Towards the Characterization of lncRNA Interactomes with RNA-Proximity Labeling

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Ronja Weissinger
aus Öhringen

Tübingen
2025

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 29.07.2025

Dekan: Prof. Dr. Thilo Stehle
1. Berichterstatter: Prof. Dr. Ralf-Peter Jansen
2. Berichterstatter: Prof. Dr. Thorsten Stafforst

Abstract

Post-transcriptional regulation of ribonucleic acid (RNA) is mainly orchestrated by the dynamic interplay between the RNA molecule itself and the interacting proteins. Through these interactions, RNAs can be transported to distinct cellular places, sequestered away into granules or, among other things, actively translated into proteins. Together with the RNA they bind to, RNA interacting proteins form ribonucleoprotein complexes (RNPs), which are reasoned to be the true regulatory units for post-transcriptional regulation. Identifying the binding partners of RNA thus provides valuable details to better understand their function.
Multiple techniques have been developed to characterize the interactions between RNAs and proteins. A more recent approach adopts a method initially developed to study protein-protein interaction: proximity labeling. The elegant RNA-BioID assay enables the detection of RNA–protein interactions and was presented by Mukherjee et al. in 2019 [1]. Starting from this publication, this thesis aims to advance, apply and follow-up on the RNA-BioID method.

The main aim of this study was to take the RNA-BioID one step further by applying it to a lesser studied group of RNAs, the so-called long non-coding ribonucleic acid (lncRNA). By setting up a similar assay as Mukherjee and colleagues, the hope was to gain insight into the protein interactome of two lncRNA candidates (NORAD and OIP5-AS1). As a first step, the fusion construct, including the labeling enzyme necessary for proximity labeling, was stably introduced into HeLa cells. This was achieved for two different labeling enzymes (microID and TurboID) and their performance was compared. In the second step, the endogenous genes of the two lncRNA candidates was tried to manipulate with several different CRISPR/Cas9 strategies. As this failed repeatedly, the approach was changed to an *in vitro* proximity labeling method to avoid the necessity of genetic manipulation of the cells. After expression and purification of the proximity labeling enzyme, together with DNA probes that should direct it to the correct RNA target, it was applied to fixed and permeabilized HeLa cells. Prior, it was shown that the probes which finally tether the enzyme to the RNA are correctly localized. The enzyme did so too, given the right probe concentration. However, having tested several different conditions, the labeling still did not appear at the expected position. Therefore, the approach was not further continued.

The last part of this thesis aimed to follow-up on the initial RNA-BioID assay. The β-actin mRNA interactome has been detected in an unbiased screen in mouse embryonic fibroblast cells. Several protein candidates which were found as potential interactors have been selected for verification (hnRNP K, TDP-43, PTBP1, and hnRNP R) to provide insights into their biological relevance. The interaction was analyzed by pull-down of the protein and RT-qPCR of the co-purified RNAs. If β-actin mRNA is bound by the RNA-binding protein, it is pulled down along with the protein. A successful enrichment of β-actin mRNA could be shown for PTBP1 and hnRNP R.

Zusammenfassung

Die posttranskriptionelle Regulierung von Ribonukleinsäure (RNA) erfolgt hauptsächlich durch ein Zusammenspiel zwischen der RNA selbst und den interagierenden Proteinen. RNAs werden dabei an verschiedene zelluläre Orte transportiert, in Granula gespeichert oder u.a. in Proteine translatiert. Zusammen mit den RNAs, an die sie binden, bilden RNA-interagierende Proteine Ribonukleoprotein Komplexe (RNPs), der als die eigentliche funktionelle Einheit für die posttranskriptionelle Regulation angesehen wird. Die Identifizierung der Bindungspartner von RNAs liefert daher wertvolle Details zum besseren Verständnis ihre Funktion.

Es wurden mehrere Techniken entwickelt, um die Bindungen und Interaktionen zwischen RNAs und Proteinen zu charakterisieren. Ein neuerer Ansatz nutzt eine Methode, die ursprünglich für die Untersuchung von Protein-Protein-Interaktionen entwickelt wurde: Proximity Labeling. Der elegante RNA-BioID Assay ermöglicht die Identifikation von RNA-bindenden-Proteinen und wurde 2019 von Mukherjee et al. vorgestellt [1]. Ausgehend von dieser Veröffentlichung zielt diese Arbeit darauf ab, die RNA-BioID Methode weiterzuentwickeln und anzuwenden, sowie die bisher gewonnenen Ergebnisse weiterzuverfolgen.
Das Hauptziel dieser Studie war es, die RNA-BioID Methode einen Schritt weiter zu entwickeln, indem sie auf eine der weniger untersuchten Gruppen von RNAs, die so genannten langen nicht-kodierenden Ribonukleinsäuren (lncRNA), anzuwenden. Mit einen ähnlichen Assay wie von Mukherjee und Kollegen vorgestellt, war das Ziel, einen Einblick in das Protein-Interaktom von zwei lncRNA-Kandidaten (NORAD und OIP5-AS1) zu erhalten. In einem ersten Schritt wurde das Fusionskonstrukt einschließlich des für das Proximity-Labeling erforderliche Enzym in HeLa-Zellen integriert. Dies geschah mit zwei verschiedenen Enzymen (microID und TurboID) und ihre Perfomance wurde verglichen. In einem zweiten Schritt wurde versucht die endogenen Gene der beiden lncRNA-Kandidaten mit mehreren unterschiedlichen CRISPR/Cas9 Strategien zu manipulieren. Da dies wiederholt fehlschlug, wurde der Ansatz auf eine *in-vitro*-Methode umgestellt, um die Notwendigkeit einer genetischen Manipulation der Zellen zu vermeiden. Nach Expression des Proximity-Labeling Enzyms und dessen Aufreinigung wurde es auf fixierte und permeabilisierte Zellen angewendet. Zuvor wurde gezeigt, dass die Sonden, die das Enzym schließlich an die RNA binden, korrekt lokalisiert sind. Nachdem nachgewisen war, dass auch das Enzym korrekt rekrutiert wird, wurden verschiedene Bedinungnen für das Proximity Labeling getestet. Das Labeling Signal wurde jedoch unter keinen Umständen an der erwarteten Stelle nachgewiesen. Daher wurde der Ansatz nicht weiter verfolgt.

Der letzte Teil dieser Arbeit zielte auf die Weiterverfolgung des ursprünglichen RNA-BioID Assays ab. Das β-Actin mRNA Interaktom wurde in embryonalen Fibroblastenzellen der Maus ermittelt. Mehrere Protein-Kandidaten, die als potenzielle Interaktoren gefunden wurden, wurden zur Verifizierung ausgewählt (hnRNP K, TDP-43, PTBP1 und hnRNP R), um ihre biologische Relevanz besser zu verstehen. Die Interaktion wurde durch Pull-down des RNA-bindenden-Proteins und RT-qPCR der mit aufgereinigten und ange-reicherten RNAs nachgewiesen. Eine erfolgreiche Anreicherung von β-Actin-mRNA konnte für PTBP1 und hnRNP R gezeigt werden.

Abbreviations

Amp	ampicillin
AP	alkaline phosphatase
APEX	ascorbate peroxidase
APS	ammonium persulfate
ATP	adenosine triphosphate
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary DNA
CDS	coding sequence
CMV	Cytomegalovirus
CRISPR	clustered regularly interspaced short palindromic repeats
DIG	digoxygenin
DMEM	Dulbecco's Modified Eagle Medium
DNA	desoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
ds	double strand
DSB	DNA double-strand break
DTT	dithiothreitole
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
E.coli	*Escherichia coli*
EtOH	ethanol
F	forward
FACS	fluorescence activated cell sorting
FCS	fetal calf serum
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
gDNA	genomic DNA
GFP	green fluorescent protein
h	hour
HA	hemagglutinin
HDR	homology-directed repair
HeLa	Henrietta Lacks
hnRNP	heterogeneous nuclear ribonucleoprotein
HRP	horseradish peroxidase
IF	immunofluorescence
Kan	kanamycin
kb	kilo basepairs
kDa	kilo Dalton
LB	Lucria Broth
lncRNA	long non-coding ribonucleic acid
MCP	MS2 coat protein
MEF	mouse embryonic fibroblast
MHA	microhomology arms
MMEJ	microhomology-mediated end joining
min	minute
mRNA	messenger ribonucleic acid
MS	mass spectrometry
NES	nuclear export signal
NHEJ	non-homologous end joining
NLS	nuclear localization signal
nt	nucleotide
OD_{600}	optical density at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PL	proximity labeling
POI	protein of interest
PS	phosphorothioate
R	reverse
RBP	RNA-binding protein
RBD	RNA-binding domains
RMCE	recombinase-mediated cassette exchange
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonucleoprotein complex
rpm	rounds per minute
RRM	RNA recognition motif
RT	room temperature
SA	Streptavidin
SDS	sodium dodecylsulfate
sgRNA	single guide RNA
smFISH	single-molecule fluorescence in-situ hybridization
ss	single strand
TEMED	tetramethylethylenediamine
UTR	untranslated region
WT	wild type

Contents

1 Overview

1.1. General introduction

Ever since the COVID-19 pandemic, ribonucleic acid (RNA) has become known to the general public. While this was not a merely positive stage that RNA entered in the form of mRNA vaccines, nonetheless it was for sure the most prominent public stage RNA got so far. For scientist, this of course did not come by surprise. Over the past two decades, RNA has increasingly received the scientific attention it deserves. Another example illustrating the growing research interest in RNA is the 2024 Nobel Prize "for the discovery of microRNA and its role in post-transcriptional gene regulation" awarded to Victor Ambros and Gary Ruvkun [2].
Their work reflects a still emerging field of research, exploring the different species of RNA like miRNA, lncRNA or circRNA, just to name a few. With the discovery of these new RNA types, naturally questions arise: Why do we need those molecules, and what are their functions? The answers to these questions have been and continue to be heavily investigated. Ultimately, this research gives us the capability to apply the gained knowledge, as demonstrated by the rapid development of the mRNA vaccines against the SARS-CoV-2 virus, which, again, won a Nobel Prize in 2023 "for their discoveries concerning nucleoside base modifications that enabled the development of effective mRNA vaccines against COVID-19" for Katalin Karikó and Drew Weissman [2].

