Aus dem Department für biomedizinische Wissenschaften

der Veterinärmedizinischen Universität Wien

Institut für medizinische Biochemie

(Leiter: Univ.-Prof. Dr. rer. nat. Florian Grebien)

Characterization of biomolecular condensation in NUP98-fusion protein-driven leukemia

Diplomarbeit

Veterinärmedizinische Universität Wien

vorgelegt von

Paula Meisel

Wien, im Oktober 2022

Supervisor: Univ.-Prof. Dr.rer.nat. Florian Grebien

University of Veterinary Medicine Vienna

Department of Biomedical Sciences

Institute for Medical Biochemistry

Veterinärplatz 1

1210 Vienna

Reviewer: Univ. Prof. Dr.med.vet. Maik Dahlhoff

University of Veterinary Medicine Vienna

Department of Biomedical Sciences

Institute of in-vivo and in-vitro models

Veterinärplatz 1

1210 Vienna

Declaration of Authenticity

I hereby declare that I composed this Diploma thesis independently. All sources I used are fully referenced and I did not use any other materials. I declare that none of the work included in this thesis was published or used for any other purpose than the acquisition of this Diploma thesis.

24.10.2022

Date

Signature

Acknowledgements

First of all, I would like to express my gratitude towards Professor Florian Grebien. Thank you for giving me the opportunity to join the lab to work on this project. I would also like to thank Melanie Allram for your patience in guiding me through the work. I am incredibly thankful for everything you have taught me this year. Last, but not least thanks to the rest of the lab for welcoming me into your group and for the great work environment.

Contents

I. Abstract (English)

Nucleoporin 98-fusion oncoproteins cause acute myeloid leukemia in around 5 % of children diagnosed with leukemia. The N-terminal part of NUP98, which contains intrinsically disordered regions of Phenylalanine-Glycine repeats, is fused to over 30 different C-terminal fusion partners in the context of leukemia. Studies of the intracellular localization of *NUP98*-fusion oncoproteins have revealed their disassociation from the nuclear membrane and their assembly in characteristic nuclear punctae. *In-vitro* experiments have shown that the N-terminus of NUP98 is capable of undergoing liquid-liquid phase separation to form biomolecular condensates. Therefore, it is likely that NUP98-fusion proteins are capable of forming biomolecular condensates that may drive leukemogenesis. As condensate composition not only depends on proteins but also on RNA molecules, the knowledge of RNAs which contribute to the formation and maintenance of oncogenic NUP98-fusion condensates may enable the identification of novel targets for the treatment of NUP98-fusion oncoprotein-driven leukemia. To study RNA molecules which associate with fusion-oncoprotein containing biomolecular condensates in murine *NUP98::KDM5A*-dependent cells, this work aims to establish the CRISPR-Cas13 system for targeted RNA knockdown. To elucidate the effect of targeted RNA knockdown a cell model was created by introducing the CasRx protein into human HEK 293T cells as well as murine *NUP98::KDM5A*-dependent AML cells. While this study laid the foundation for the establishment of the CRISPR-Cas13 system in these cellular models, additional work will be required to fully establish and validate this system for targeted RNA knockdown.

II. Abstract (German)

In etwa 5 % aller von Leukämie betroffenen Kindern sind Translokationen des Genes Nucleoporin98 (*NUP98*) nachweisbar. Der N-Terminus von NUP98 enthält intrinsisch ungeordnete Regionen aus Wiederholungen der Aminosäuren Phenylalanin und Glyzin (FG) und ist in Patienten mit akuter myeloische Leukämie mit über 30 verschiedenen C-terminalen Fusionspartnergenen verbunden. Studien über die intrazelluläre Lokalisation von NUP98-Fusionsproteinen haben gezeigt, dass diese nicht mit der Kernmembran assoziiert sind, sondern kleine Ansammlungen im Kern bilden. *In-vitro* Experimente haben gezeigt, dass der N-Terminus von NUP98 zur Phasentrennung fähig ist und dadurch biomolekulare Kondensate bilden kann. Diese Strukturen könnten ursächlich mit der Leukämogenese hämatopoetischer Zellen verbunden sein. Die Zusammensetzung von biomolekularen Kondensaten ist abhängig von Proteinen und RNA. Daher könnte die Identifikation von RNA Molekülen, welche die Bildung und Aufrechterhaltung von NUP98-Fusionsprotein Kondensaten beeinflussen, dazu beitragen einen neuen Therapieansatz für Patienten mit *NUP98* Translokationen zu finden. Das Ziel dieser Arbeit ist es, ein CRISPR-Cas13 Model zu etablieren, um RNA Moleküle zu untersuchen, welche mit *NUP98::KDM5A* Kondensaten interagieren. Um den Effekt des gezielten RNA knockdown zu analysieren wurde ein Zellmodel etabliert, indem das Protein CasRx in humane HEK 293T Zellen und murine *NUP98::KDM5A*-getriebe Leukämiezellen eingebracht wurde. Diese Arbeit legt den Grundstein zur Etablierung des CRISPR-Cas13 Systems in diesen Zellmodellen. Um das System vollständig zu etablieren und es funktionell zu validieren sind jedoch noch weitere Experimente notwendig.

III. List of abbreviations

AAVS1: adeno-associated virus integration site 1

ALL: acute lymphoblastic leukemia

AML: acute myeloid leukemia

APS: Ammonium persulphate solution

Cas: CRISPR-associated

CasRx: Cas13d derived from *Ruminococcus flavefaciens*

CDK1: Cyclin-Dependent Kinase 1

CLL: chronic lymphocytic leukemia

CML: chronic myeloid leukemia

CREBP: cAMP response element-binding protein

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

crRNA: CRISPR RNA

CTRL: control

ddH_2O: double-distilled water

dCasRx: catalytically dead CasRx

DMEM: Dublbeccos' Modified Eagle Medium

DR: direct repeat

EIF3B: eukaryotic translation initiation factor 3 subunit B

eGFP: enhanced green fluorescence protein

EP300: E1A Binding Protein P300

FACS: Fluorescence Activated Cell Sorting

FG-repeats: Phenylalanine-Glycine repeats

GFP: green fluorescent protein

GLEBS: Gle2p-binding-domain

GLFG-repeats: Glycine-Leucine-Phenylalanine-Glycine-repeats

HA-tag: human influenza hemagglutinin tag

HD: homodomain

HEK: human embryonic kidney cells

HEPN: higher eukaryotes and prokaryotes nucleotide binding domain

HSC: hematopoietic stem cell

IDH1/IDH2: isocitrate dehydrogenase 1/2

IDR: intrinsically disordered region

iRFP670: near infrared fluorescent protein with an emission peak of 670 nm

KDM5A: lysine specific demethylase 5A

LSC: leukemia stem cell

LTR: long terminal repeat

NPC: nuclear pore complex

NUP98: Nucleoporin 98

p-bodies: processing-bodies

PBS: phosphate buffered saline

PEI: polyethylenimine

PIF: primary induction failure

PLK1: polo-like Kinase 1

PML bodies: promyelocytic leukemia bodies

pre-crRNA: precursor CRISPR RNA

PTFE: polytetrafluorethylene

RAE1: Ribonucleic Acid export 1

RIPA buffer: Radioimmunoprecipitation assay buffer

RNAi: RNA interference

RT-qPCR: reverse transcriptase-quantitative PCR

SDS-PAGE: sodium dodecylsulfate polyacrylamide gel electrophoresis

shRNA: short hairpin RNA

SNRNP200: small nuclear ribonucleoprotein U5 subunit 200

TBS-T: Tris-buffered saline with Tween20

TEMED: Tetramethylethylendiamin

XPO1: Exportin 1

1. Introduction

1.1. Leukemia

Leukemia is a malignant disorder of the hematopoietic system characterised by the abundance of abnormally differentiated precursor cells in the bone marrow and the peripheral blood [1]. In the process of hematopoietic differentiation, common myeloid and lymphoid progenitor cells are derived from multipotent hematopoietic stem cells (HSCs). While myeloid progenitor cells differentiate into Thrombocytes, Erythrocytes, Mast cells, Granulocytes and Monocytes, lymphoid progenitor cells will differentiate into Natural Killer cells, B- and T-lymphocytes (Figure 1). Leukemia is believed to arise from single malignant cells referred to as the leukemia stem cells (LSCs), which can arise at several stages of hematopoietic differentiation. A variety of mutations allow LSCs to gain self-renewal capacity and enhanced proliferation properties while also interfering with terminal differentiation [2].

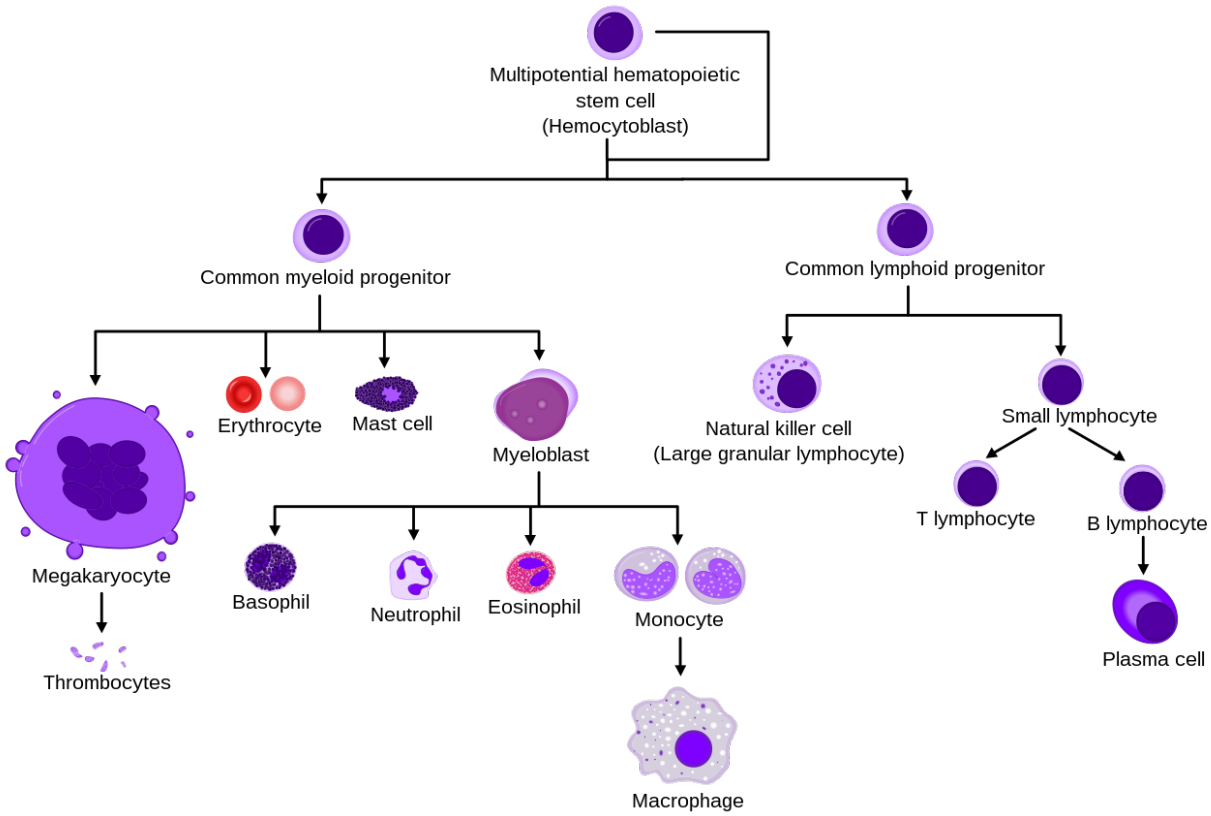


Figure 1: Differentiation of hematopoietic stem cells into the main blood cell types [3]

Leukemia is divided into nine groups with over 70 subtypes by the WHO [4]. However, only four types of leukemia are commonly distinguished based on cell of origin and progression of the disease: acute myeloid leukemia (AML), chronic myeloid leukemia (CML), acute lymphoblastic leukemia (ALL) and chronic lymphocytic leukemia (CLL). About 80 % of all children diagnosed with leukemia suffer from ALL, while in adults CLL and AML are most common [5], [6].

