp. 233–245. doi: 10.1038/nrc2091.

Mitreá, D. M. and Kriwacki, R. W. (2016) 'Phase separation in biology; Functional organization of a higher order Short linear motifs - The unexplored frontier of the eukaryotic proteome', *Cell Communication and Signaling*. Cell Communication and Signaling, 14(1), pp. 1–20. doi: 10.1186/s12964-015-0125-7.

Molliex, A. *et al.* (2015) 'Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization', *Cell*. Cell Press, 163(1), pp. 123–133. doi: 10.1016/J.CELL.2015.09.015.

Navarro, M. G.-J. *et al.* (2019) 'RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates', *Nature Communications*. Springer US, 10(1), pp. 1–13. doi: 10.1038/s41467-019-11241-6.

Oran, B. and Weisdorf, D. J. (2012) 'Survival for older patients with acute myeloid leukemia: A population-based study', *Haematologica*, 97(12), pp. 1916–1924. doi: 10.3324/haematol.2012.066100.

Patel, A. *et al.* (2015) 'A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation', *Cell*. Elsevier Inc., 162(5), pp. 1066–1077. doi: 10.1016/j.cell.2015.07.047.

Pavlović, S. and Tosić, N. (2019) *Genetic and Epigenetic Profiling in Personalized Medicine: Advances in Treatment of Acute Myeloid Leukemia*. Springer Nature Switzerland AG 2019. doi: 10.1007/978-3-030-16465-2_28.

- De Propriis, M. S. *et al.* (2011) 'High CD33 expression levels in acute myeloid leukemia cells carrying the nucleophosmin (NPM1) mutation', *Haematologica*, 96(10), pp. 2–5. doi: 10.3324/haematol.2011.043786.
- Qamar, S. *et al.* (2018) 'FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation- π Interactions', *Cell*, 173(3), pp. 720–734. doi: 10.1016/j.cell.2018.03.056.
- Sabari, B. R. *et al.* (2018) 'Coactivator condensation at super-enhancers links phase separation and gene control', *Science*, 361(6400), pp. 1–17. doi: 10.1126/science.aap9195.
- Schmidt, H. B. *et al.* and Görlich, D. (2015) 'Nup98 FG domains from diverse species spontaneously phase-separate into particles with nuclear pore-like permselectivity', *eLife*, 4, pp. 1–30. doi: 10.7554/eLife.04251.
- Seita, J. and Weissman, I. . (2010) 'Hematopoietic Stem Cell: Self-renewal versus Differentiation', *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(5), pp. 1–20. doi: 10.1002/wsbm.86.Hematopoietic.
- Shin, Y. *et al.* (2017) 'Spatiotemporal control of intracellular phase transitions using light-activated optoDroplets', *Cell*. doi: 10.1016/j.physbeh.2017.03.040.
- Soverini, S. *et al.* (2018) 'Chronic myeloid leukemia: The paradigm of targeting oncogenic tyrosine kinase signaling and counteracting resistance for successful cancer therapy', *Molecular Cancer*. *Molecular Cancer*, 17(1), pp. 1–15. doi: 10.1186/s12943-018-0780-6.
- Staal, F. J. T. *et al.* (2016) 'Aberrant Wnt Signaling in Leukemia', *Cancers*, 8(78). doi: 10.3390/cancers8090078.
- Thanasopoulou, A., Tzankov, A. and Schwaller, J. (2014) 'Potent co-operation between the NUP98-NSD1 fusion and the FLT3-ITD mutation in acute myeloid leukemia induction', *Haematologica*, 99, pp. 1465–1471. doi: 10.3324/haematol.2013.100917.
- Toretsky, J. A. and Wright, P. E. (2014) 'Assemblages: Functional units formed by cellular phase separation', *Journal of Cell Biology*, 206(5), pp. 579–588. doi: 10.1083/jcb.201404124.

- Vikal, Y. and Kaur, D. (2017) 'Nucleus', in *Plant Cells and their Organelles*, pp. 146–208.
- Wang, G. G. *et al.* (2007) 'NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis', *Nature Cell Biology*, 9(7), pp. 804–812. doi: 10.1038/ncb1608.
- Wang, G. G. *et al.* (2009) 'Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger', *Nature*, 459(7248), pp. 847–851. doi: 10.1038/nature08036.Haematopoietic.
- Watts, J. and Nimer, S. (2018) 'Recent advances in the understanding and treatment of acute myeloid leukemia [version 1; referees: 3 approved]', *F1000Research*, 7. doi: 10.12688/f1000research.14116.1.
- Wheeler, R. J. *et al.* (2019) 'Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease', *bioRxiv*. doi: 10.1101/721001.
- Xu, H. *et al.* (2016) 'NUP98 Fusion Proteins Interact with the NSL and MLL1 Complexes to Drive Leukemogenesis', *Cancer Cell*. Elsevier Inc., 30(6), pp. 863–878. doi: 10.1016/j.ccell.2016.10.019.
- Xu, S. and Powers, M. A. (2013) 'In vivo analysis of human nucleoporin repeat domain interactions', *Molecular Biology of the Cell*, 24(8), pp. 1222–1231. doi: 10.1091/mbc.E12-08-0585.
- Yung, E. *et al.* (2011) 'Delineating domains and functions of NUP98 contributing to the leukemogenic activity of NUP98-HOX fusions', *Leukemia Research*, 35(4), pp. 545–550. doi: 10.1016/j.leukres.2010.10.006.DELINEATING.
- Zhao, Y., Wang, Y. and Ma, S. (2018) 'Racial Differences in Four Leukemia Subtypes: Comprehensive Descriptive Epidemiology', *Scientific Reports*. Springer US, 8(1), pp. 1–10. doi: 10.1038/s41598-017-19081-4.