In order to understand the function and fate of different RNA types and classes, RNA-binding proteins (RBPs) play an important role. These RBPs influence the complete life cycle of RNAs from maturation and distribution within cells to degradation [3], [4]. While some RBPs are specific for only a subgroup of RNAs, others are rather unspecific and attach to even a subclass of RNAs. The complexes formed by RNAs and RBPs are called ribonucleoprotein complexes (RNPs) and can be of transient or stable nature. The study of those complexes, or from the

RNA's perspective, the RNA interactome, is of major interest as they serve as functional units for post-transcriptional regulation [5], [6].
This thesis attempts to further apply and advance strategies to understand the RNA interactome. In particular, it builds upon the RNA-BioID assay developed by Mukherjee et al. in 2019, which was used to perform an unbiased screen of β-actin mRNA interactors [1].

1.2. Structure of this thesis

In the following section, an overview of the structure of this thesis is given, which consists of three parts. These are broken down as follows:
Part I introduces the background required for the later chapters and related work of this thesis. At first, ribonucleic acid (RNA) and RNA-binding protein (RBP) are characterized. The RNA interactome is defined in the next section, followed by the introduction of the concept of proximity labeling and the RNA-BioID assay. This will form a key theme in the rest of this thesis, where each of the following chapters will focus on different aspects of this method.
Part II provides an overview of all materials used and methods applied during this work.
Part III comprises the experimental work done for this thesis.
In chapter 8 the main project of this thesis, establishing an *in vivo* proximity labeling assay for lncRNA, is presented. It begins with a brief description of the subclass of lncRNA. To introduce an essential part for the RNA-BioID assay, the DNA tagging system CRISPR/Cas9 is described in more detail. In the results part, first the stable introduction of the biotin ligases TurboID and microID into HeLa cells, including the characterization of the new cell lines, is covered. The second part of the section elaborates on the different CRISPR strategies that have been tried to manipulate the endogenous genes of two different lncRNA candidates.
The chapter 9 focuses on the establishment of an *in vitro* proximity labeling assay for lncRNA. In the beginning, a third lncRNA candidate is introduced, next to a distinct labeling method so far not discussed: the biotin labeling by peroxidases instead of ligases. It follows a slightly different mechanism, but leads to a similar outcome. The *in vitro* approach described here comes with the advantage that the manipulation of endogenous genes is not necessary, but naturally comes with its own pitfalls, which will be further described and shown by microscopy in the results part.
The chapter 10 takes a different perspective and follows the result of the first RNA-BioID experiment. Therefore, the most commonly known RNA class, mRNAs, is discussed in more detail. Based on that, the RNA of interest, β-actin mRNA, together with its protein interactome found in the RNA-BioID screen, is introduced. To validate candidates of interest, co-immunoprecipitations and eventually RT-qPCRs are conducted.
The final chapter 11 summarizes the content of the previous three chapters critically and suggests future directions for the work. A so far undiscussed variant of proximity labeling is proposed to improve the method.

Part I.

Background

2 RNA - characterization and life cycle

2.1. RNA chemistry

Ribonucleic acid (RNA) is a fundamental biomolecule composed of a sequence of four different nucleotides. Each nucleotide consists of a ribose sugar, a phosphate group, and one of four nitrogenous bases: adenine (A), cytosine (C), guanine (G), or uracil (U). The ribose serves as a linker, and is connected to the phosphate group via the 5'C atom while the 1'C is bound to the base. The four nucleotides constitute the monomers of the linear RNA polymer. RNA is typically single-stranded, unlike deoxyribonucleic acid (DNA), which is double-stranded. RNA also utilizes uracil instead of thymine as a base, and has a hydroxyl (-OH) group attached to the 2' carbon of its ribose. This makes RNA chemically less stable than DNA, as the free hydroxyl group can lead to self-cleavage because it is prone to hydrolysis. The phosphate group is not only linking the monomers together, it also gives the whole molecule a negative charge, a feature which RNA and DNA share [3].
These chemical properties give RNA the unique capability to both store information and catalyze reactions, which lead to the *RNA world hypothesis*. This hypothesis, formulated in 1986, proposes that early life on earth relied solely on RNA [7]. Supporting this hypothesis is RNA's ability to adopt complex secondary and tertiary structures through base pairing interactions. Thereby, double-stranded RNA, even consisting of two independent RNA molecules, although rare in eukaryotes, can form. Double-stranded RNA can form helices and often includes wobble pairing e.g. between G/U. Classical secondary structures are hairpins or stem-loop structures with a 5' and 3' flanking region, including internal loops/ bulges, which occur through mismatched regions in the stem structure or pseudoknots [8]. This intricate 3D structure enables some RNA molecules to perform enzymatic functions, an examples being the evolutionary conserved rRNA in ribosomes [3].

2.2. RNA classes

A huge variety of RNA types exists, they are divided into various classes, with new species still being discovered. Each type has distinct and sometimes not well-understood roles in cellular processes. Messenger RNA (mRNA) is the most studied and commonly known class of RNA. It carries genetic information transcribed from DNA and serves as a template for protein synthesis during translation. This makes it the main protein-coding RNA.
After the human genome project came to an end in the early 2000s, the discrepancy between the number of protein-coding genes in humans (∼25,000) compared to the length of the total DNA became apparent [9]. The debate and research about the surplus "junk" DNA boosted the non-coding RNA (ncRNA) field. In contrast to mRNA, ncRNAs are usually not translated into proteins. They can be subdivided based on function and size (see Figure 1). The best understood classes are the so-called housekeeping RNAs, transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA). They are typically expressed constitutively because their function is required in all cells to maintain core cellular activities [10].
tRNA deliver specific amino acids to the ribosome, matching mRNA codons with corresponding amino acids, thereby enabling the synthesis of the protein primary sequence. Each tRNA is ∼75 nt long and has a distinct secondary structure containing several modified nucleotides. rRNAs are key components of ribosomes. Similar to tRNAs, they have a very defined structure, which determines the shape of the ribosome. Together with several proteins that help stabilizing the huge complex, they build the machinery for translating mRNA to protein. The third housekeeping ncRNA is snRNA, an integral component of the spliceosome. The spliceosome complex is responsible for carrying out the splicing of pre-mRNA, a process that removes introns from the nascent RNA transcript and joins exons together to form mature mRNA [3]. The discovery of the splicing mechanism also helped to explain a further mystery revealed by the human genome project: the difference between the ∼25,000 protein-coding genes in humans compared to the many more proteins found in the cell (>90,000) [11]. The last class of housekeeping ncRNA is snoRNA. snoRNAs are primarily involved in the chemical modification and processing of rRNA, snRNA and sometimes even tRNA in the nucleolus. The modifications are essential for the maturation, stability and proper function of these RNAs in translation and splicing, making snoRNAs crucial for fundamental cellular processes [3].

On the other hand, there are regulatory non-coding RNAs, which are further classified into two groups based on their length. RNAs with less than 200 nucleotides are mostly involved in gene regulation, while those with more than 200 nucleotides fulfill a variety of functions.
One example of short regulatory ncRNAs are microRNAs (miRNAs). They are typically 21-23 nucleotides in length and regulate gene expression post-transcriptionally by binding to their target sequence within mRNAs. This often leads to their translational repression or degradation. Thereby, miRNAs specifically fine-tune the gene expression of their target mRNAs.
Another class of small, double-stranded RNA molecules are small interfering RNAs (siRNAs), which are usually about 20-25 base pairs in length. They are an important player in the RNA interference (RNAi) pathway, where they guide the degradation of complementary target

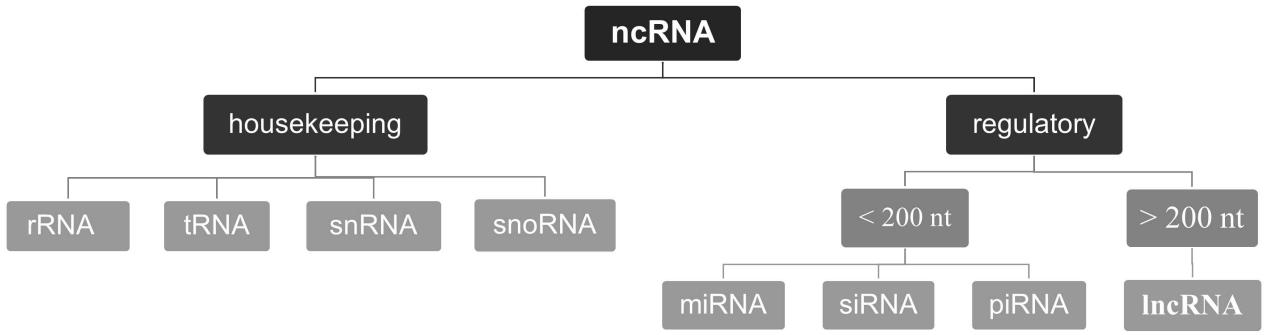


Figure 1.: Classification of non-coding RNA (ncRNA), based on function and length.
The group of housekeeping ncRNAs shown on the left side consists of ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA). The regulatory ncRNAs depicted on the right are further divided into RNAs shorter or longer than 200 nt. The group of short RNAs comprises e.g. microRNAs (miRNAs), small interfering RNAs (siRNAs) and piwi-interacting RNAs (piRNAs) while long non-coding RNAs (lncRNAs) are the largest and most diverse group, unified only by their size.

mRNA sequences. Thus, like miRNAs, they are involved in the regulation of gene expression and in silencing specific genes. This process is also vital for the defense against viral infections, where siRNAs bind to and trigger the destruction of viral genes [3], [10].
Long non-coding RNAs (lncRNAs), transcripts which are more 200 nucleotides in length, are involved in diverse regulatory functions. This includes guiding chromatin-modifying enzymes to specific genomic locations to facilitate chromatin remodeling, as well as transcriptional control or serving as a scaffold to assemble protein complexes. More recently, small open reading frames within lncRNAs being translated to micropeptides have been discovered, although their function remains to be shown [12]. In general, the role of lncRNAs is often context-dependent and they have diverse interaction partners like DNA, other RNA molecules or proteins [13]. More detailed information on lncRNAs can be found in subsection 8.1.1.
These various RNA classes, with many more interesting species like circular RNA (circRNA), highlight the diversity and versatility of RNA for both information transfer and regulatory functions within the cell. Each class contributes to complex cellular processes, reflecting RNA's essential role in life.