1.2. Acute Myeloid Leukemia (AML)

Acute myeloid leukemia is characterized by the abundance of abnormally differentiated myeloid cells. These immature myeloid blasts include myeloblasts, monoblasts or megakaryoblasts. Their abnormal accumulation impairs normal hematopoiesis, leading to bone marrow failure and anemia, as well as an increased risk of infection and hemorrhage. AML commonly affects older patients and the mean age of patients at diagnosis is 68 years. However, also children and infants can be affected by AML. Patients usually show a rapid disease onset and are diagnosed by abnormal white blood cell counts, morphological assessment of bone marrow biopsies and screening for genetic alterations associated with haematological malignancies. For some of the 21 subtypes of AML a minimum of 20 % blasts in the bone marrow or peripheral blood is required for diagnosis, while for others specific genetic alterations are sufficient for diagnosis [1]. The WHO distinguishes 13 subtypes of AML with disease-defining genetic abnormalities as well as eight subtypes of AML which are defined by morphological signs of (abnormal) differentiation [4].

AML patients show a great variety of different genetic alternations. Studies on over 1,500 AML patients have identified 5,234 driver mutations across 76 genes. Most patients harbour at least two or three of these driver mutations [7]. Recurrent mutations associated with AML affect genes involved in epigenetic modifiers, for example *DNMT3A*, *TET2* or *IDH1 and IDH2*, but also genes involved in signal transduction, such as *FLT3* or *NRAS* and transcription factors, including *CEBPA* or *RUNX1*. Furthermore, chromosomal rearrangements occur frequently in AML patients and can be found in three out of four children diagnosed with AML [8].

The standard therapy for AML is intensive chemotherapy which is administered in a 7 + 3 regimen of 7 days of cytarabine together with the addition of an anthracycline for the first 3 days of treatment. In patients older than 60 years complete remission can be achieved in 40–60 %, while 60–85 % of younger patients achieve remission [1]. However, AML patients are frequently affected by disease relapse and relapse rates range from 40–50 % in young patients

and are even higher in elderly patients. In up to 40 % of AML patients relapses are observed despite receiving hematopoietic stem cell transplantation after achieving remission [9]. Moreover, the intensive induction chemotherapy treatment is not suitable for all patients, partly due to comorbidities associated with advanced age. Therefore, in AML patients ≥ 60 years the median survival time is only 6–9 months [1]. Furthermore, in 15 % of paediatric AML patients primary induction therapy is unsuccessful and does not lead to morphological signs of remission in the bone marrow. These children display so-called primary induction failure (PIF). Many of the patients affected by PIF suffer from AML driven by Nucleoporin 98 (NUP98)-fusion oncoproteins, which are generally associated with poor prognosis [10].

1.3. NUP98-fusion proteins

Recurrent chromosomal rearrangements in the context of AML involve the Nucleoporin 98 (*NUP98*) gene [8]. In the most recent update of the WHO classification of hematolymphoid tumours AML with *NUP98* rearrangements is recognized as distinct disease subtype [4]. NUP98-fusion proteins also occur in other hematopoietic disorders, such as myelodysplastic syndrome, CML in blast crisis, or mixed-phenotype acute leukemia. Nevertheless, they are often associated with AML with a particularly poor prognosis. While *NUP98*-fusion genes are only detected in around 2 % of all AML patients, they are more frequent in childhood and infant leukemia [8, 11].

Nucleoporin 98 is part of the nuclear pore complex (NPC), which mediates the selective bidirectional transport of RNA and proteins between the nucleus and the cytoplasm. The NPC consists of over 30 different nucleoporins. Some of those harbour repetitive amino acid stretches which are referred to as intrinsically disordered regions (IDRs) and consist of Phenylalanine-Glycine repeats (FG-repeats). Nucleoporins with FG-repeats make up the central pore of the NPC and form a permeability barrier for structures larger than 40 kDa. While small molecules can pass through the NPC by passive diffusion, larger molecules need to selectively bind to karyopherins in order to be transported across the nuclear membrane [12]. The NUP98 protein features an IDR which consists of Phenylalanine-Glycine and Glycine-Leucine-Phenylalanine-Glycine (GLFG-repeats). The export factor XPO1 as well as the cofactors CREB-binding protein (CREBBP) and EP300 interact with the IDR of NUP98. The N-terminus of NUP98 also includes a Gle2p-binding domain (GLEBS) which interacts with the mRNA export factor RAE1. The C-terminal part of NUP98 contains autoproteolytic cleavage sites as well as RNA binding sites (Figure 2A). Alternative splicing of the *NUP98* transcript also

enables NUP98 to interact with chromatin and the anaphase promoting complex independently of the NPC [11].

Fusion genes are results of chromosomal translocations and lead to the expression of novel fusion proteins that can have aberrant functions. NUP98-fusion proteins consist of the N-terminus of NUP98, which is fused to the C-terminal part of over 30 different fusion partner genes. The fusion partners can be divided into two groups: homeodomain (HD) proteins and non-homeodomain proteins. Most of the non-HD fusion partners share a coiled-coil domain and many also feature domains that are associated with histone writing or reading, such as plant homeodomains (PHD) [11]. Common fusion partners of *NUP98* in AML patients include members of the HOX gene family, *NSD1*, *PSIP1*, *DDX10* and *KDM5A*.

In monocytic, erythroid and megakaryoblastic leukemia NUP98 is recurrently fused to the C-terminal part of KDM5A [8]. *KDM5A* is located at chromosome 12 and encodes for the lysin specific demethylase 5A, which demethylates di- and trimethylated histone H3 at position lysin 4. The third PHD (Plant homeodomain) finger of wildtype KDM5A is conserved in NUP98::KDM5A fusion proteins (Figure 2B). This domain is important for binding of methylated residues in histone H3, allowing the removal of inhibitory histone methylation marks in genes such as *HOXA5*, *HOXA7*, HOXA*9* and HOXA*10*. The expression of *HOXA* genes results in inhibition of regular cell differentiation and contributes to leukemogenesis [13].

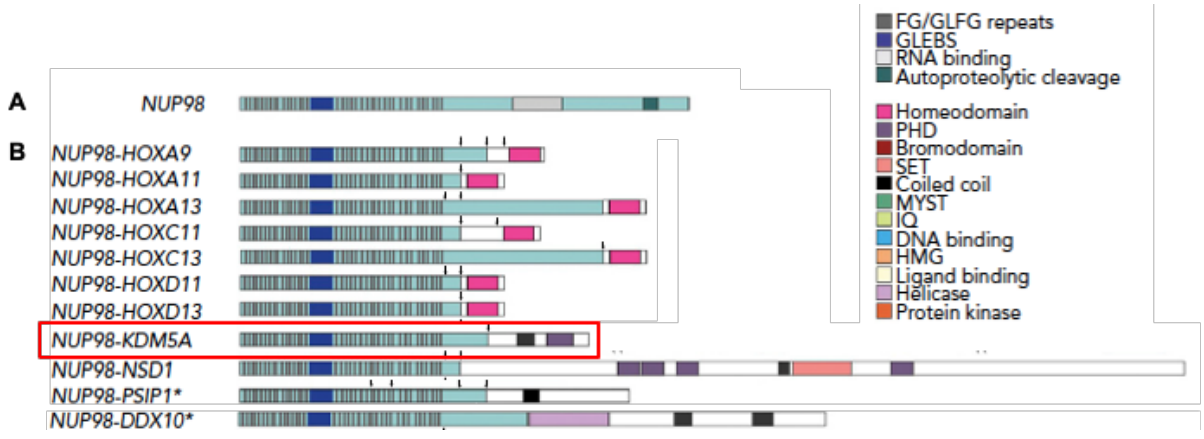


Figure 2: Domain Structure of (A) NUP98 and (B) NUP98::KDM5A (in red) and other recurrent NUP98 fusion proteins [8]

While the NUP98 protein mostly localizes to the nuclear membrane as part of the NPC, it can also be found in focal clusters throughout the nucleus [11]. Several NUP98-fusion proteins including NUP98*::*KDM5A have also been shown to form nuclear condensates, while they do not co-localize with endogenous NUP98 at the NPC [14].

1.4. Biomolecular condensates

Biomolecular condensates are cellular compartments, which lack a lipid membrane that encloses traditional organelles such as the mitochondria or the Golgi apparatus. In eukaryotic cells more than 35 different biological functions, including transcription, DNA damage repair and RNA metabolism, are known to be compartmentalized within biomolecular condensates [15]. Biomolecular condensation can occur both in the nucleus as well as the cytoplasm. Examples for nuclear condensates are the Nucleolus, Cajal bodies and PML bodies, whereas cytoplasmic condensates include Stress granules and P-bodies [15].

The formation of biomolecular condensates is driven by proteins with particular structural features, including low complexity domains like intrinsically disorder regions. In addition, nucleic acids can also induce biomolecular condensation. IDRs and the negative charge of nucleic acids allow for weak, multivalent low affinity interactions in a heterotypic and homotypic manner to form condensates. Once condensation has been initiated, other molecules can be recruited into the condensate. Dynamic exchange of molecules between the condensate and its surrounding is possible and allows for quick changes of the condensate composition as well as rapid assembly and disassembly [16].

The NUP98 N-terminus harbours an IDR which could allow both the normal protein as well as NUP98-fusion proteins to undergo biomolecular condensation. Terlecki-Zaniewicz *et al.* showed that the expression of artificial fusion proteins consisting only of FG repeats fused to the C-terminus of KDM5A can phenocopy gene expression patterns of NUP98-fusion oncoproteins. This indicates that FG-repeats are essential for leukemogenesis, as other amino-acid repeats fused to the same partner gene did not have the same effect [14]. Several other studies discovered that NUP98-fusion proteins are localized in nuclear puncta [17], [18]. Furthermore, NUP98::HOXA9 fusion proteins have been shown to form condensates by liquid-liquid phase separation *in-vitro* [17], suggesting that these proteins may also be able to form condensates in live cells. Biomolecular condensation has been reported as an important feature of NUP98-fusion protein driven AML as immunofluorescence imaging of a *NUP98::KDM5A* PDX model revealed that the NUP98::KDM5A fusion protein also locates in

nuclear puncta in human AML cells [17]. Furthermore, mutations within the NUP98-IDR that lead to a loss of condensation potential failed to induce leukemia *in vivo*. Therefore, fusion of *NUP98* to a chromatin targeting domain might induce relocation of NUP98-fusion proteins from the nuclear membrane to nuclear puncta, where they may act as transcription factors. Aberrant recruitment of cofactors and activation of transcription might ultimately drive leukemogenesis [16].

Biomolecular condensation is often mediated by proteins and RNA, and proteins involved in biomolecular condensation often feature RNA binding domains. The negative charge of RNA promotes the separation of polymers due to opposite electrical charge, which can mediate condensate formation together with RNA-RNA interactions and interactions with RNA-binding proteins [19]. In P-bodies of *Drosophila melanogaster* proteins have been shown to form biomolecular condensates at lower concentration if RNA is present [20]. While this implies that RNA molecules contribute to biomolecular condensation, a high concentration of RNA molecules was shown to prevent this process on some occasions due to charge repulsion effects [20]. The composition, as well as the structure, length and expression levels of RNA may contribute to a variety of characteristics of condensates including size, shape, viscosity, or protein composition [19]. In the context of biomolecular condensates RNA can have an important scaffolding role by mediating multi-valent low-affinity interactions. However, the roles of RNA in in the condensation of oncogenic fusion proteins has not been investigated.

1.5. CRISPR-Cas13 mediated RNA knockdown

RNA interference (RNAi) has been widely utilized as tool for targeted gene knock-down to study cellular functions of specific genes. RNAi is a biological process which induces gene silencing by mRNA degradation and/or inhibition of translation and plays an important role in both transcriptional and post-transcriptional gene regulation in eukaryotes [21]. Most frequently small hairpin RNAs (shRNAs) are used for gene knockdown, but their high expression is associated with increased off-target effects which may lead to cellular toxicity [22]. Furthermore, mRNA has been shown to be resistant to RNAi when sequestered in the nucleus of human cells [23].

The CRISPR-Cas system (Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR Associated) can be used as an alternative approach to induce the efficient knockdown of target genes. CRISPR-Cas systems are essential for the immune response against bacteriophages and invading plasmids in bacteria and archaea. Class II systems derived from

bacteria feature one multi-subunit effector, while class I systems, which are regularly found in archaea, feature multiple effector proteins [24]. The CRISPR array in the bacterial genome has been first identified in 1987. During the adaption phase fragments of foreign genomic material are integrated into the bacterial genome as protospacer sequences. Precursor CRISPR RNA (pre-crRNA) is then transcribed and matured into crRNA by enzymatic cleavage. The Cas enzyme interacts with the crRNA. This interaction guides the Cas enzyme to the target genomic sequence, which is complementary to the spacer sequence included in the crRNA. After their identification targeted nucleic acid sequences are cleaved by Cas enzymes [25] (Figure 3).