2.3. RNA life cycle

Although different in length, expression pattern and function, the life cycle of all RNA molecules begins with transcription inside the nucleus. During this procedure, an RNA polymerase synthesizes an RNA strand complementary to the DNA template. Depending on the type of RNA, in eukaryotes, one out of three DNA-dependent RNA polymerases binds to a promoter after the DNA double helix is unwound at the gene of interest by a helicase. Then, the four different RNA nucleotides are successively added complementary to the DNA template strand while the RNA polymerase moves downstream, resulting in an immature RNA molecule, which is usually further processed [14].
The best studied example is the RNA polymerase II enzyme, producing so-called pre-mRNA which undergoes several maturation steps co-transcriptionally. To the newly synthesized mRNA a 7-methylguanosine cap is added at the 5' end. Further, it has to undergo splicing, where introns are cut out, and exons are ligated together. At the end, the RNA is polyadenylated. This process adds a poly(A) tail to the 3' end in order to protect the RNA from degradation, thereby stabilizing the RNA molecule. The poly(A) tail also acts synergistically with the methylated cap of the mRNA to stimulate translation by forming a 'closed-loop' structure and recruiting translation initiation factors once exported from the nucleus [15]. After processing, most mature RNA molecules are transported to the cytoplasm. Already during maturation in the nucleus, the RNA is incorporated into complexes (so-called ribonucleoprotein complexes (RNPs)). The composition of these complexes varies with length, sequence, secondary structure, and of course, event of RNA processing and helps to determine the selection of the export pathway as well as the further transport and fate of the RNA.
In the cytoplasm, the RNP is immediately remodeled and export facilitating proteins are redirected to the nucleus. The RNA may now be translated into protein (in the case of mRNA) or fulfill its regulatory function (in the case of non-coding RNAs). The correct localization plays an important role for the RNA to be able to fulfill its correct function. There are three main mechanisms for RNA localization. The first, passive diffusion only works for limited distances. Secondly, protection from degradation at a specific locus in the cell is an elegant way to create an RNA gradient. Thirdly, the most intricate mechanism is directional transport via the help of transport proteins and often the cytoskeleton (for more detail, see subsection 10.1.1) [16], [17]. Finally, RNA molecules are subjected to decay, a process that controls their lifespan and ensures proper regulation of cellular RNA levels. There are several different RNA decay pathways, some playing a role already in quality control while some are targeted degradation pathways like those mediated by RNA-induced silencing complexes in the case of miRNA-targeted mRNAs. Usually, the poly(A) tail is trimmed by deadenylases, while other enzymes take over the decapping, before exonucleases come into action. The RNA transcript can be either degraded in a 5'–3' direction by Xrn1, an exoribonuclease or in a 3'–5' direction by the exosome, a multi-protein complex [3], [15]. Together, these processes maintain RNA function and turnover, supporting cellular homeostasis and gene regulation.

3 RNA-binding proteins and their RNA-binding domains

3.1. Classes of RNA-binding proteins

As already described in the section above from the perspective of an RNA, RNA molecules are in a tight interplay with RNA-binding proteins (RBPs) throughout their entire lifetime. The binding, complex formation, detachment and complex remodeling of RBPs influence RNA function and fate in several ways. As mentioned before, by attaching to specific sites on the RNA molecule, RBPs regulate RNA maturation, control nuclear export, direct localization within the cell, and affect the RNA half-life [18].
Approximately 5% of the total human proteome was identified as potential RNA-binding proteins, and more and more highly complex RNA–protein interaction networks are discovered [19]. RNA-binding proteins are mostly bi-functional, as they bind to both RNA and interact with proteins in a complex, enabling these networks. They can be divided into different classes based on their function, RNA targets and shared structures.
A selection of RNA-binding protein classes is shortly introduced below to demonstrate their versatility.
One of the most abundant RBPs in eukaryotic cells are heterogeneous nuclear ribonucleoproteins (hnRNPs), a selection of candidates is further introduced in subsection 10.1.3. They are involved in various stages of RNA processing, including splicing, transport, and stability and thus mainly bind to pre-mRNAs in the nucleus. For example, by binding to pre-mRNA and stabilizing their secondary structures, hnRNPs influence the splicing machinery's access to specific splice sites. This can either promote or inhibit the inclusion of certain exons. Thereby, hnRNPs influence alternative splicing and increase proteomic diversity [20], [21].
Another class of RBPs are Poly(A)-binding proteins (PABPs) that bind specifically to the polyadenylated tails of mRNA, and, as described above, protect them from degradation and enhance their translation. PABP1, a well-studied member of this family, binds the poly(A) tail

via multiple RNA recognition motif (RRM) domains (detailed below in section 3.2). PABPs can also form a complex with the 5' cap-binding proteins, leading to a circulation of the mRNA. This configuration facilitates the re-initiation of translation after the ribosome reached the poly(A) tail. Consequently, the translation efficiency is increased [15].
Yet another group of RNA-binding proteins which was already mentioned above, are RNA helicases. They use ATP to unwind secondary structures in RNA, facilitating access for other RBPs or enzymes. This unwinding activity is essential during splicing, ribosome biogenesis, and RNA decay, where RNA structures must be unfolded for efficient processing [3].

The selection of RBPs introduced above illustrate both, their broad functional diversity and regulatory complexity. RBP-RNA interactions are often dynamic and regulated by post-translational modifications, such as phosphorylation or methylation. This allows RBPs to quickly adapt to cellular changes and regulate gene expression by influencing RNA binding affinity as well as protein-protein interactions. Many RBPs share conserved structural motifs for RNA binding, which are characterized in the following section.

3.2. RNA-binding motifs

RBPs can be classified based on their function, as illustrated above, but also by their structural domains and RNA targets. The RNA-binding domain within the protein is the direct contact site between the protein and RNA molecule, determining binding specificity and serving as a basic classification feature.
The most common motif is the RNA Recognition Motif (RRM). It is characterized by a structure of 90 amino acids on average that form two conserved folding units (RNP1 and RNP2). This motif recognizes linear sequences of typically 2-8 nucleotides in length on its target RNA. Some RRM domains are rather sequence specific, while others have a broader binding spectrum but often multiple RRMs exist per protein [5]. Found in proteins like PTBP1 (introduced in section 10.1.3.3), the RRM is widely distributed.
The K homology (KH) domain is another frequently observed RNA-binding domain, first identified in the human heterogeneous nuclear ribonucleoprotein K (hnRNP K) (see 10.1.3.1 for detailed characterization). The KH domain is found in all domains of life, eukaryotes, bacteria and archaea. It is typically about 70 amino acids long and contains a unique β-α-α-β fold with a conserved "GXXG" RNA-binding motif located between the two α helices. This structure allows it to recognize and bind RNA (or sometimes DNA) sequences via hydrogen bonds, usually on a single-stranded region. Proteins with KH domains, such as FMR1 (linked to fragile X syndrome), recognize and interact with RNA by recognizing short sequence motifs of 4 nt, often not sequence specific. Many RBPs have multiple KH domains per protein [5].

In general, it can be distinguished between single-stranded RNA binding proteins and double-stranded RNA binding proteins, with the latter being the smaller group. The domains described above both bind to single-stranded RNA (ssRNA), while the double-stranded RNA-binding

domain (dsRBD) is specialized for binding double-stranded RNA (dsRNA). dsRNA is found in structured regions of mRNA, non-coding RNAs, and in the RNA interference pathways and proteins containing a dsRBD are found in all domains of life, including viruses. The dsRBD consists of about 65 to 70 amino acids arranged in a conserved α-β-β-β-α fold. Proteins involved in the synthesis of microRNAs, such as Drosha and Dicer, depend on this motif to detect and process structured RNAs with great specificity[5].

Besides the described structural domains, there are many more, like the large family of Zinc Finger (ZnF) domains or intrinsically disordered regions (IDRs). Furthermore, a huge diversity of viral RBPs exists, with unique protein structures evolved for binding highly structured viral RNA sequences. The variety of binding domains reflects the diversity of RNA sequences and 3D-structures. Often, one RNA-binding domain alone does not give enough specificity or affinity and therefore, multiple RNA-binding domains are combined in one RBP [5].

Understanding which RBPs interact with a given RNA is crucial for deciphering its fate. While RNA-binding motifs and protein classes are well characterized, they do not usually allow for precise prediction of RBP-RNA interactions. Given the significant influence of RBPs on RNA stability, localization, and translation, identifying these interactions is essential for uncovering the functional roles of specific RNAs.