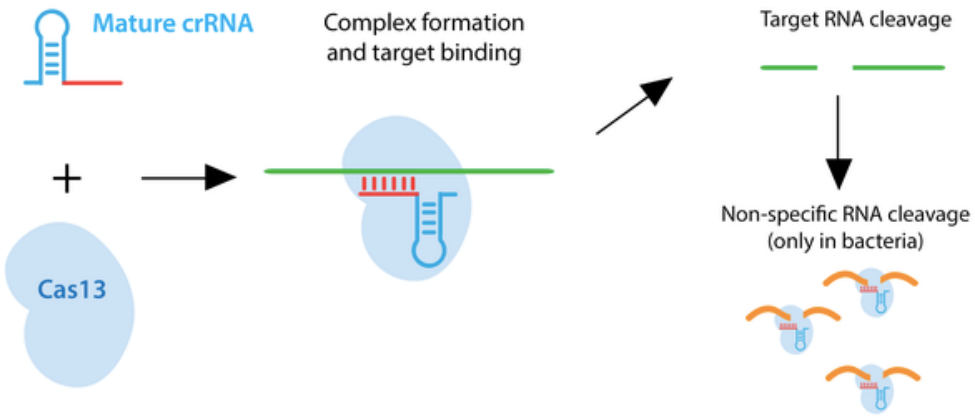


Figure 3: Schematic overview of targeted RNA knockdown using CRISPR-Cas13 [26]

While CRISPR-Cas9 is the most frequently used system for genetic engineering, other Cas systems have also been identified. Cas13 proteins derived from *Leptotrichia shahii* (LshCas13a) were initially discovered in 2015. They belong to class II systems alongside with Cas9 and have also been successfully used for genome editing. While other CRISPR-Cas proteins target DNA, Cas13 proteins, which are also referred to as type VI Cas proteins, can be used for RNA manipulation as they exclusively cleave single-stranded RNA and pre-crRNA. All four subtypes of Cas13 proteins share two HEPN domains (higher eukaryotic and prokaryotic nucleotide binding domain), which confer ribonuclease activity (Figure 4) [27].

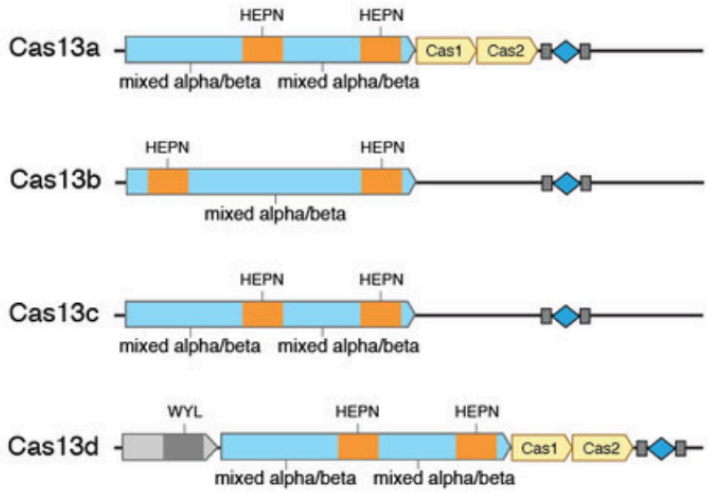


Figure 4**: **Comparison of domain structures of Cas13 proteins. The common HEPN domains confer ribonuclease activity [28]

In bacteria Cas13 is activated upon targeted RNA cleavage, which leads to non-specific cleavage of other RNAs [25]. However, no collateral Cas13 activity has been observed in mammalian cells [29]. Hence, the Cas13 system may be used for targeted RNA degradation in mammals [25]. There are several subtypes of Cas13 which have been utilized for RNA knockdown *in-vitro*, including Lwa13a, derived from *Leptrotrichia wadeii*, PspCas13b from *Prevotella sp. P5-125* and RfxCas13d (CasRx) from *Ruminococcus flavefaciens* [27]. Cas13d has the most favourable features for use in RNA editing, including its small size of and the lack of sequence constraints and the Cas13d family member CasRx has been shown to induce targeted RNA knockdown in human cells with the highest efficiency [25].

1.6. Aim of this study

Recent research suggests that biomolecular condensation plays a crucial role in NUP98-fusion oncoprotein-driven leukemia. However, little is known about the composition of oncogenic NUP98-fusion protein driven condensates and the role of RNAs in the context of oncogenic biomolecular condensation has not been addressed in the past. With this work we hope to establish an *in-vitro* model to study the contributions of different candidate RNA molecules to the formation and dynamic changes of biomolecular condensation in NUP98-fusion protein-driven leukemia. By introduction of CasRx into murine AML cells harbouring the *NUP98::KDM5A* fusion gene it will be possible to study effects of targeted RNA knockdown on condensate dynamics, formation and composition.

2. Materials and Methods

2.1. Cell culture

For this study HEK 293T cells and a mouse AML cell line harbouring the *NUP98::KDM5A* fusion gene were used. HEK 293T cells were cultivated in DMEM (Gibco, USA) supplemented with 10 % fetal bovine serum (Thermo Fisher Scientific, USA), 1 % L-Glutamine (Gibco, USA) and 1 % Penicillin/Streptomycin (Gibco, USA). *NUP98::KDM5A*-dependent AML cells were cultivated in RPMI 1640 (Gibco, USA), supplemented with 10 % fetal bovine serum (Thermo Fisher Scientific, USA), 2 % L-Glutamine (Gibco, USA), 1 % Penicillin/Streptomycin (Gibco, USA), 1 % Sodium pyruvate (Sigma-Aldrich), 0.1% 2-Mercaptoethanol (Gibco, USA) and 2 % HEPES (Sigma-Aldrich, USA). All cells were cultivated at 37 °C, 5 % CO_2 and 95 % humidity and were split every 2–3 days to adjust cell numbers to 0,5 * 10^6 cells/ml. For splitting HEK 293T cells were washed with phosphate buffered saline (PBS) and detached from the plate using Trypsin (Gibco, USA) and resuspended in culture medium after 3–5 minutes of incubation at 37 °C.
LentiX were used for lentiviral production and were cultivated in DMEM (Gibco, USA) supplemented with 10 % fetal bovine serum (Thermo Fisher Scientific, USA), 2 % L-Glutamine (Gibco, USA) and 1 % Penicillin/Streptomycin (Gibco, USA).

2.2. Plasmids and Cloning

2.2.1. CasRx plasmid

For the introduction of CasRx into target cells plasmid pXR001, EF1a-CasRx-2A-EGFP (plasmid #109049, Addgene, USA) was used. In addition, a catalytically dead CasRx (dCasRx) mutant was introduced to the same cell lines (plasmid pXR002, EF1a-dCasrx-2A-EGFP, plasmid #109050, Addgene, USA). Plasmid DNA was isolated from bacteria after overnight culture in LB-medium supplemented with 100 ng/ml Carbenicillin using the Qiagen Midi Prep Kit (Qiagen, Netherlands) following the manufactures' instructions.

2.2.2. Guide RNA plasmid cloning

A lentiviral backbone for the introduction of guide RNAs was created using the backbone pLenti-hU6-sgRNA-IT-PGK-iRFP670 which was digested with the restriction enzymes AgeI-HF (NEB, USA) and Esp3I (NEB, USA). After restriction enzyme digest gel electrophoresis

was performed and the linearized plasmid DNA was extracted using the MiniPex 3 in 1 Kit (IMP, Austria) following the manufactures' protocol. Subsequently, the CasRx direct repeat sequence (DR30) was introduced into the backbone via Gibson Assembly using Gibson Assembly Master Mix (NEB, USA). 100 ng of digested lentiviral backbone and 4,82 ng of the 229 bp fragment which includes the DR30 as well as BsmBI restriction sites for guide sequence were used. Backbone and fragment were mixed in a 1:2 molar ratio, using 100 ng of digested lentiviral backbone (9505 bp) and 4,28 ng of the insert (229 bp). The molar ratio was calculated using the formula pmols = (weight in ng) x 1,000 / (base pairs x 650 Daltons). DNA, 10 µl of Gibson Assembly Master Mix (2X) and water was assembled in 20 µl total volume and incubated at 55 °C for 1 hour, before bacterial transformation with the ligation product was performed.

Table 1*: g-block sequence used to modify the plasmid pLenti-hU6-sgRNA-IT-PGK-iRFP670*

g-block	atcttgtggaaaggacgaaacaccgaacccctaccaactggtcggggtttgaaacggagacggacgtctctttttttg ctagcgatctaattctaccgggtaggggaggcgcttttcccaaggcagtctggagcatgcgctttagcagccccgctgg gcacttggcgctacacaagtggcctctggcctcgcacacattccacatccaccggtaggcgccaaccggct

For guide RNA cloning the modified lentiviral backbone was digested with 3 µl of the restriction enzyme BsmBI-v2 per 5 µg of plasmid DNA (NEB, USA), incubated at 55 °C for 2 hours and dephosphorylated with Antarctic phosphatase (NEB, USA) for 1 hour at 37 °C. Restriction enzyme digests were performed in duplicates or triplicates. Fragments were separated by gel electrophoresis using a 0.7 % agarose gel. The linearized plasmid was extracted from the gel using the MiniPex 3 in 1 Kit (IMP, Austria).

Seventeen guide RNAs were used (Table 2) and designed with a 5'-AAACG overhang on the forward strand and a AAAA-3' overhang on the reverse strand. For each guide RNA 100 µM Forward and 100 µM Reverse oligos were annealed and phosphorylated using 1 µl of 10X T4 DNA Ligation buffer (NEB, USA), 0.5 µl T4 Polynucleotide Kinase (NEB, USA) and water in a total volume of 10 µl for 30 minutes at 37 °C and 5 minutes at 95 °C in the thermocycler. Afterwards oligos were diluted 1:200 and ligation of guide RNAs and backbone was performed using 1 µl of the diluted oligos, 0.5 µl of T4 DNA ligase (NEB, USA), 1 µl of 10X T4 Ligation Buffer (NEB, USA) and water in a total volume of 10 µl. After 10 minutes of incubation time at room temperature 2 µl of the ligation product and 15 µl of bacteria was used for bacterial transformation. Upon failure to introduce guide RNA sequence, the procedure was repeated with the restriction enzyme Esp3I (NEB, USA) and CutSmart Buffer (10X) (NEB, USA), again

using 3 µl of restriction enzyme per 5 µg of DNA and incubating the reaction for 1 hour at 37 °C, before proceeding as described.

Table 2: *List of guide RNA sequences*

Target	Sequence
AAVS1.1	GGGGCCACTAGGGACAGGAT
AAVS1.2	GCTCCGGAAAGAGCATCCT
AAVS1.3	GCTGTGCCCCGATGCACAC
CDK1.1	AAAGATTCCACTTCTGGCCACACTTCATTA
CDK1.2	GGAATCCTGCATAAGCACATCCTGAAGACT
CDK1.3	GATGAGATATAACCTGGAATCCTGCATAAG
CDK1.4	AAGATTCCACTTCTGGCCACACTTCATTAT
NT_rc.1	GCTCCCGCCGAGATGCCGCCGATTTCT
NT_rc.2	AAGCTATAACCTCCGTAAAATAAGAAA
NT_rc.3	CCCTACTTTGTATTCCAGTTACGTGTA
PLK1.1	GAAGAAGTTGATCTGCACGCTGCCGTTGCT
PLK1.2	AGAACTCGTCATTAAGCAGCTCGTTAATGG
PLK1.3	AAGAACTCGTCATTAAGCAGCTCGTTAATG
PLK1.4	AAGGTTGCCCAGCTTGAGGTCTCGATGAAT
EIF3B.1	TTCTTTCATCTCGACCACATCCACAGG
EIF3B.2	GCATTCTGCCTCTTCAAGCCAGTAACG
SNRNP200.1	CTCATCCGTCTCAAGGGCAGCACGGTA
SNRNP200.2	CCCAGCATCTGCAGAATGTCCAGTGCT

2.3. Bacterial Transformation

NEB Stable Competent E. coli (High Efficiency) (NEB, USA) were used for transformation. 15 µl or 30 µl of Bacteria were thawed on ice before adding 2 µl or 3 µl of plasmid DNA. Cells were incubated on ice for 10 minutes. A heat shock was performed at 42 °C for 45 seconds, and the mixture was placed back on ice for 2 minutes. Afterwards 200 µl or 400 µl of LB-media was added and cultures were incubated at 37 °C, for 1 hour with shaking (650 rpm). After incubation bacteria were plated on LB-Agar plates with 100 µg/ml Carbenicillin (Carl Roth, Germany). After overnight incubation at 37°C single colonies were picked and cultivated in 4 ml LB-media with 100 µg/ml Carbenicillin. They were then either incubated overnight to isolate plasmids using the Mini Pex 3 in 1 kit (IMP, Austria) or transferred to 50 ml

LB-media before overnight incubation for plasmid isolation using Qiagen Midi Prep Kit (Quiagen, Netherlands). After plasmid isolation Sanger sequencing was performed (Microsynth, Switzerland) to confirm the presence of the desired sequence.