4 Proximity labeling - finding the interactome

Proximity labeling (PL) describes a toolbox of methods that, in recent years, became more and more popular due to key advances of it's tools. PL is applied to answer a broad range of questions and is used to investigate otherwise hard to explore areas of the cell, for example organelle contact sites between the ER and Golgi [22] or virus host interactions [23]. In contrast to traditional co-purification strategies, proximity labeling approaches can identify even weak and/or transient interactions without cross-linking in order to stabilize these interactions. In a classical affinity pull-down, either antibodies to enrich the protein of interest or tagged probes with an affinity handle to pull-down the RNA of interest are used. The whole purification process has to be very gentle in order not to lose the direct interactors. Conversely, PL does not rely on preserving the (RNA-)protein complex, with the risk of identifying only tight and direct binding interactors, nor on bait-specific antibodies or probes. Furthermore, PL can be used to elucidate interactions in (subcellular) compartments that defy standard biochemical isolation methods, e.g. fractionation by centrifugation like stress granules [24] or lipid droplets [25].
The principle behind the PL assay can be seen in Figure 2 and is described in the following. The basis of the assay is the fusion or attachment of a labeling enzyme to a molecule of interest. This molecule can be either an RNA or, more commonly, any protein of interest. Although there are some protocols employing enzymes that use a peptide as label [26], most PL techniques are based on biotin.
Biotin (also known as vitamin H or B7) is an essential coenzyme for humans that is mainly synthesized in plants, bacteria and some fungi. It is used as a co-factor in enzymes such as biotin carboxylases and decarboxylases, which play a role in cellular processes like amino acid metabolism, lipogenesis, or gluconeogenesis. The labeling enzyme used in PL assays can be either a biotin ligase or peroxidase, in both cases mutated and engineered in a way that the enzyme promiscuously biotinylates its near surrounding. The reactive biotin species have been shown to be membrane-impermeant [27] and experimentally determined labeling radii for proximity-labeling enzymes are estimated to be $\sim$10 nm in living cells [28], [29].

It all started in 2012, with the *E. coli* carboxylase BirA being the first tool in the box. A mutant version of BirA, with a substitution of one amino acid (R118G), became known as BirA*. It prematurely releases reactive biotinyl-AMP, which then reacts with adjacent primary amines (mostly lysines) and forms a covalent bond, releasing AMP. This whole process is known as biotinylation [30], [31].
Later, peroxidases were also used as labeling enzyme. These enzymes are capable of converting biotin-conjugated compounds, like biotin-phenol with the help of H_2O_2 into short-lived radical species (e.g. phenoxyl-biotin) which react with electron-rich amino acids like tyrosine on neighboring proteins [28].
Biotin has an exceptionally high affinity for streptavidin, avidin, and neutravidin, forming one of the strongest non-covalent interactions known in nature (Kd = 10^{-15} M). Consequently, biotinylated proteins are efficiently isolated by using streptavidin or avidin beads. A key advantage of the PL method is that the interaction partners are covalently tagged with biotin while still in their native complex. As a result, the labeled proteins can be isolated under harsh extraction and washing conditions without requiring the preservation of the original complex [32]. After the enrichment, the protein samples are prepared for the analysis by mass spectrometry, usually by proteolysis [33] (see Figure 2).

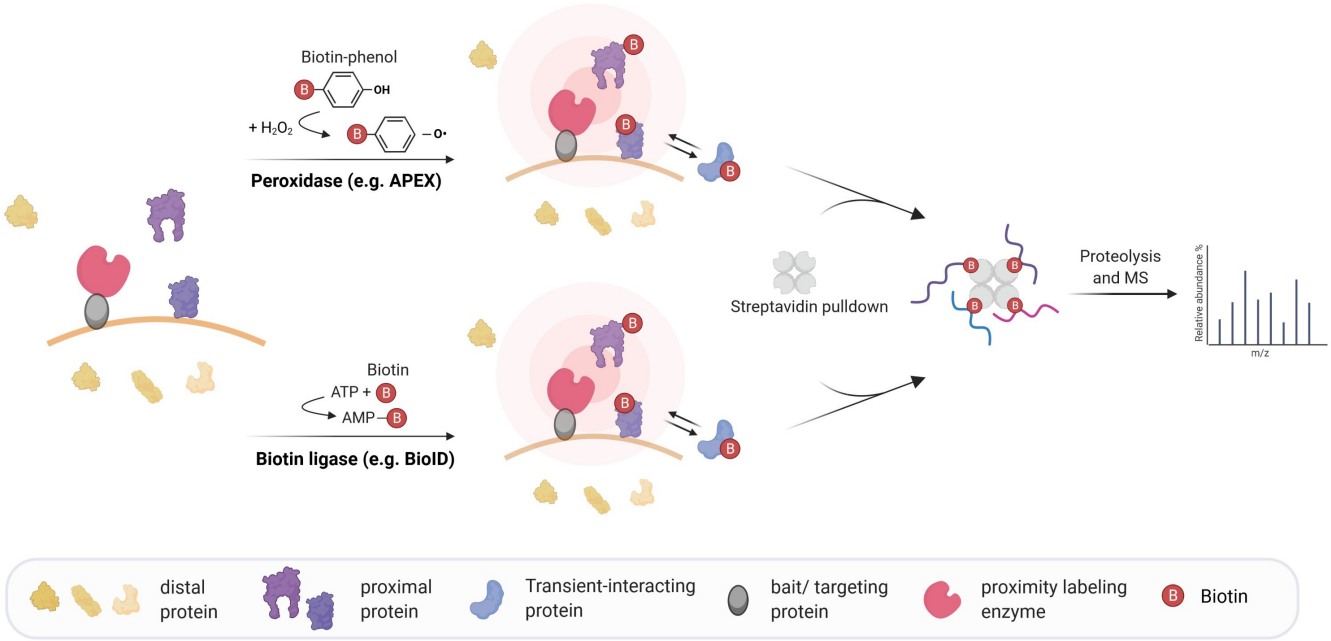


Figure 2.: Simplified workflow of the proximity labeling assay to identify interactors. After the labeling enzyme (a biotin ligase or peroxidase) is fused to a protein of interest and expressed, biotin or a biotin conjugated compound is added. During the biotinylation process, peroxidases oxidize biotin–phenol to reactive phenoxyl radicals using hydrogen peroxide while biotin ligases utilize ATP and biotin to catalyze the formation of reactive biotin-5'-AMP. Next, the activated biotin is released from the enzyme and covalently labels proteins in the proximity (depicted in blue/ purple), but not distal proteins (shown in yellow). After lysis, the biotinylated proteins are enriched by a streptavidin pull-down followed by proteolysis and analysis by mass spectrometry. Figure from [6].

4.1. (Directed) evolution of labeling enzymes

The first described promiscuous biotin ligase, BirA*, is a monomeric protein with 35.3 kDa and a mutated version of the *E. coli* biotin protein ligase BirA. BirA is a DNA-binding protein that functions as a transcriptional repressor for the biotin biosynthetic operon. The enzyme uses bicarbonate as a source to create CO_2, and then catalyzes the transfer of a carboxyl group with the help of biotin as carboxyl carrier [34], [35]. The catalysis is a two-step process. In the first step of biotin ligation, BirA generates biotinoyl-5'-AMP ('activated biotin') from biotin and ATP. The reaction intermediate is retained in the active site of the enzyme until, in a second step, it reacts with a nearby primary amine of its protein substrate [32], [36]. The Burke lab [35] took advantage of the reduced affinity of the mutated BirA* (BirA R118G) to biotinyl-5'-AMP. The enzyme prematurely releases the reactive biotinyl-AMP and, in a next step, it can react with nearby primary amines. This process adds a covalent tag to the proximal proteins and is called biotinylation [30], [31].
The originally used labeling enzyme BirA* has one major disadvantage: its activity is very low, so the incubation of samples with biotin ranges from 8 to 24 hours in order to get enough biotinylated material for analysis. Improving the slow labeling kinetic and thereby reducing the labeling time would enable the capture of snapshots of interactions or complexes under certain conditions triggered by small molecules or drugs. So from 2012 until now, several improved enzyme versions have been published. The original BirA* enzyme was either optimized or other enzymes were developed like BioID2 [37], AirID [38], BASU [39], microID and ultraID [40] or TurboID and miniTurboID [41].
TurboID was developed by Branon et al. through yeast-display-based directed evolution from *E. coli* BirA. The final variant contains 15 mutations relative to wild type BirA. A smaller 28 kDa version, miniTurbo, was also engineered by the authors. TurboID exhibits significantly enhanced activity, producing similar biotinylated material in 10 minutes as BioID, BioID2, or BASU generate in 18 hours, while miniTurbo was a bit less active. TurboID turned out to function superior across various cellular compartments, including the nucleus, mitochondrial matrix, endoplasmic reticulum lumen, and membrane. Furthermore, proteomic datasets generated from 10-minute TurboID (or miniTurbo) labeling are comparable in specificity and coverage to those obtained using BioID with an 18-hour labeling period. However, miniTurbo turned out to suffer from stability problems [41]. The high activity of TurboID enabled the successful extension of biotin-based PL assays to model organisms such as yeast [42], worms [41], flies [41], and plants [43], [44]. However, while they could not see decreased survival in *C. elegans* expressing the ligase, some developmental delay was observed [41]. The main drawback of TurboID is the strong degree of background labeling even before exogenous biotin is added or fed, likely due to its high affinity to biotin and the resulting usage of even low levels of biotin present in standard media. This necessitates caution when applying TurboID *in vivo*, as continuous biotinylation in the absence of supplemented biotin could lead to biotin depletion and potential cellular stress [41], [45]. Additionally, TurboID, like BirA, is relatively large (35 kDa), which may interfere with the localization, function, and interactions of fusion proteins.
Kubitz et al. presented an alternative approach for improving PL enzyme performance. Unlike TurboID, which was optimized for high activity, microID was engineered to minimize enzyme

size while maintaining robust biotinylation activity. It is derived from a truncation of BioID2, and has a molecular weight of 19.7 kDa, making it the smallest known PL enzyme. This compact size reduces the likelihood of interfering with the function or trafficking of fusion proteins, making microID a highly versatile tool for PL experiments. In addition to its small size, microID also exhibits enhanced activity compared to BioID. In a 10-minute labeling experiment, the clearest biotinylation signal was observed for BASU, TurboID, and microID. Like TurboID, microID functions effectively in a wide range of cellular contexts and has been successfully applied in bacteria, yeast, and mammalian cells [40].