2.4. Lentiviral Transduction

The LentiX packaging cell line was seeded at about 70 % confluency in DMEM without FBS supplemented. For transfections, 4 µg of the plasmid pXR001 or pXR002 were assembled with packaging plasmids pCMVR8.74 (2 µg) and VSV-G (1 µg) together with Polyethylenimine (PEI) (Polyscience, USA) in a 3:1 ratio of DNA:PEI. After 20 minutes of incubation at room temperature the mixture was added to the LentiX cells in a dropwise fashion. The next day the medium was changed to DMEM and RMPI (with supplements as described above) for HEK 293T cells and NUP98::KDM5A-driven AML cells, respectively. 48 and 72 hours after transfection the supernatant of LentiX cells containing lentivirus was harvested, filtered through a 0.45 µm filter (TPP, Switzerland) and stored at 4 °C. Target cells where transduced with lentiviral supernatants in a ratio of 1:5 to 1:10 supplemented with 10 µg/ml Polybrene (Merck, Germany). Spinfection was performed at 1000 x g for 90 minutes at room temperature. The culture medium containing lentiviral particles was removed after 24 hours and replaced with fresh medium. For HEK 293T cells spinfection was performed once, while for *NUP98::KDM5A*-dependend AML cells the procedure was repeated after 24 hours using a pool of the first and second virus harvest.

2.5. Selection of single cell clones

After lentiviral transduction cells were sorted by fluorescence assisted cell sorting (FACS) for GFP-positive cells (GFP+), which reports the presence of the construct for CasRx expression. GFP+ cells were cultivated for one week to allow for recovery. Then, cells were diluted to 2.5 cells / ml in conditioned medium, which was harvested from cells after transfection and filtered through a 0.22 µm filter before usage. After diluting the cells, the suspension was filtered through a 0.70 µm strainer subsequently seeded into three 96-well plates per condition in 200 µl per well. Plates were incubated for 10 days before single cell clones were expanded. GFP expression in clones was analysed using the IntelliCyt iQue Screener Plus (Sartorius, Germany).

2.6. HEK 293T transfection

After expansion of single cell clones to > 1 million cells, the GFP signal of CasRx-positive HEK 293T cells was confirmed using flow cytometry. Three single cell clones were selected and further expanded for transfection with 6 guide RNAs each. Cells were transfected with Guide RNAs targeting *CDK1*, *PLK1* and *AAVS1* as well as one additional non-targeting guide RNA using Polyethylenimine (PEI). 1 µg of plasmid DNA and 3 µl of PEI were mixed and added dropwise to HEK 293T cells after 20 minutes of incubation at room temperature. After 48 hours cells were analyzed for the presence of iRFP670 signal using a fluorescence microscope. 72 hours after transfection cells were collected and spun down at 300 x g for 5 minutes at 4 °C and washed twice with 1 ml of ice-cold PBS. Cell pellets were kept on ice at all times, shock frozen in liquid nitrogen and stored at - 80 °C until further use.

2.7. RT-qPCR

To confirm CasRx activity, we performed qPCR analyses to test the efficiency of target RNA knockdown upon transfection of guide RNAs targeting *CDK1* and *PLK1*. As negative control one guide RNA targeting the *AAVS1* locus was used. Cell pellets were thawed on ice and RNA extraction was performed using the Monarch RNA extraction kit following the manufactures' protocol. RNA was eluted in 50 µl nuclease free water and kept on ice at all times. RNA concentration was measured using the Tecan Spark Spectrophotometer. cDNA conversion was performed with 1 µg of RNA using the Revert-Aid first-strand cDNA synthesis kit (Thermo Fisher Scientific, USA) following the manufactures' instructions. RT-qPCR was performed in triplicates with 1 µl of cDNA, 1µl of forward and reverse primers (10µM), 5 µl SsoAdvanced Universal SYBR green Supermix (Bio-Rad, USA) and nuclease free water in a total volume of 10 µl. For RT-qPCR the Bio-Rad CFX96-Real-Time PCR Detection System (Bio-Rad, USA) was used. Gene expression was determined relative to GAPDH using the formula $2^{(-\Delta C(t))} * 100$.

***Table 3**: primers used for RT-qPCR*

target gene	forward primer	reverse primer
GAPDH	TGCACCACCAACTGCTTAGC	GGCATGGACTGTGGTCATGAG
PLK1	GGGCACAGTTTCGAGGTGGA	CGGGGTTGATGTGCTTGGGAA
CDK1	CACTTGGCTTCAAAGCTGGCTC	TGGGTATGGTAGATCCCGGC

2.8. Western Blot

1 million CasRx-expressing *NUP98::KDM5A*-driven AML cells were collected, spun down at 4 °C at 300 x g for 5 minutes and cell pellets were washed with ice cold PBS. Cells were spun down again at 4 °C, 300 x g for 5 minutes and after removal of the supernatant, cell pellets were resuspended in 100 µl of RIPA buffer (10 mM Tris pH 8,1 mM EDTA, 1% NP-40, 0.1% SDS, 140 mM NaCl, 0.1% Na-Deoxycholate), supplemented with Protease Inhibitor Cocktail, 500 mM sodium fluoride, 100 mM phenylmethylsulfonyl fluoride, 1 M dithiothreitol, 25 U/µl Benzonase and 5 mg/ml Tosyl phenylalanyl chloromethyl ketone on ice. Once the cell pellet was fully resuspended the mixture was incubated on ice for 20 minutes and centrifuged for 30 minutes at 12500 rpm at 4 °C. The supernatant was transferred to a fresh 1.5 ml Eppendorf tube and stored at -80 °C.

The protein concentration of lysates was measured using the Bradford Assay. Coomassie Protein Reagent was used and samples for the standard curve were prepared using 5, 10, 15, 20, 30 and 50 µg/ml of γ -globulin. 1 µl of protein lysate was added to 1 ml Coomassie Reagent and the absorbance at 595 nm was determined. Protein lysates were diluted in RIPA buffer and the concentration of samples was adjusted to 3.5 µg / µl in Lämmli Buffer supplemented with β-mercaptoethanol. Samples were boiled at 99°C for 10 minutes and spun down after cooling to room temperature. A 10 % sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel was used to separate proteins according to their size. The running gel was prepared with 1.5 M Tris-HCl pH 8.8, 10% SDS, 30% acrylamide and ddH_2O before adding 10 % APS and TEMED to initiate polymerisation of the gel. The stacking gel was prepared using 0.5 M Tris-HCl pH 6.8 and 30% acrylamide with ddH_2O. 10 % APS and TEMED were added to induce polymerisation of the gel. 70 µg of protein were loaded in each pocket and the gel was run at 80 V for 15 minutes and then at 120 V for 1 hour before proteins were transferred to a nitrocellulose PTFE membrane.

For transfer a buffer consisting of 50 mM Tris[hydroxymethyl]aminomethane, 380 mM glycine, 10 % methanol was used. Filter paper and nitrocellulose membrane were soaked in transfer buffer and assembled in a transfer chamber. The BioRad Trans-Blot Turbo system was used at 25 V, 1 A for 30 minutes.

After transfer the membranes were blocked for 1 hour in 5 % skim milk in TBS-T supplemented with 0.1 % Tween20 (Sigma-Aldrich, USA). Primary antibodies were diluted 1:1000 in 5 % skim

milk and the membranes were incubated overnight at 4 °C. Membranes were washed three times for 10 minutes in TBS-T before being placed in 5 % skim milk supplemented with secondary antibody at a 1:10000 dilution for 1.5 hours at room temperature.
The membranes were washed three times in TBS-T for 10 minutes and placed on transparent foil. A mixture of Amersham ECL prime luminol enhancer solution and Amersham ECL prime peroxide solution was added to the membranes in a 1:1 ratio before imaging was done using the Vilber Fusion FX instrument.

***Table 4**: Antibodies used for Western Blotting*

Antibody	Company and catalogue number
rabbit anti HA-Tag (C29F4) mAb	Cell Signaling Technology, USA (#3724)
mouse anti GAPDH (0411) mAb	Santa Cruz Biotechnology, USA (#47724)
goat anti rabbit IgG HRP-linked	Cell Signaling Technology, USA (#7074)
goat anti mouse IgG HRP-linked	Thermo Fisher Scientific, USA (#31439)

3. Results

3.1. Cloning of a lentiviral construct for CasRx-specific guide RNA expression

To generate an efficient tool for stable expression of guide RNAs we combined three important features in a lentiviral vector: two long terminal repeats (LTRs) to allow lentiviral integration after transduction, a 30 nucleotide direct repeat (DR) stretch that serves as CasRx-specific crRNA and a gene encoding the near infrared fluorescence protein (iRFP670) for detection of transduced cells.

The lentiviral vector pLenti-hU6-sgRNA-IT-PGK-iRFP670 which is regularly used for sgRNA expression in the Cas9 system was modified for the expression of CasRx-specific guide RNAs. This backbone plasmid includes the near infrared fluorescence protein with an emission peak at 670 nm (iRFP670), which allows the detection of transfected/transduced cells by fluorescence microscopy or flow cytometry. In order to utilize this backbone plasmid for the expression of CasRx guide RNAs a restriction enzyme digest was performed to remove sequences that are specific for Cas9 guide RNAs and introduce sequences that allow cloning and expression of CasRx guide RNAs. To confirm the successful restriction enzyme digest and to isolate the linearized plasmid gel electrophoresis was performed (Figure 5).

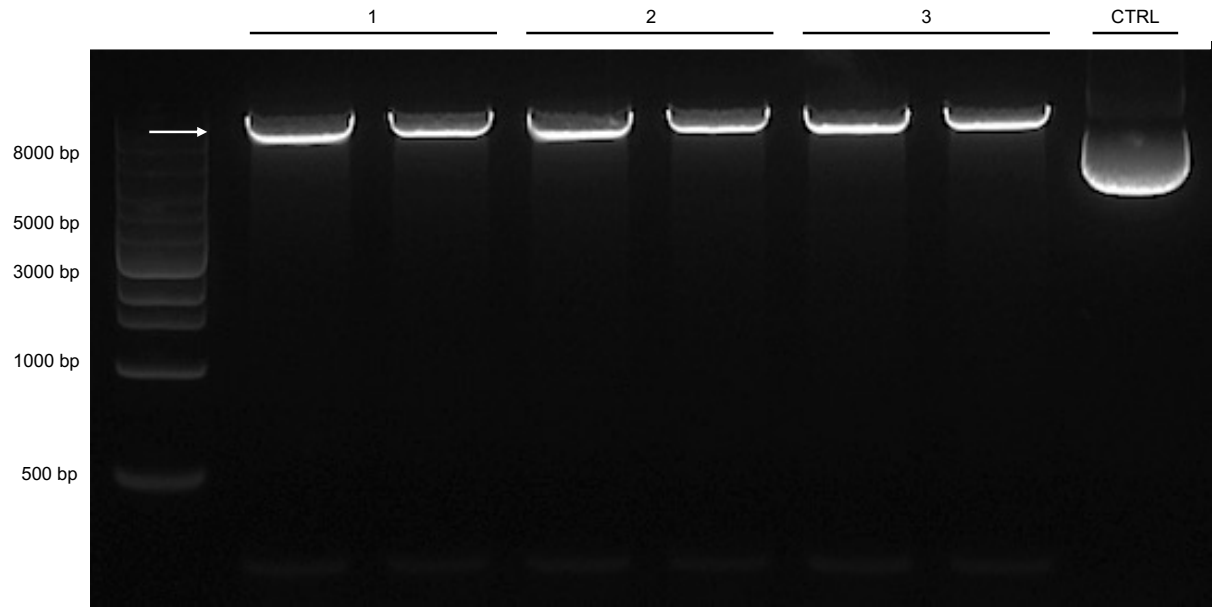
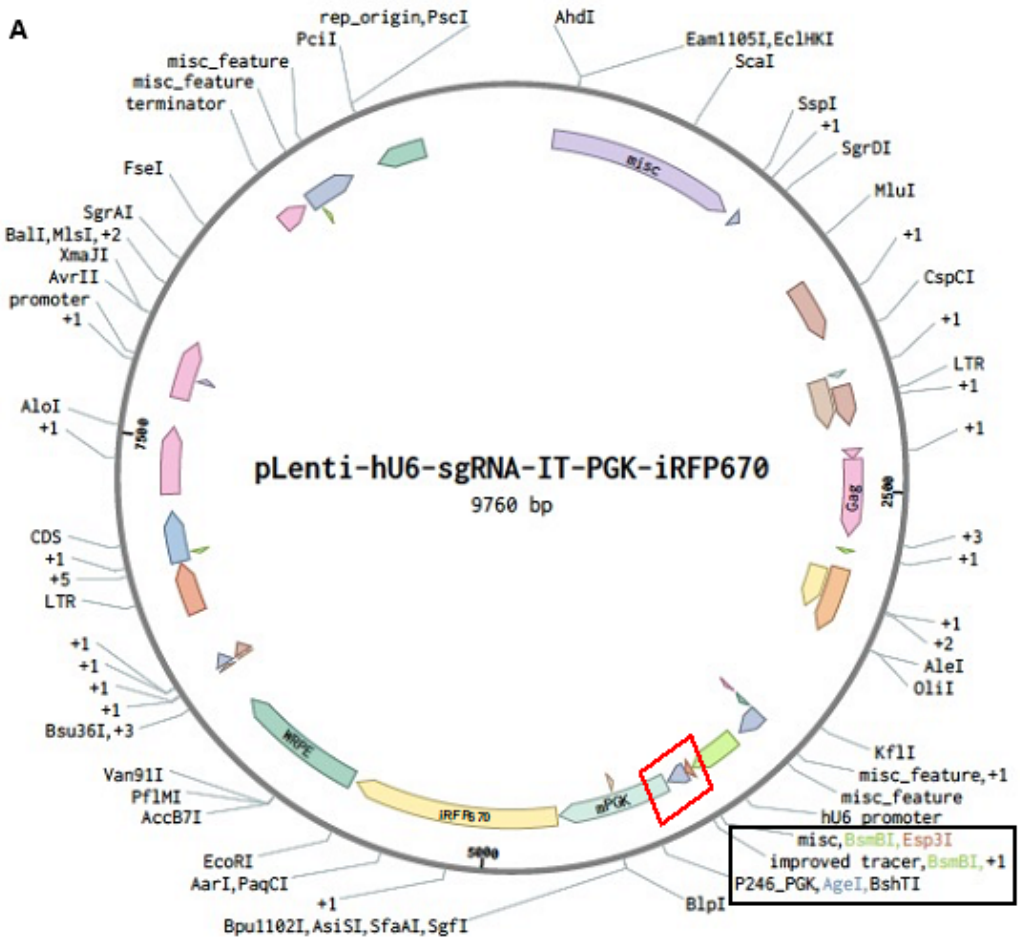


Figure 5: Gel electrophoresis after restriction enzyme digest using AgeI-HF and Esp3I of pLenti-hU6-sgRNA-IT-PGK-iRPF670*, *performed in triplicates (1-3). *Undigested backbone serves as negative control (CTRL). Arrow indicates the isolated fragment (~ 9505 bp).*

The 255 bp fragment which is removed from the original backbone includes the sequence of the Cas9 tracr RNA (improved tracer, IT) and two BsmBI restriction sites that are used for sgRNA cloning. Figure 6 shows a schematic overview of the modifications to pLenti-hU6-IT-PGK-iRFP670 which we performed to establish a plasmid for CasRx guide RNA introduction. Figure 6B shows the sequence of the fragment that was inserted into the backbone via Gibson Assembly. The inserted fragment reconstitutes the mPGK sequence, which was partially removed by the restriction enzyme digest. The sequence of the hU6 promoter as well as parts of the mPGK sequence also function as overhangs to align the insert to the linearized plasmid in order to ligate the two fragments using Gibson Assembly. Furthermore, the insert re-introduces two BsmBI restriction sites which we can use for the introduction of CasRx guide RNA. The correct integration of the inserted fragment and therefore the successful modification of pLenti-hU6-sgRNA-IT-PGK-iRFP670 was confirmed by Sanger sequencing.



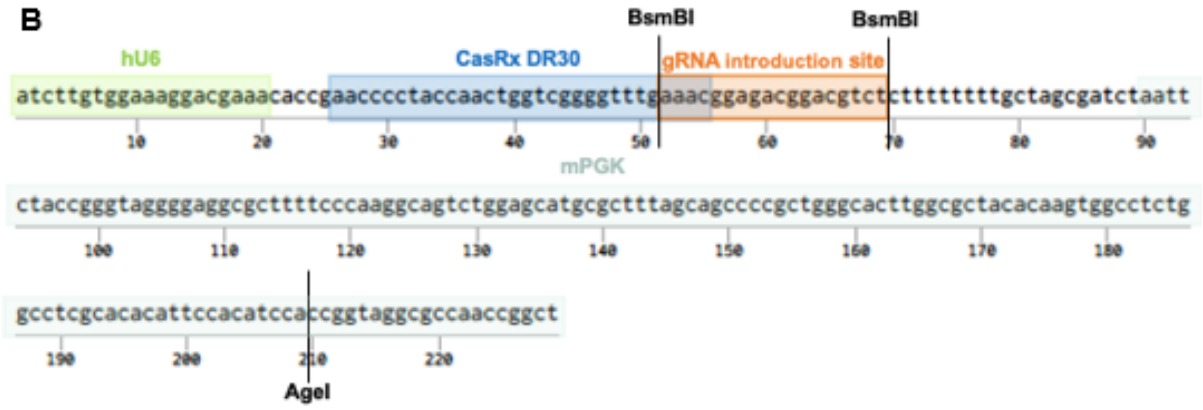


Figure 6: Modification of pLenti-hU6-sgRNA-IT-PGK-iRFP670. ***(A)*** *The red box indicates the 255 bp fragment which is removed and exchanged for:* ***(B)*** *Inserted fragment (229 bp) to introduce CasRx DR30 (blue) and a guide RNA (gRNA) introduction site (orange)*

3.2. Guide RNA cloning

For targeted RNA knockdown using CasRx 17 guide RNAs were selected. The guide RNAs can be divided into three categories depending on their target mRNAs. We selected guide RNAs targeting essential genes, such as *CDK1* and *PLK1*, while guides which are intended as negative control target either *AAVS1* or do not target any region in the human genome (non-targeting). Furthermore, we also selected guide RNAs targeting *EIF3B* and *SNRNP200*. These genes are not essential but known to be expressed in HEK 293T cells. Two, three or four guide RNAs were designed for each of the target genes using the RNA targeting online tool (www.rna-targeting-design.org) [30] to select target sequences of 30 nucleotides with predicted high efficiency. For each gene of interest we selected the top scoring guide RNA sequences. Overhangs that are compatible to restriction enzyme ends created by BsmBI were introduced on both ends of the guide RNA sequence to allow the ligation of the linearized plasmid backbone and the guide RNA sequence. For preliminary testing the cloning of guide RNAs targeting *CDK1*, *PLK1, AAVS1* was prioritized. The modified lentiviral plasmid for CasRx guide RNA expression was digested using the restriction enzyme BsmBI. Successful digest was confirmed using gel electrophoresis (Figure 7).

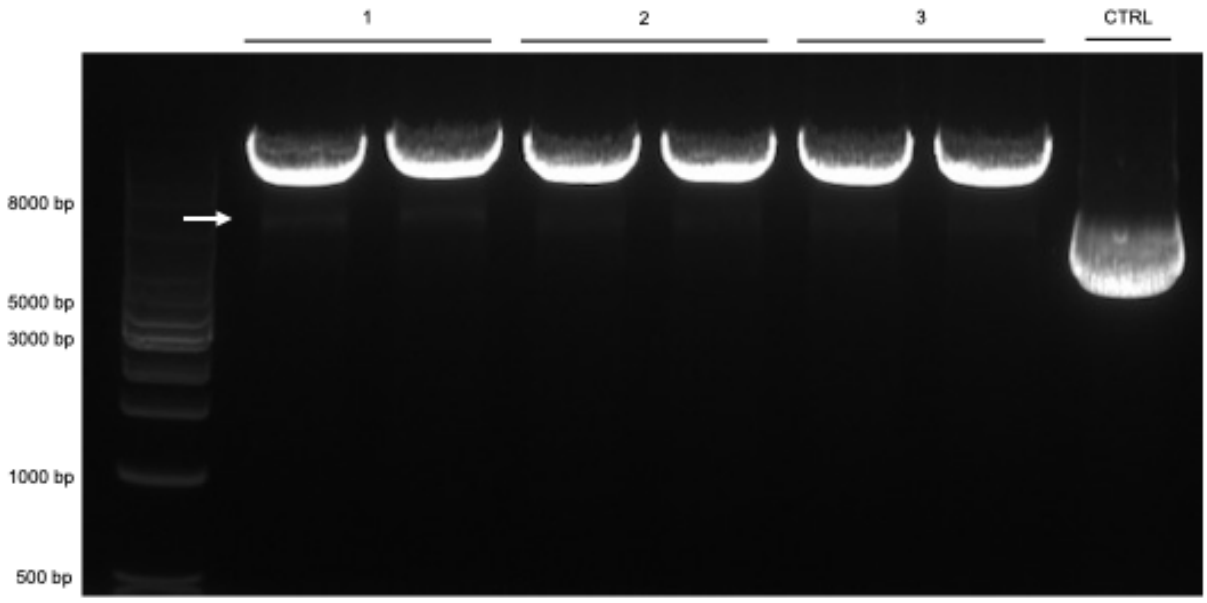


Figure 7: Gel electrophoresis after restriction enzyme digest. *The CasRx guide RNA backbone was digested using BsmBI-v2, performed in triplicates (1-3). Undigested backbone serves as negative control (CTRL), a faint band is visible at approximately 7500 bp, as indicated with an arrow.*

While the comparison to the undigested lentiviral plasmid (CTRL) confirms that the backbone was digested, the fragment that is released by the of BsmBI digest is only 20 bp long and can therefore not be efficiently visualized on an agarose gel. Therefore, the appearance of a linearized backbone does not confirm the correct enzymatic digest at both restriction sites as incomplete digestion cannot be ruled out. The plasmid backbone contains an Ampicillin resistance gene. This allows bacteria to grow in presence of Carbenicillin if the plasmid is circularized while the presence of linearized plasmid does not lead to a survival benefit. Ligation of guide RNA sequences with the linearized backbone should enable the plasmid to regain its circular form and allow the growth in presence of Carbenicillin. To ligate RNA guide sequences with the digested backbone the guide RNAs were designed with overhangs which correspond to the ends created by restriction enzyme digests with BsmBI.

Sanger sequencing was performed to confirm the presence and the correct sequence of the respective guide RNA. However, most of the results did not show the expected guide RNA sequence, but rather the sequence of a re-ligated backbone. While optimizing the cloning protocol, the best results were eventually achieved using the restriction enzyme Esp3I instead of BsmBI-v2. This is possible as BsmBI and Esp3I are isoschizomers and therefore share both

recognition and restriction sites and create identical ends. Figure 8 shows the gel electrophoresis result comparing Esp3I and BsmBI-v2 digest side by side.