Peroxidases provide an alternative to biotin ligases in the proximity labeling assay. Typical peroxidases can oxidize biotin–phenol in a hydrogen peroxide dependent reaction to reactive phenoxyl radicals ($t_{1/2} < 1$ ms) that biotinylate proximal proteins at electron-rich amino acid side chains [27], [46].
One peroxidase which was first utilized as tracer in electron microscopy before it was used for early proximity labeling assays is the heme enzyme horseradish peroxidase (HRP) [47], [48]. The major downside of HRP is its exclusive extracellular application, as the enzyme has disulfide bonds and calcium-binding sites and hence cannot be expressed in the reducing cytosolic environment of cells without loss of activity. In 2013, the Ting lab showed that an engineered ascorbate peroxidase called APEX, originally also used as a tag for electron microscopy [28], can be applied *in cellulo* for proximity labeling of the mitochondrial matrix proteome [27]. However, the main limitation of APEX is its low sensitivity when expressed at low levels. This has been overcome by generating APEX2 with the help of yeast-display evolution. APEX2 shows improved catalytic efficiency and higher activity [49]. APEX2-based proximity labeling has been successfully applied to study dynamic biological processes in protein signaling networks, such as ligand-stimulated remodeling of G-protein-coupled receptors [50]. (For more details see subsection 9.1.2.)
Additionally, APEX2 was shown to directly biotinylate RNA, by forming a covalent adduct between the RNA base guanosine and biotin phenol. This technique enables to capture proximal RNAs and proteins at the same time, providing a powerful tool to analyze the dynamics of RNAs and proteins in macromolecular complexes [51], [52].
Although all peroxidase-based proximity approaches allow biotinylating proteins or RNA within a short time ($\sim$ 1 min), a major drawback of these enzymes is the potential oxidative stress and toxicity from H_2O_2, which makes them challenging to apply *in vivo*.

4.2. Protein-centric proximity labeling

In the beginning, the proximity labeling (PL) assay was designed to identify protein-protein interactions by covalently adding a biotin label to proteins in proximity to the POI. This was mostly done in cultured cells, allowing the detection of native interactions and complexes. The biotin ligase in charge of the labeling was fused to the POI either C- or N-terminally, which is why this approach is called protein-centric proximity labeling. Many interesting discoveries

have been made in more than a decade of protein-centric proximity labeling. This includes the characterization of insoluble protein complexes (for example, the nuclear envelope [32]) or capturing short-lived interactions like enzyme–substrate reactions [53].

Over time, several modifications of the basic assay have been developed. Nowadays, the proximity labeling approaches can be broadly classified into either protein-centric or RNA-centric. While the first focuses on a specific protein of interest (POI) and seeks to find the interacting proteins and RNA molecules, identified by MS or RNA sequencing, respectively, the second starts with an RNA of interest and aims to identify the interacting proteins by mass spectrometry (MS).
Additionally, DNA-centric proximity labeling has be mentioned here. This assay is based on the DNA binding specificity of the Cas9 enzyme, but takes advantage of a catalytically inactive version (dCas9) fused to BirA* and GFP. Cells expressing the fusion protein of BirA*-dCas9-eGFP together with a sgRNA targeting the whole construct to the sequence of the desired genome locus, can biotinylate proteins in the direct vicinity of the specific DNA locus [54]. The same idea also proved to work later using an APEX2-dCas9 fusion protein [55], [56].

4.3. RNA-centric proximity labeling

RNA-centric proximity labeling was developed to characterize the proteins associated with a specific RNA. There have been several assays developed to serve this purpose. In RaPID [39], a short RNA of interest is flanked by two BoxB aptamers. Its co-expression with a fusion of the BoxB recognizing λN-peptide and BirA* or BASU allows the biotinylation of proteins bound to or associated with the flanked RNA sequences. This study identified proteins that bind to well-characterized RNA motifs (e.g. the IRE, TNF-CDE or PUF RNA motifs) and analyzed how mutations of these motifs affect protein binding [39], [57]. In contrast, RNA-BioID [1] characterizes the proteome of an endogenous RNA. Here, BirA* fused to the MS2 coat protein (MCP) was stably expressed in fibroblasts from a transgenic knock-in mouse line [58] where MS2 aptamers had been inserted into the β-actin gene locus. (More details on RNA tagging can be found in subsection 8.1.2.)
In the RNA-BioID assay, the modified β-actin RNA, expressed at endogenous level, contains 24 MS2 aptamers in its 3'UTR. The MCP is fused to a fluorescent protein and the BirA* enzyme including a nuclear localization signal (NLS), and is co-expressed with the tagged RNA. The NLS is included so the MCP-fusion protein can already bind to the MS2 stem loops of the RNA co-transcriptionally in the nucleus and is then co-exported to the cytoplasm together with its cognate RNA. A second reason to include the NLS is that unbound MCP-fusion proteins are restricted to the nucleus, resulting in a nuclear background fluorescence and biotinylation, but no cytoplasmic background.
RNA-BioID not only identified all the RBPs previously reported to bind to β-actin mRNA but also novel functional interactors, including FUBP3/MARTA2, were found [1]. However, the major technical hurdle of RNA-BioID is the need to genomically insert the MS2 aptamer array.

Due to genetic modification of the mRNA, it can lead to alterations in the expression and/ or stability of the mRNA itself [59]. The development of CRISPR-based knock-in tools might offer a more user-friendly way to overcome this [60].
It is critical to note that in both RNA-centric proximity labeling presented here, the tested RNAs are highly expressed, either by over-expression from a strong heterologous promoter (RaPID) or due to their high endogenous level of expression (RNA-BioID). This facilitates the efficient biotinylation of associated proteins and makes it easier to obtain enough material for subsequent analysis. Thus, it will be interesting to see if these methods can be successfully applied to low abundant RNAs, especially considering the problem of the observed labeling of non-RNP proteins. In principle, it should be possible to increase the labeling of RNA-associated proteins by increasing the number of aptamers, resulting in the recruitment of more labeling enzymes to the RNA and thus an increase in the signal-to-noise ratio. However, as this might affect the function or the stability of the RNA [61], [62], a careful examination of the impact of large aptamer arrays on the RNA will be required (discussed in more detail in section 8.1.2). Another risk in applying these assays is the fact that the fusion protein, including the biotin ligase, is not stably bound to the RNA although the affinity of MCP was increased and the dissociation kinetics decreased. This could potentially lead to an increase in background biotinylation by the free proximity labeling enzyme after the dissociation from the RNA.

5 Motivation and aim of this work

The main objective of this study was to identify RNA-binding proteins that interact with lncRNAs, as this class of RNA remains relatively understudied. The localization and function of an RNA molecule in the cell is determined by its sequence and secondary structure. However, this information alone is insufficient to predict the fate of an RNA transcript [63]. Modern, computer-based tools such as protein structure prediction help to understand protein shape and support functional studies, but RNA structure is still not easy to predict, not even speaking of drawing conclusions about the function of an RNA based on its sequence. Identifying their binding partners provides valuable details to better understand their function. As RBPs act in concert to achieve their regulatory role, characterization of the whole set of proteins that stably or transiently interact with RNAs should be taken into consideration for the elucidation of their molecular function. Although there is substantial knowledge about how RBPs bind RNA, much less is known about the reverse. Additional methods to study RNA molecules in complex with RBPs are therefore needed [5].

Hence, the main goal of this thesis is to uncover the interactomes of the lncRNAs candidates OIP5-AS1 and NORAD. Thereby, an insight into their molecular functions, based on their interaction partners should be gained. While key interactors of NORAD (PUM1 and PUM2) have been identified and can serve as positive controls, OIP5-AS1 remains poorly understood, and this work aimed to help elucidate its cellular function.

To achieve this, the RNA-BioID assay, originally introduced by Mukherjee et al. (2019) and applied to β-actin mRNA [1], should be adapted and advanced. The experimental plan involved stably integrating a biotin ligase into cells, followed by CRISPR/Cas9-mediated tagging of the endogenous OIP5-AS1 and NORAD loci with 24xMS2 loops. After optimization of biotin labeling and streptavidin-based pull-down, enriched proteins are to be identified by mass spectrometry.

As an alternative or complementary strategy, the HyPro assay introduced by Yap et al. (2021) was pursued [64]. While the original method was only applied to nuclear ncRNAs, this project aimed to expand its use to cytoplasmic lncRNAs, such as NORAD and OIP5-AS1. As an

initial step, NEAT1, which was also used in the original study, should serve as a benchmark to establish the method. Therefore, the HyPro enzyme is first expressed, purified and functionally validated *in vitro*. This is followed by verification of probe specificity, enzyme localization and biotinylation via immunofluorescence. Once optimized, streptavidin pull-down and mass spectrometry were planned for interactome profiling.
Together, these two approaches aimed to establish strategies that enable the identification of protein interactors for the selected cytoplasmic lncRNAs.