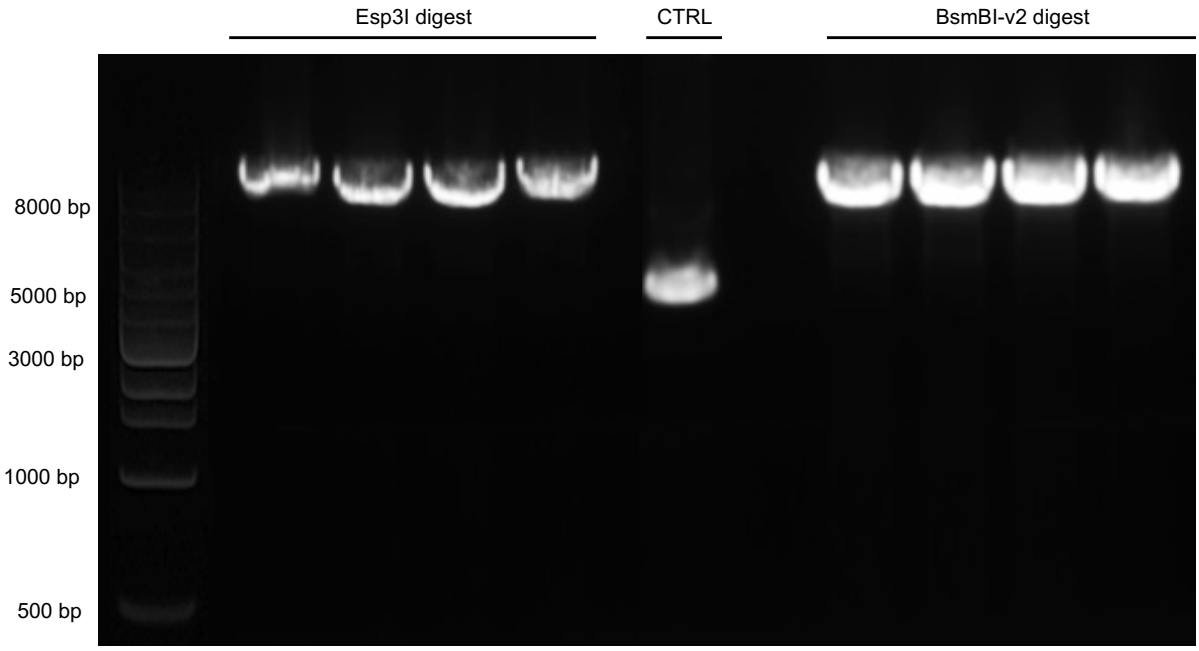


Figure 8: Comparison of restriction enzymes efficiency. *Gel electrophoresis after digest of the CasRx guide RNA backbone using Esp3I (left) and BsmBI-v2 (right). Undigested backbone serves as negative control (CTRL).*

While the use of an Esp3I-digested guide RNA backbone for the ligation did not improve the yield of bacterial colonies obtained it decreased the percentage of bacterial colonies which harboured a re-ligated plasmid backbone instead of a plasmid with the correctly integrated guide RNA sequence. Nevertheless, the cloning protocol needs to be further improved, as only 12 out of 17 guide RNAs could be cloned into the guide RNA backbone during this thesis.

3.3. Generation of single cell clones of HEK 293T cells expressing CasRx

To test the CRISPR-Cas13 system in a human cell line model a construct harbouring GFP and CasRx was stably integrated into the HEK 293T cell line by lentiviral transduction. After FACS sorting of GFP-positive cells, single cell clones were expanded for 6 weeks. To investigate CasRx expression we analysed 25 single cell clones using flow cytometry for GFP. We expected to find homogenous populations of cells which were either GFP-positive or negative. GFP negative cells could have induced silencing of the CasRx vector and therefore not express GFP any longer. On the other hand, a robust GFP signal could report the expression of the

CasRx gene, as it is present on the same vector. Despite GFP-positive cells were sorted after lentiviral introduction of the CasRx vectors, 75 % of all clones investigated showed less than 20 % GFP-positive cells, indicating that most clones had lost CasRx expression. Figure 9 shows the GFP expression of selected HEK 293T + CasRx single cell clones (one exemplary GFP-negative clone and two GFP-positive clones with different patterns of GFP expression). The single cell origin of several clones is unlikely, as more than one cell population could be identified according to the GFP signal (e.g. clone #6 in Figure 9). This could arise from an experimental error where more than one cell was seeded.

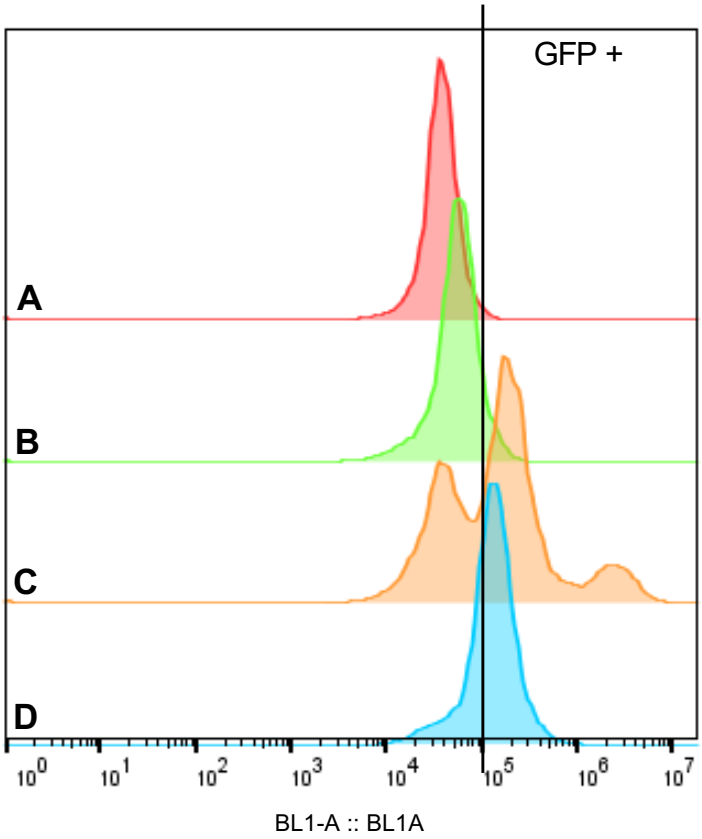


Figure 9: GFP signal of HEK 293T + CasRx single cell clones. *GFP-positive cell populations are shown on the right side of the vertical line (A) GFP-negative control cells. (B) clone #4 which does not express GFP. (C) clone #6 which shows three subpopulations with different GFP expression levels. (D) clone #2 which expresses GFP in a homogenous fashion.*

3.4. RT-qPCR analysis of mRNA levels upon RNA knockdown

The enzymatic activity of CasRx of CasRx positive HEK 293T single cell clones was tested using a total of five guide RNAs targeting the two essential genes *CDK1* and *PLK1*. A guide RNA targeting *AAVS1* was used as a negative control. We selected two GFP-positive clones

for testing. Clone #2 seemingly portraits the ideal candidate for establishing a stable CasRx expressing cell line, as it was the single cell clone with the highest GFP signal. In contrast, clone #6 most likely does not originate from one single cell, as analysing GFP signal revealed three sub-populations of cells with no, intermediate and high GFP signal (Figure 10).

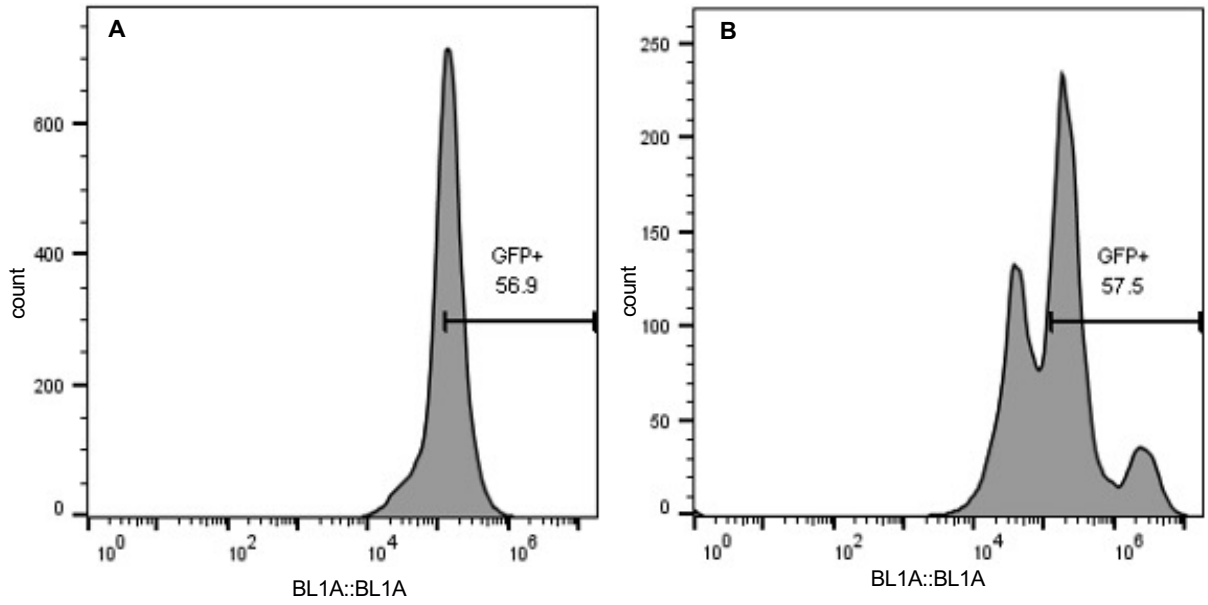


Figure 10: GFP signal of HEK 293T + CasRx single cell clones #2 and #6 *(A) clone # 2 shows a single population of GFP positive cells (B) clone # 6 shows three sub-populations with different GFP expression levels.*

We introduced the guide RNAs via PEI transfection and confirmed their presence using a fluorescence microscope to visualize iRPF670 signal. However, we did not quantify the iRFP670 signal to estimate transfection efficiency, nor did we sort for iRFP670 positive cells. Figure 11 shows that no reduction in mRNA levels upon guide RNA introduction was found. While introduction of guide RNAs targeting either *PLK1* or *CDK1* is expected to cause reduction of mRNA levels of the targeted gene transcript this was not observed for the two single cell clones tested. As the presence of guide RNAs was not quantified a low transfection efficiency may be the reason for the absence of any reduction of target mRNAs. However, it could also be possible that CasRx is either not expressed, expressed at a very low level, or has acquired mutations in either HEPN domain leading to a loss of ribonuclease activity.

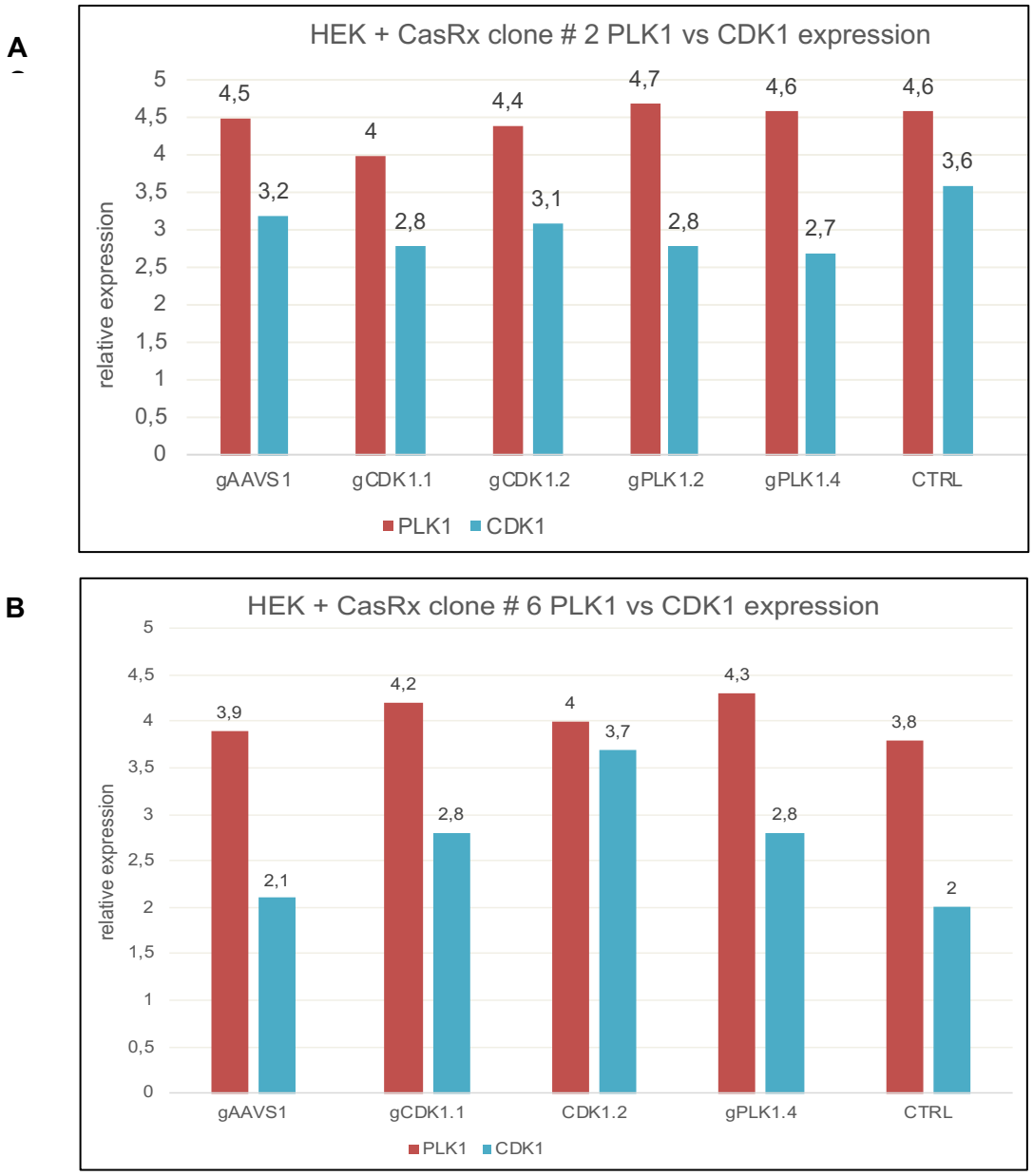


Figure 11: Relative expression level of PLK1 and CDK1 72 hours post transfection of guide RNAs. *Two clones were tested for CasRx efficiency. Cells were transfected with guides against AAVS1, CDK1 and PLK1 and mRNA levels were measured 72 hours post transfection via qPCR. No reduction of mRNA levels of either target gene could be detected.*