The goal of the last part of this study was to prove and further analyze the interactions between β-actin mRNA and four (potential) RNA-binding candidates previously identified by Mukherjee et al. (2019) in the RNA-BioID screen [1]. The chosen candidates are hnRNP K, TDP-43, PTBP1 and hnRNP R.
To achieve this goal, a protocol for RNA co-immunoprecipitation (Co-IP) with MEF cell lysate is established for each of the four candidates. For RNA interaction analysis, after successful Co-IP, the protocol is performed again under RNase-free conditions to enrich for RNA-protein interactions. Pulled-down RNAs are then isolated and converted to cDNA via reverse transcription. By performing RT-qPCR, the association of β-actin mRNA with the target protein can be confirmed.
This approach aims to validate β-actin mRNA-protein interactions in MEF cells and provide insights into their biological relevance.

Part II.

Material and Methods

6 Material

6.1. Buffers and solutions

Manually prepared buffers and their composition in alphabetic order are listed. For all buffers deionized water was used.

Buffer name	**Composition**
AIR buffer	150 mM NaCl
	20 mm HEPES, pH 7.4
	15 mM glucose
	15 µg/ml BSA
	20 mM Trehalose
	5.4 mM KCl
	0.85 mM $MgSO_4$
	0.7 mM $CaCl_2$x2 H_2O
Blocking buffer	5 g milk powder
	in 50 ml TBS-T buffer
Coomassie destaining solution	40% (v/v) ethanol
	7% (v/v) acetic acid
Coomassie staining solution	0.25% (w/v) Serva Blue R
	40% (v/v) ethanol

	7% (v/v) acetic acid
FACS buffer	1xPBS
	0.5 µM EDTA
	2% Fetal bovine serum
5x Laemmli sample buffer	250 mM Tris-HCl, pH 6.8
	10% (w/v) SDS
	40% (w/v) glycerol
	200 mM β-mercaptoethanol
	0.005% (w/v) BPB Bromophenol blue
Lucria Broth (LB) medium	10 $\frac{g}{L}$ NaCl
	5 $\frac{g}{L}$ yeast extract
	10 $\frac{g}{L}$ Bacto Tryptone
	(1.5% (w/v) agar)
Lysis buffer	50 mM Tris, pH 7.5
	150 mM NaCl
	2.5 mM $MgCl_2$
	1 mM DTT
	1% Triton X-100
	1x proteinase inhibitor
Lysis buffer for gDNA isolation	100 mM NaCl
	25 mM EDTA pH 8
	10 mM Tris-HCl
	0.5% SDS
	500 µg/mL Proteinase K
NT2 buffer (RNase-free)	50 mM Tris, pH 7.4
	150 mM NaCl
	1 mM $MgCl_2$
	0.01% NP-40
1x PBS	2 mM KH_2PO_4
	10 mM Na_2HPO_4
	137 mM NaCl

	2.7 mM KCl
Ponceau S solution	5% (v/v) acetic acid
	0.1% (w/v) Ponceau S
Resolving gel buffer	1.5 M Tris-HCl, pH 8.8
	0.4% SDS
SDS-PAGE running buffer	250 mM Tris-HCl, pH 8.3
	192 mM glycine
	1% (w/v) SDS
Stripping buffer	0.1 M glycine-HCl, pH 2.2
	0.1% (w/v) SDS
TAE buffer	40 mM Tris-HCl
	20 mM glacial acetic acid
	1 mM EDTA, pH 8
TBS buffer	50 mM Tris-HCl, pH 7.5
	150 mM NaCl
	1 mM $MgCl_2$
TBS-T buffer	50 mM Tris-HCl, pH 7.5
	150 mM NaCl
	1 mM $MgCl_2$
	0.01% (v/v) Tween-20
Towbin buffer	250 mM Tris-HCl, pH 8.6
	192 mM Glycine
20x SSC buffer, pH 7.0	3 M NaCl
	0.3 M Sodium Citrate

Manually prepared buffers for Southern blot and their composition in alphabetic order. For all buffers deionized water was used.

Buffer name	**Composition**
Denaturation solution	0.5 M NaOH
	1.5 M NaCl
10xDIG1 buffer, pH 7.5	1 M maleic acid
	1.5 M NaCl
DIG1 wash buffer	1x DIG1 buffer
	0.3% Tween-20
DIG1 blocking solution	1x DIG1 buffer
	1x DIG blocking solution
DIG3 wash buffer, pH 9.5	0.1 M NaCl
	0.1 M Tris
High-stringency wash buffer	0.1x SSC
	0.1% (w/v) SDS
Neutralization solution	0.5 M Tris, pH 7
	1.5 M NaCl
Southern blot wash buffer	2x SSC
	0.1% (w/v) SDS
Stripping buffer	0.2 M NaOH, pH 2.2
	0.1% (w/v) SDS

Solutions and reagents for sm(i)FISH were purchased RNAse-free to avoid RNA degradation. Powders were dissolved in HPLC grade water.

Name of the reagent	**Manufacturer**
Bovine serum albumin	New England Biolabs GmbH, (GE)
Dextran sulfate	Carl Roth GmbH + Co. KG, (GE)
Formamide	Honeywell International Inc., (US)
NEB buffer r3.1	New England Biolabs GmbH, (GE)
10x PBS	Thermo Fisher Scientific GmbH, (GE)

Ribonucleoside-vanadyl complexes (VRC)	New England Biolabs GmbH, (GE)
20x SSC	Thermo Fisher Scientific GmbH, (GE)
tRNA (from E. coli MRE 600)	Roche Deutschland Holding GmbH, (GE)

6.2. Laboratory instruments

Used laboratory equipment, their name and manufacturer in alphabetic order.

Instrument	**Name, Manufacturer**
Cell counter	CellDrop™, DeNovix Inc. (US)
Centrifuges	5415 R, Eppendorf SE (GE)
	5424, Eppendorf SE (GE)
	5810 R, Eppendorf SE (GE)
	5702 R, Eppendorf SE (GE)
Dry bath	heat block, Thermo Fisher Scientific GmbH (GE)
Electrophoresis system	Mini-PROTEAN®Tetra Cell,
	Bio-Rad Laboratories GmbH (GE)
Electroporator	4D-Nucleofector™Core adn X Unit, Lonza AG (GE)
Freezing container	Mr. Frosty™, Thermo Fisher Scientific GmbH (GE)
Gel documentation	ChemiDoc™MP Imaging System,
	Bio-Rad Laboratories GmbH (GE)
Hybridization oven	Cytiva Germany GmbH (GE)
Incubator	Safe 2020, Thermo Fisher Scientific GmbH (GE)
Microscope	CKX53, Olympus Europa SE & Co. KG (GE)
	Axio Observer Z1, Carl Zeiss AG (GE)
PCR thermocycler	T100™Thermal Cycler, Bio-Rad Laboratories GmbH (GE)
	GE9612T-S™, Bio-Gener Technology Co. Ltd (CH)
pH-meter	FiveEasy, Mettler-Toledo GmbH (GE)
Pipettes	Eppendorf AG (GE)
Power Supply	PowerPac™Basic, Bio-Rad Laboratories GmbH (GE)

Rotating Mixer	Benchmark Scientific Inc. (US)
Shaker	DRS-12, neoLab Migge GmbH (GE)
Sterile bench	Safe 2020, Thermo Fisher Scientific GmbH (GE)
Thermo shaker	Mixing Block MB-102, Bioer Technology GmbH (GE)
UV/Vis Spectrometer	GENESYS 10 Bio, Thermo Fisher Scientific GmbH (GE)
	NanoDrop One, Thermo Fisher Scientific GmbH (GE)
Vortex mixer	VORTEX-GENIE®2, Scientific Industries, USA
Water bath	Isotemp 210, Thermo Fisher Scientific GmbH (GE)

6.3. Bacterial strains

Table 5.: Bacterial strains whit their genotype and source in alphabetical order.

E.coli strain	**Genotype**	**Source**
BL21 (DE3)	F$^-$ *omp*T *gal dcm lon hsd*S$_B$(r$_B{}^-$,m$_B{}^-$) λ(DE3)	Thermo Fisher Scientific
DH5α	F$^-$ *end*A1 *glnV*44 *thi*-1 *rec*A1 *rel*A1 *phoA* *gyr*A96 *deoR* *nupG* *purB*20 Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169, *hsd*R17(r$_K{}^-$,m$_K{}^+$), λ^-	Thermo Fisher Scientific
TOP10	F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*Z ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G λ-	Thermo Fisher Scientific

6.4. Enzymes

Restriction enzymes used for cloning or southern blot were ordered from New England BioLabs or Thermo Fisher Scientific GmbH and used with the buffer recommended by the manufacturer.

Table 7.: Enzymes used for PCR, cloning, RNA or protein digestion or CRISPR with their concentration and manufacturer in alphabetical order.

Enzyme	Concentration	Manufacturer
Alt-R™S.p. Cas9 Nuclease V3	10mg/mL	Integrated DNA Technologies, Inc.
Herculase II	2500 U/mL	Agilent Technologies Deutschland GmbH
Proteinase K	20 mg/mL	Thermo Fisher Scientific GmbH
20x reverse transcriptase	-	Thermo Fisher Scientific GmbH
RNase A	-	Thermo Fisher Scientific GmbH
RQ1 DNase	1 U/µL	Thermo Fisher Scientific GmbH
T4 DNA Ligase	400000 U/mL	Thermo Fisher Scientific GmbH

6.5. Plasmids

Table 8.: Plasmids used in this work. ID = database entry number.