3.5. Protein expression level of CasRx in murine *NUP98::KDM5A*-driven cells

To test the system in a cancer-relevant setting, GFP-CasRx (HA-tagged) was lentivirally transduced into *NUP98::KDM5A*-depentent AML cells. After seeding of GFP-positive single cells we obtained nine viable clones. We then performed flow cytometry to analyse GFP expression of single cell clones to assess the expression of CasRx. This analysis showed that

only two clones were positive for GFP (*NUP98::KDM5A* + CasRx clone #3 and #7) (Figure 12). However, we observed two subpopulations for both clones, of which one consists of GFP-negative cells. This could arise from an experimental error of seeding more than one cell into one well. Alternatively, a fraction of the cells in the clone could have lost or silenced CasRx expression during the expansion period, as GFP-negative cell populations may not express CasRx.

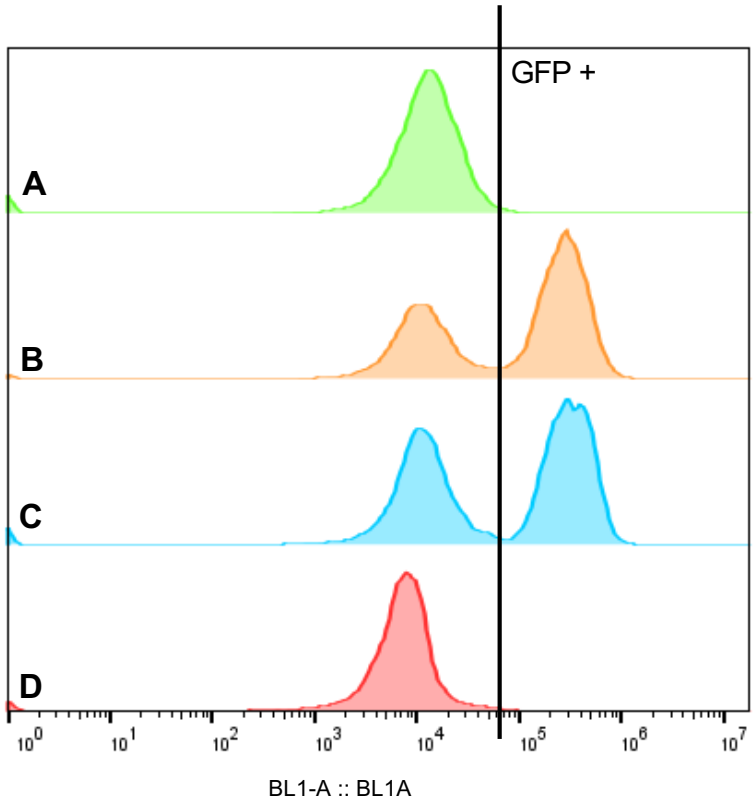


Figure 12: GFP expression of NUP98::KDM5A + CasRx single cell clones. *GFP-positive cell populations are shown on the right site of the line (A) GFP-negative control cells (B) clone #3 (C) clone #7 (D) clone #5*

In addition, we evaluated the presence of CasRx by Western Blotting for five single cell clones. Figure 13 shows CasRx-HA protein expression of single cell clones of murine *NUP98::KDM5A*-driven AML cells. As controls, we used a cell line expressing HA-tagged *NUP98::KDM5A*. Only for clone #5 a faint band at the expected size of the C-terminally HA-tagged CasRx protein (113 kDa) was visible, while no HA signal was detected in the other clones. Although clone #5 was GFP-negative, the presence of this band could indicate expression of the C-terminal HA tagged CasRx protein. However, repetition of this experiment for the sample CasRX-HA clone #5 is necessary to confirm CasRx expression.

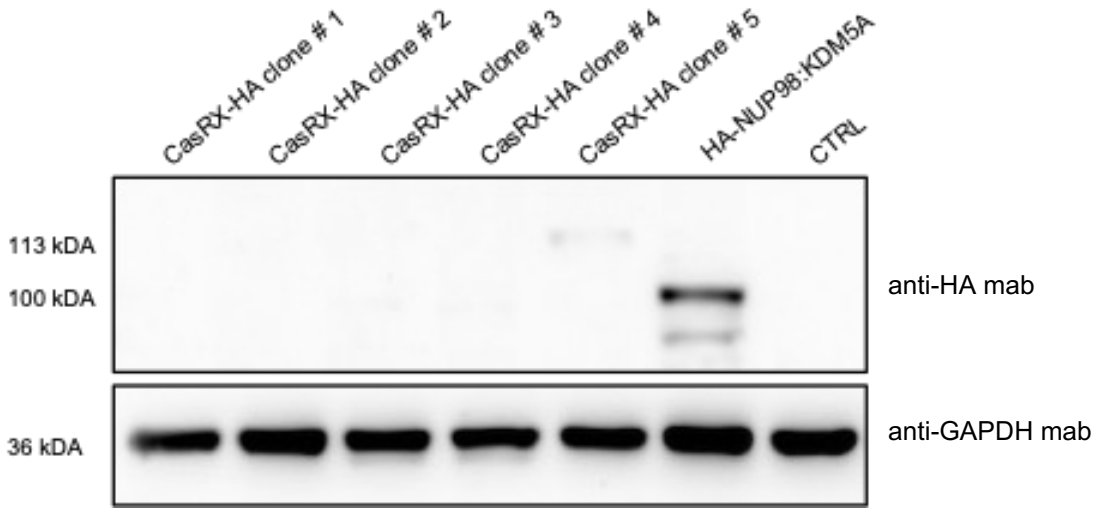


Figure 13: Protein expression of CasRx in murine AML cells. *Protein lysates were subjected to Western Blotting and the membrane was stained with an anti-HA-tag antibody and anti-GAPDH antibody as loading control.*

4. Discussion

Biomolecular condensation has been shown to play an important role in leukemogenesis in NUP98-fusion oncoprotein driven leukemia. However, little is known about the changes in the composition of biomolecular condensates upon expression of *NUP98*-fusions. As condensate composition can be influenced by RNA and proteins with particular structural features such as IDRs, we wanted to establish an *in-vitro* model to study the effects of targeted RNA knockdown in murine *NUP98::KDM5A*-dependent AML cells to gain an understanding of crucial components of NUP98-fusion oncoprotein-driven biomolecular condensates.

RNAi is frequently used to induce targeted downregulation of RNA molecules. However, we chose to establish a cellular model which expresses CasRx to investigate RNA knockdown by using the CRISPR-Cas13 system. While Cas13 has only been identified in 2015, several studies suggest advantages of this approach over the conventional RNAi technology. Cas13 proteins can be fused to a nuclear localization signal, which enables the knockdown of nuclear RNAs such as non-coding RNAs, which cannot be efficiently targeted using RNAi. Furthermore, Cas13-mediated RNA knockdown was shown to be more efficient if compared to shRNAs, while causing less off-target effects [31]. In an attempt to identify the ideal strategy for guide RNA design to induce targeted RNA knockdown using Cas13d, a screen using >20,000 guide RNAs was performed by Wessels *et. al*. It was revealed that for a core region of the guide RNA a single mismatch is sufficient to abolish RNA cleavage [32]. Furthermore, guide RNAs shorter than 20 nucleotides have been deemed inactive as little RNA cleavage was observed when shorter guide RNAs were used [33] [25]. When comparing PguCas13b, PspCas13b and RfxCas13d (CasRx), CasRx showed the strongest target knockdown, particularly when its fusion to a nuclear localization signal [32].

Furthermore, Cas13 proteins are attractive candidates for other applications. A catalytically dead Cas13 fused to a GFP tag has been utilized for RNA imaging in live cells as only the cleavage capability, but not the capability to bind to target sequences was lost [34]. Furthermore, guide RNAs with single base mismatches can be used in the RNA editing Programmable A to I Replacement (REPAIR) system, in which Cas13 is fused to an adenosine deaminase acting on RNA deaminase (ADAR2) domain to edit gene transcripts of interest [35]. Therefore, not only the introduction of CasRx, but also of dCasRx is of interest.

However, other studies reported off-target effects of RNA knockdown with CasRx. Activation of the HEPN domains, which can be located at the surface of the CasRx protein, can lead to

unspecific cleavage of bystander RNAs, hence causing off-target effects [29]. While such collateral cleavage activity is known to occur in bacteria and *in-vitro,* it was believed that no off-target cleavage occurs in mammalian cells. Ai *et al*. investigated potential off-target effects of CasRx and PsbCas13b, which both have been reported to show high sensitivity in mammalian cells. While the authors performed the majority of their study in *Drosophila* cells, they also compared their findings to results from HEK 293T and HeLa cells. These data provide clear evidence for off-target RNA cleavage if highly expressed RNAs are targeted [36].
For this work, we choose the CasRx-2A-eGFP and dCasRx-2A-eGFP plasmids to introduce the CasRx protein into our target cell lines. As only few studies which show CasRx being used for targeted RNA knockdown are available so far, we decided to first introduce the CasRx system into HEK 293T cells to establish a robust workflow to validate this system.

The plasmid used for CasRx expression also includes a gene that encodes eGFP. This allowed us to perform FACS to sort for cells which emit a GFP signal in order to exclude any cells which were not successfully transduced. However, also other selection genes are available. IN some instances, the use of Blasticidin-S or Puromycin resistance genes might be more efficient. While FACS generally provides an efficient selection of cells which contain the desired plasmid there might be a need for a different selection method when CasRx is introduced into cells. In particular for the visualization of biomolecular condensates one might need to use antibiotic resistance genes instead of eGFP to select cells which express CasRx, because the presence of GFP limits the range of antibodies for immunofluorescence and the amount of stainings which can be performed simultaneously in this cell line. Furthermore, it is possible that the GFP signal does not correlate to the expression levels of CasRx within cells. Therefore, a different plasmid which includes an antibiotic resistance gene instead of eGFP could have been used.

To identify a single cell clone which shows high enzymatic activity of CasRx RT-qPCR was performed to analyse mRNA levels of two target genes 72 hours after introduction of guide RNAs. While five CasRx-expressing HEK 293T clones featuring different levels of GFP expression were transfected with guide RNAs, only two CasRx-expressing clones were used for RT-qPCR analysis due to material shortage and time reasons. Both single cell clones tested did not show any reduction in mRNA levels upon guide RNA introduction. One possible reason might be that CasRx is not expressed in these cells, or that CasRx expression levels are too low for efficient RNA knockdown. However, out of all 25 single cell clones analysed the clones

used in these experiments showed a high percentage of GFP positive cells. As GFP is co-expressed with CasRx from the same promotor, expression of GFP is expected to be closely correlated to CasRx expression. Nevertheless, despite all cells emitting a GFP signal, the expression level of CasRx might still be too low to induce RNA cleavage. Another reason could be a low transfection efficiency of guide RNAs. This cannot be ruled out as the iRFP670 signal was low when transfected cells were analysed for iRFP670 under the fluorescence microscope. While HEK 293T cells can be transfected very efficiently using transfection agents such as PEI, one would need to analyse the actual transfection efficiency to correctly interpret the qPCR results. Therefore, the transfection should be repeated to estimate the transfection efficiency. Alternatively, the guide RNA plasmids could also be introduced by lentiviral transfection instead of using PEI transfection. In addition, cells could be sorted for iRFP670 positive cells to exclude non-transfected cells, which would mask the reduction of mRNA levels. However, it is known that point mutations in the HEPN domains cause the CasRx protein to lose its enzymatic activity. Although unlikely, it is possible that such a mutation was acquired during the cultivation period in both single cell clones tested and this could also be a possible reason for the lack of mRNA degradation. In any case, more guide RNAs should be tested to identify guide RNAs which can induce reliable RNA knockdowns, although this has been performed for HEK293 cells already [32].