ID	Plasmid name	Vector	Resistance
p1953	NLS-2xMCP-eGFP-BirA	phageUBC	Amp
p1999	Impa1 2bp mCherry	pcDNA3.1(+)	Amp
p2039	NLS-2xMCP-eGFP-BirA	pcDNA3.1(+)	Amp
p2042	PGK-FLP-obpA	-	Amp
p2047	GST-GFP-Igf2bp1	pGEX	Amp
p2082	mCherry-Fubp3	pcDNA3.1(+)	Amp
p2122	Flag-C-BirA-FRB_Myc-N-BirA-FKBP	pSF3	Amp
p2123	Flag-C-BirA-FRB_FKBP-Myc-N-BirA	pSF3	Amp
p2124	Flag-FRB-C-BirA_Myc-N-BirA-FKBP	pSF3	Amp
p2125	Flag-FRB-C-BirA_FKBP-Myc-N-BirA	pSF3	Amp

p2142	Fubp3	pETM-14	Kan
p2231	split-microID	pSF3	Amp
p2246	V5-TurboID-NES	pcDNA3	Amp
p2250	NLS-2xMCP-eGFP-microID	pcDNA3.1(+)	Amp
p2268	actb HDR construct	puc57	Amp
p2275	CMV-Cre Recombinase-eGFP		Amp
p2281	sgRNA-BbsI_sgRNA-donor plasmid	-	Amp
p2283	Fluc-eGFP	-	Kan
p2285	CMV sTurboID(C)-HA-FRB-ERM	pLX208	Amp
p2286	CMV OMM-FKBP-V5-sTurboID(N)	pLX304	Amp
p2309	hnRNP R-HA	pcDNA3	Amp
p2314	LHA beta actin	pUC57	Amp
p2321	arc-HDR repair	puc57	
p2327	HyPro expssion plasmid	pML433	Kan
p2339	OIP5-AS1 LHA-MCS-RHA	pUC (Genewiz)	Amp

Table 9.: Generated plasmids in this work. The plasmid maps can be found in the supporting information. ID = database entry number.

ID	Plasmid name	Vector	Resistance
p2263	2xMCP-eGFP-TurboID	pcDNA3	Amp
p2264	2xMCP-eGFP-TurboID	phageUBC	Amp
p2279	2xMCP-eGFP-microID	pSF3	Amp
p2280	2xMCP-eGFP-TurboID	pSF3	Amp
p2313	donor-CMV-puro-loxP-24xMS2 V5-loxP	-	Amp
p2315	hnRNP R-HA-eGFP	-	Kan
p2324	Amp-NORAD	pUC57	Amp
p2325	Amp-OIP5-AS1	pUC57	Amp
p2340	OIP5-AS1 HDR-repair-construct_with HAs	pUC	Amp
p2341	OIP5-AS1 HDR repair template_with HAs+loxP sites	pUC	Amp
p2362	FRB-TurboID(C)-HA_FKBP-V5-TurboID(N)	pSF3	Amp
p2363	FRB-TurboID(C)-HA_FKBP-V5-TurboID(N) long	pSF3	Amp

p2391	FRB-TurboID(C)-HA_FKBP-V5-TurboID(N)_ TetOn-T2A-Puro	pSF3	Amp
p2393	2xMCP-TurboID(C)-HA_FKBP-V5-TurboID(N)_TetOn-T2A-Puro	pSF3	Amp
p2394	2xMCP-TurboID(C)-HA_ZBP1-V5-TurboID(N)_TetOn-T2A-Puro	pSF3	Amp
p2395	2xMCP-TurboID(C)-HA_Fubp3-V5-TurboID(N)_TetOn-T2A-Puro	pSF3	Amp
p2404	gRNA1a in p2281	2281	Amp
p2405	gRNA1b in p2281	2281	Amp
p2406	gRNA1c in p2281	2281	Amp
p2407	gRNA2a in p2281	2281	Amp
p2408	gRNA2b in p2281	2281	Amp
p2409	gRNA Ia in p2281	2281	Amp
p2410	gRNA Ib in p2281	2281	Amp
p2411	gRNA II in p2281	2281	Amp
p2412	gRNA IIIa in p2281	2281	Amp
p2413	gRNA IIIb in p2281	2281	Amp
p2414	gRNA IVa in p2281	2281	Amp
p2415	gRNA IVb in p2281	2281	Amp

6.6. Oligonucleotides

All oligonucleotides were ordered from Merck KGaA, (GE).

Table 10.: Primers used in this work for cloning, genotyping PCR or sequencing. The name of each Primer is followed by it's orientation, with F indicating a forward and R indication a reverse direction. The asterisk (*) indicates nucleotides that are joined by phosphorothioate internucleotide linkage. ID = database entry number.

ID	Name	Sequence (5' → 3')
3529	BGH 3-R	AGAAGGCACAGTCGAGG
3654	GAPDH_250-F	GTTCGACAGTCAGCCGCATCT
3655	GAPDH_250-R	TCTCAGCCTTGACGGTGCCA
4676	M13_pUC_rev_primer	AGCGGATAACAATTTCACACAGG

4927	N-term mCherry-R	GTCTGGGTGCCCTCGTAG
5147	Sequencing primer within EGFP	CTGAACTTGTGGCCGTTTA
5188	pSF3 sequencing antisense-R	GATATCCTGTCTTTAACAAATTGG
5623	In EGFP Seq primer-F	CACATGGTCCTGCTGGAGTTCGTG
5761	1-Gapdh_F_RT	AAGAGGGATGCTGCCCTTAC
5762	1-Gapdh_R_RT	ATCCGTTCACACCGACCTTC
5763	2-Gapdh_F_RT	GAGGGATGCTGCCCTTACC
5764	2-Gapdh_R_RT	AAATCCGTTCACACCGACCT
5765	3-Gapdh_F_RT	AGAGGGATGCTGCCCTTACC
5766	3-Gapdh_R_RT	CCAAATCCGTTCACACCGAC
5769	qRT-F-zip	CGGACTGTTACTGAGCTG
5783	1. qRT-β-actin-F	ACTGAGCTGCGTTTTACACC
5784	1. qRT-β-actin-R	GCCTTCACCGTTCCAGTTTTT
5803	psF3 cmv seq-R	GTGAAAGAAGCTTGTAGGC
5804	psF3 cmv seq-F	TGGTAAACTCGAGGGTAGG
5815	1. Amp_qRT-F	AGTGATAACACTGCGGCCAA
5816	1. Amp_qRT-R	TCCGGTTCCCAACGATCAAG
5855	bGHpolyA-seq-R	CACCTTCCAGGGTCAAG
6050	psF3-seq-RNA-F	CGTCTTTGGATTAGGCAG
6143	5. Fubp3-RT-F	CCCTCCTAACACCGACCCTA
6201	qRT-NORAD-R1	CGCTCTTGTAGACCTGTGGT
6202	qRT-NORAD_R1	CGCTCTTGTAGACCTGTGGT
6203	qRT-NORAD-F2	CCTCTGCCCTGCAGATAACG
6204	qRT-NORAD-R2	ATTAGTCACCGAGGTGGGGT
6205	qRT-NORAD-F3	AGGCCATTAAGGCGTGGAC
6206	qRT-NORAD-R3	GACTAGGTAGGTCTCGCCCA
6284	1-FubpF1	CTTAGTCACGGGGATGGTGG
6202	qRT-NORAD3-gRNA-R	AAACTTCCGACATCGATCCACGAT
6395	RBS-Seq-F	TTTGTTTAACTTTAAGAAGGAGA
6376	puro-Seq-F	GCAACCTCCCCTTCTACGAGC
6416	Norad-assembly-F3	CTCGAGAGGCTGAGGCAGGAGAATCG
6498	mCherry-Mid-F	ACGGCGAGTTCATCTACAAGG
6593	NORAD-seq-F2	AAGCTGCTCTCAACTCCACC
6595	NORAD-seq-F4	AGGCTGTCTTCTGAAAAGGAC
6596	NORAD-seq-F5	GTGGCGCATGCCTGTAATCC

6756	18S qRT-F	GCTTAATTTGACTCAACACGGGA
6757	18S qRT-R	AGCTATCAATCTGTCAATCCTGTC
6802	Actb-Northern-Both	GGTCATCTTTTCACGGTTGG
6851	Seq-MS2-F	CCAAATGGGTTCAGCAC
6874	1. qRT-ACTb long-F	GTGGATCAGCAAGCAGGAGTA
6875	2. qRT-ACTb long-R	AAAACGCAGCTCAGTAACAGTC
6881	qRT-ACTb 5UTR-F	CACTGTCGAGTCGCGTCC
6882	qRT-ACTb 5UTR-R	TCATCCATGGCGAACTGGTG
6957	BamHI nested GSP	TCAGGATCCTTGACATCCGTAAAGACCT CTATGC
6976	TurboID ubc slic-F	GCATGGACGAGCTGTACAAGCTCGAGAA AGACAATACTGTGCCTCTGAAGC
6977	TurboID ubc slic-R	CCAGAGGTTGATTAGGATCTATCGATTCA CTTTTCGGCAGACCGCAGA
6984	px459 seq-F	TTTATGGCGAGGCGGCGG
7041	microID pcDNA3.1 R	AACGGGCCCTCTAGACTCGAGCGGCCGC TTACTTCTCCTTGAACTTCTTCAGG
7066	TurboID slic pcDNA3-F	CACTCTCGGCATGGACGAGCTGTACAAG CTCGAGAAAGACAATACTGTGCCTCTGA AGC
7067	TurboID slic pcDNA3-R	AACGGGCCCTCTAGACTCGAGCGGCCGC TTACTTTTCGGCAGACCGCAGA
7095	OIP5-AS1-F	TGCTCATCCATGTTAGCTGG
7096	OIP5-AS1-R	GTGAGCAAAGATTGTGCCAC
7103	OIP5-AS1 nested-F	ACCGGGATATGTGCAATAGA
7107	OIP5-AS1 nested-R	AGTGAAACTCTATCTCCCCC
7112	chicken b-actin Promoter Seq-F	GTTTCCTTTTATGGCGAGGC
7113	CMV enhancer Seq-R	CAAGTGGGCAGTTTACCGTA
7129	C-term Cas9-F	TGTACGAGACACGGATCGAC
7130	CMV enhancer-Seq-F	CCTACTTGGCAGTACATCTACG
7131	Cas9-Seq-R	CGTTGCTGAAGATCTCTTGC
7132	CMV_fwd_seq	GGCACCAAAATCAACGGGAC
7133	WPRE_rev_seq	CATAGCGTAAAAGGAGCAACA
7137	Gapdh_fwd_RT-qPCR	TGACCTCAACTACATGGTCTACA
7138	Gapdh_rev_RT-qPCR	CTTCCCATTCTCGGCCTTG