As a novel guide RNA expression vector was created in this work, it remains unclear if this new construct contains all features to efficiently induce the expression of CasRx guide RNAs in cells. While this vector was designed to include all known regulatory elements necessary for a CasRx guide RNA expression vector, however, as little research has been done using CasRx for targeted RNA knockdown, there may be more elements which were not considered. Validation of the vector to confirm its ability to allow stable expression of guide RNAs is yet to be done.

As mentioned before, low expression levels of CasRx could explain the lack of mRNA degradation. Therefore, analysis of CasRx protein expression could be performed. This was attempted for murine AML cells, which were transduced with the same CasRx-2A-eGFP plasmid that was also used for the HEK 293T cells using Western Blot. The CasRx protein has a C-terminal HA-tag which can be used to visualize the protein. While GFP expression was only found in clone #3, the presence of the HA tag could not be found in this clone by Western Blotting. In contrast, for single cell clone #5, which did not show any GFP signal, a band at the

expected size of 113 kDa was detectable. Nevertheless, the interpretation of protein expression level remains questionable as none of the other samples can be used for a comparison of protein levels. Thus, the Western Blot would need to be repeated to confirm the presence of an HA-tagged protein of 113 kDa in the lysate of CasRx-expressing clone #5.

Seeding of HEK 293T single cell clones yielded a total of 45 viable clones. While no single cell clone with high enzymatic activity of CasRx could be identified so far, the majority of single cell clones has not been tested yet. As protein expression is essential for enzymatic activity and may not fully correlate with GFP expression one could either optimize the antibody staining for Western Blot or measure CasRx mRNA levels using RT-qPCR to select clones which show the highest CasRx mRNA levels for further testing. While CasRx-mediated RNA knockdown appears to be a reasonable approach to study the contributions of specific RNA molecules to biomolecular condensates mediated by NUP98-fusion proteins, further testing of the available single cell clones is necessary to evaluate the enzymatic activity to establish a functional CasRx expressing cell line. In HEK 293T cells CasRx mediated RNA knockdown was shown to be highly specific, it remains to be investigated if this is also true for murine *NUP98::KDM5A*-driven AML cells. Otherwise, the use of other Cas13 proteins might need to be taken into consideration.

In summary, we stably introduced CasRx as well as a catalytically inactive CasRx mutant into HEK 293T cells and *NUP98::KDM5A*-dependend murine AML cells by lentiviral transduction. We expanded clones to establish cell lines of single cell origin. As CasRx expression is linked to GFP, we performed flow cytometry to investigate expression level and enzymatic activity of CasRx in selected single cell clones. We selected two HEK 293T + CasRx single cell clones to test five guide RNAs targeting either the AAVS1 locus, *CDK1* or *PLK1*. For the introduction of these guide RNAs we created a novel expression vector which consists of all necessary elements for CasRx mediated RNA knockdown that are known so far. However, we were unable to show any reduction of mRNA levels of either target gene using RT-qPCR. A validation of our guide RNA expression vector is yet to be done, as we could not prove its functionality in this thesis. Furthermore, we investigated expression levels of CasRx in murine *NUP98::KDM5A*-driven AML cells using Western Blot and identified one single cell clone, which might express CasRx. Further experiments will show why no mRNA degradation could be achieved with the single cell clones we tested. More clones will need to be evaluated to establish a system for reliable CasRx mediated RNA knockdown.

5. References

[1] Short NJ, Rytting ME, Cortes JE. Acute Myeloid Leukemia. Lancet. 2018; 392 (10147): 593-606. DOI: 10.1016/S0140-6736(18)31041-9

[2] Chopra M, Bohlander SK. The cell of origin and the leukemia stem cell in acute myeloid leukemia. Genes Chromosomes Cancer. 2019;58(12):850-858. DOI: 10.1002/gcc.22805

[3] Häggström M. Medical gallery of Mikael Häggsröm. Wikiversity Journal of Medicine. 2014; 1 (8). DOI: 10.15347/wjm/2014.008

[4] Khoury JD, Solary E, Abla O, Akkari Y, Alaggio R, Apperley JF *et al*. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myleoid and Histiocytic/Dendritic Neoplasms. Leukemia. 2022;36(7):1703-1719. DOI: 10.1038/s41375-022-01613-1

[5] Puumala SE, Ross JA, Aplenc R, Spector LG. Epidemiology of childhood acute myeloid leukemia. Pediatr Blood Cancer. 2013;60(5): 728-33. DOI: 10.1002/pbc.24464

[6] Bailey C, Richardson LC, Allemani C, Bonaventure A, Harewood R, Moore AR *et al.* Adult leukemia survival trends in the United States by subtype: A population-based registry study of 270,994 patients diagnosed during 1995-2009. Cancer. 2018;124(19): 3856-3867. DOI: 10.1002/cncr.31674

[7] Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. New England Journal of Medicine. 2016;374(23):2209-2221. DOI: 10.1056/NEJMoa1516192

[8] Michmerhuizen NL, Klco JM, Mullighan CG. Mechanistic insights and potential therapeutic approaches for NUP98-rearranged hematologic malignancies. Blood. 2020;136(20): 2275-2289. DOI: 10.1182/blood.2020007093

[9] Thol F, Ganser A. Treatment of Relapsed Acute Myeloid Leukemia. Current Treatment Options Oncology. 2020;21(8):66. DOI: 10.1007/s11864-020-00765-5

[10] Barresi V, Di Bella V, Andriano N, Privitera AP, Bonaccorso P, La Rosa M *et al.* NUP-98 Rearrangements Led to the Identification of Candidate Biomarkers for Primary Induction Failure in Pediatric Acute Myeloid Leukemia. International Journal of Molecular Science. 2021;22(9): 4575. DOI: 10.3390/ijms22094575

[11] Gough SM, Slape CI, Aplan PD. NUP98 gene fusions and hematopoietic malignancies: common themes and new biologic insights. Blood. 2011;118(24): 6247-57. DOI: 10.1182/blood-2011-07-328880

[12] Capitanio JS, Montpetit B, Wozniak RW. Human Nup98 regulates the localization and activity of DExH/D-box helicase DHX9. Elife. 2016;6:e18825. DOI: 10.7554/eLife.18825

[13] Yang GJ, Zhu MH, Lu XJ, Liu YJ, Lu JF, Leung CH *et al.* The emerging role of KDM5A in human cancer. Journal of Hematology and Oncology. 2021;14(1):30. DOI: 10.1186/s13045-021-01041-1

[14] Terlecki-Zaniewicz S, Humer T, Eder T, Schmoellerl J, Heyes E, Manhart G *et al.* Biomolecular condensation of NUP98 fusion proteins drives leukemogenic gene expression. Nature Structual & Molecular Biology. 2021;28(2):190-201. DOI: 10.1038/s41594-020-00550-w

[15] Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. Nature Reviews Molecular Cell Biology. 2017;18(5):285-298. DOI: 10.1038/nrm.2017.7

[16] Cai D, Liu Z, Lippincott-Schwartz J. Biomolecular Condensates and Their Links to Cancer Progression. Trends in Biochemical Sciences. 2021;46(7): 535-549. DOI: 10.1016/j.tibs.2021.01.002

[17] Chandra B, Michmerhuizen NL, Shirnekhi HK, Tripathi S, Pioso BJ, Baggett DW *et al.* Phase Separation Mediates NUP98 Fusion Oncoprotein Leukemic Transformation. Cancer Discovery. 2022;12(4):1152-1169. DOI: 10.1158/2159-8290.CD-21-0674.

[18] Ahn JH, Davis ES, Daugird TA, Zhao S, Quiroga IY, Uryu H *et al.* Phase separation drives aberrant chromatin looping and cancer development. Nature. 2021;595(7868): 591-595. DOI: 10.1038/s41586-021-03662-5

[19] Roden C, Gladfelter AS. RNA contributions to the form and function of biomolecular condensates. Nature Review Molecular Cell Biology. 2021;22(3):183-195. DOI: 10.1038/s41580-020-0264-6

[20] Huynh N, Depner N, Larson R, King-Jones K. A versatile toolkit for CRISPR-Cas13-based RNA manipulation in Drosophila. Genome Biol. 2020; 21(1):279. DOI: 10.1186/s13059-020-02193-y

[21] Yu JY, DeRuiter SL, Turner DL. RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99(9): 6047-52. DOI: 10.1073/pnas.092143499

[22] Xu W, Jiang X, Huang L. RNA Interference Technology. Comprehensive Biotechnology. 2019:560–75. DOI: 10.1016/B978-0-444-64046-8.00282-2

[23] Zeng Y, Cullen BR. RNA interference in human cells is restricted to the cytoplasm. RNA. 2002;8(7):855-60. DOI: 10.1017/s1355838202020071

[24] Fapohunda FO, Qiao S, Pan Y, Wang H, Liu Y, Chen Q, Lü P. CRISPR Cas system: A strategic approach in detection of nucleic acids. Microbiology Research. 2022;259:127000. DOI: 10.1016/j.micres.2022.127000

[25] Konermann S, Lotfy P, Brideau NJ, Oki J, Shokhirev MN, Hsu PD. Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. Cell. 2018;173(3): 665-676.e14. DOI: 10.1016/j.cell.2018.02.033

[26] Addgene. CRISPR plasmids RNA targeting. 2020. [online]. Available at: https://www.addgene.org/crispr/rna-targeting/ (accessed: 15/09/2022, 12.43 pm)

[27] Zhang B, Ye Y, Ye W, Perčulija V, Jiang H, Chen Y *et al.* Two HEPN domains dictate CRISPR RNA maturation and target cleavage in Cas13d. Nat Commun. 2019 ;10(1): 2544. DOI 10.1038/s41467-019-10507-3

[28] Abudayyeh O, Gootenberg J.Tips and Tricks for Cas13.2017. [online]. Available at: https://zlab.bio/cas13 (accessed: 15/09/2022, 11.29 am)

[29] O'Connell MR. Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR-Cas Systems. Journal of Molecular Biology. 2019;431(1):66-87. DOI: 10.1016/j.jmb.2018.06.029

[30] Wei J, Lotfy P, Faizi K, Wang E *et al.* Deep learning and CRISPR-Cas13d ortholog discovery for optimized RNA targeting. bioRxiv. [pre-print]. 2022. Available at: https://doi.org/10.1101/2021.09.14.460134 (accessed: 25/10/2022, 2.18 pm)

[31] Saifullah, Sakari M, Suzuki T, Yano S, Tsukahara T. Effective RNA Knockdown Using CRISPR-Cas13a and Molecular Targeting of the EML4-ALK Transcript in H3122 Lung Cancer Cells. Int J Mol Sci. 2020;21(23):8904. DOI: 10.3390/ijms21238904.

[32] Wessels HH, Méndez-Mancilla A, Guo X, Legut M, Daniloski Z, Sanjana NE. Massively parallel Cas13 screens reveal principles for guide RNA design. Nature Biotechnology. 2020;38(6): 722-727. DOI: 10.1038/s41587-020-0456-9

[33] Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ *et al.* RNA targeting with CRISPR-Cas13. Nature. 2017;550(7675): 280-284. DOI: 10.1038/nature24049

[34] Yang LZ, Wang Y, Li SQ, Yao RW, Luan PF, Wu H *et. al.* Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 Systems. Molecular Cell. 2019;76(6):981-997.e7. DOI: 10.1016/j.molcel.2019.10.024

[35] Cox DBT, Gootenberg JS, Abudayyeh OO, Franklin B, Kellner MJ, Joung J, Zhang F. RNA editing with CRISPR-Cas13. Science. 2017;358(6366):1019-1027. DOI: 10.1126/science.aaq0180

[36] Ai Y, Liang D, Wilusz JE. CRISPR/Cas13 effectors have differing extents of off-target effects that limit their utility in eukaryotic cells. Nucleic Acids Research. 2022;50(11): e65. DOI: 10.1093/nar/gkac15

6. List of Figures

7. List of Tables