7157	Bsp119I2xMCPsTurboID(C+N)-F	TAAGCATTCGAAGCTAGCCGTTAAAATG GCTTCTAAC
7158	Bsp1407Istop2xMCPsTuID(C-R	TGCTTATGTACATTAGTAGATGCCGGAG TTTGCT
7159	BamHI2xMCPsTurboID(N-R	GTCTTAGGATCCGTAGATGCCGGAGTTT GCTG
7171	BamHI_2xMCP-R	GTCTTAGGATCCTCTAGAATCCGCGTAG ATGCC
7208	PCR_MS2-R	TTGTGTATTGTACCCCTTGATTG
7224	NheI-hnRNP R_mouse-F	CGCGCTAGCATGGCTAATCAGGTGAATGGT
7225	hnRNP R-BglII_mouse-R	GCGAGATCTTTTCCACTGTTGCCCATAAGT
7226	hnRNP R-5'UTR_mouse-F	CGCGCCATTTCTTGTCGTT
7227	hnRNP R 3'UTR_mouse-R	CTCTGCAAGCCCAATAATGTCC
7228	Tpm1-qPCR_mouse-F	GCTGAAGCTGACGTAGCTTC
7229	Tpm1-qPCR_mouse-R	CTCACTCTCATCTGCAGCCT
7230	Mink1-qPCR_Mouse-F	TGGTGATGGAGTTCTGCGGT
7231	Mink1-qPCR_mouse-R	CCTGAGAATCTCCCTGCAGATG
7236	MS2-V5 F	CACCCGACACTTCACAATCA
7237	MS2-V5 R	ATAGTTTGCGCTGGACGAAA
7238	MS2 R	TCCGTTTGTCGTGTTTTTGTGA
7258	M13 F	GTAAAACGACGGCCAGT
7259	AscI-MS2-V5-F	ACTGGCGCGCCGCCTGGAGCGTGAATC GATA
7260	SgrDI-MS2-R	GCTCGTCGACGGCAGCGTACAATTCCT GCAG
7261	NheI-hnRNP R-Kozak-F	GCGCTAGCACAGCCACCATGGCTAATCA GGTGAATGGT
7262	HA-BglII-R	CGCAGATCTAGCGTAATCTGGAACATCG TATGG
7263	7SK-ncRNA_mouse-F	AGGACGACCTTCCCCGAATA
7264	7SK-ncRNA_mouse-R	GCGCCTCATTTGGATGTGTC
7268	NORAD-PCR-R	ATCTTCCCAGAGGGTGGTG
7269	NORAD-PCR-F2	TGGTAGAGGTGGGAAAGGGA
7270	NORAD-PCR-R2	GAGGGGTGTTACTTAAGCCCTC
7277	MS2 V5-F	CTCCTACCTCAAACCTCTTCCC
7289	BSD F	CAAGCCTTTGTCTCAAGAAGAAT

7290	BSD-F2	TCGTCGCGATCGGAAATGAG
7291	BSD-R	TAGCCCTCCCACACATAACC
7292	MS2 V5-R	GCCCTCGTTTGACGTATATTGC
7293	arc HDR repair template-F	CCTACCCCTCATCTGTCTG
7294	arc-RHA HDR repair template-R	GAGCAGAATGAGGAAGCC
7303	NORAD-PCR-R4	CCCATCTCCATCAACCCAGAAG
7312	NORAD_SpeI-R	GACTAGTCTGCGACAGGCCGTTGTACAC
7313	NORAD_MfeI-F	GCAATTGAGGCTGTCTTCTGAAAAGGAC
7314	OIP5-AS1-15-MfeI-F	CGCAATTGGTGGTCTCATGTTTACTCTTC
7315	OIP5-AS1-16-AscI-R	CAGGCGCGCCGAGTCATGCATTAGAATG TTCG
7322	NORAD-DIG-probe-R	GCCTAGAACCCTTCCCTTTC
7323	NORAD-DIG-probe-F	GATAACATCAGCTCTAAGTGACAC
7324	Puro-seq-out-F	CCGGTGCCTGACAGATCTTAG
7354	MS2-V5 middle-F	CACTCGAGCATACATTGTGCC
7355	MS2-V5 up-F	CCTTCGCGCCATTCTTACAC
7356	MS2-V5+T7-Promoter-R	GAAATTAATACGACTCACTATAGGCTGTT GGTGCTCGGATGTG
7422	GAPDH-F	TTCGACAGTCAGCCGCATCTTCTT
7423	GAPDH-R	GCCCAATACGACCAAATCCGTTGA
7529	ClaI-loxP-F	CGATATAACTTCGTATAGCATACATTATA CGAAGTTATC
7530	XmaI-loxP-F	CCGGGATAACTTCGTATAGCATACATTAT ACGAAGTTATT
7531	loxP-XbaI-R	CTAGAATAACTTCGTATAATGTATGCTAT ACGAAGTTAT
7532	loxP-MfeI-R	AATTGATAACTTCGTATAATGTATGCTAT ACGAAGTTAT
7549	NotI-sTurbo(C)-R	TCAGTGGCGGCCGCGCCACACCGCTCC CACTA
7550	PacI-SbfI-sTurboID(C)-F	TCAGTCTTAATTAACCTGCAGGAACTGT CGGGATCCGCCAC
7551	MluI-sTurboID(N)-F	TCAGTCACGCGTAAACTGGAAGGGAGTGGC
7552	MluI-sTurboID(N)-F2	TGACTGACGCGTAGTGGCAGCGGAGGTAGT
7553	HpaI-sTurboID(N)-R	GACTCTGTTAACTGTCGACTTAACGCGCCAC
7691	SpeI-TetOn-Puro-R	GACTGACTAGTGGAAGGGAGACCAGATGAGC

7692	BsrGI-TetOn-Puro-F	ACTGTGTACATCTAAGTGGATCTGCGATCGC
7693	SpeI-TetOn-Puro-F	CAGTCACTAGTCTAAGTGGATCTGCGATCGC
7694	BsrGI-TetOn-Puro-R	ACTGTGTACAGGAAGGGAGACCAGATGAGC
7695	FseI-MCP-F	GTCAGGCCGGCCGCTAGCCGTTAAAATG GCTTC
7696	SbfI-MCP-new-R	GTCTCCTGCAGGCCTCTAGAATCCGCGT AGATGCC
7697	ClaI-M-ZBP-1-F	ACTGATCGATATGAACAAGCTTTACATC GGCAAC
7698	MluI-ZBP-1-R	ACTGACGCGTTTCCTTCCTCCGAGCCTG
7699	Fubp3-Bsp119I-F	ACTGTTCGAAGACGAGCTGTACAAGCT TAAGC
7700	Fubp3-MluI-R	GACTCACGCGTCGCCTGCTCCTGGCTGTG
7752	Norad_new-R	CCAAGAGATGCATACATCAATGGC
7753	Norad_new2-R	GAGGTGGCTTAACAATTCCATGT
7754	arc HDR repair-MHA-F	T*A*A*AGATTAAGAACAGAGTTTCTCTAA ATTCCTACCCCTCATCTGTCTG
7755	arc HDR repair-MHA-R	A*A*T*TAAGTATTTTAATGTGCCTTAGTG GAAGAGCAGAATGAGGAAGCC
7848	Norad MHA-p2341 repair-F	T*C*T*G*G*GAGTGACCTTTATTAGTCCAC CCCTTGGAGCTAGACATCCTGTACGAG TTTCTCTAAATTCCTCCTGC
7849	Norad MHA-p2341 repair-R	T*A*C*T*A*GCAAAGAGCCCTTCCCTTCAC CCTTCCTCTTCTCCCTCTTCCACCTCTG TGATGTCTGGTTTCCC

Table 11.: Oligonucleotides used for smiFISH experiments. Shown is only the primary sequence which is directly annealing to the RNA, the FLAP-Y sequence: 5'-TTACACTCGGACCTCGTCGACATGCATT-3' is added to the 3'end of each oligonucleotide. ID = database entry number.

ID	Name	Sequence (5' → 3')
7890	MS2-V5 FLAP-Y 01	TGATTGTGAAGTGTCGGGTG
7891	MS2-V5 FLAP-Y 02	TCCACCCTTGTGTATTGTAC
7892	MS2-V5 FLAP-Y 03	TGTAATGTGTCTGGAGGGTG
7893	MS2-V5 FLAP-Y 04	GCTTCTGTTTGATTGGATTT

7894	MS2-V5 FLAP-Y 05	GATGGTGATTCCTTGTTGTA
7895	MS2-V5 FLAP-Y 06	GTATATTGCACAGGGAATCC
7896	MS2-V5 FLAP-Y 07	GATATTCGGGAGGCGTGATC
7897	MS2-V5 FLAP-Y 08	ACGCACTGAATTCGAAAGCC
7898	MS2-V5 FLAP-Y 09	ATTCGACTCTGATTGGCTGC
7899	MS2-V5 FLAP-Y 10	CTCTTCGCGAAAGTCGACTT
7900	MS2-V5 FLAP-Y 11	TAAGAATGGCGCGAAGGCTG
7901	MS2-V5 FLAP-Y 12	GTAGGGGAGAGTGTGGTTTG
7902	MS2-V5 FLAP-Y 13	CAGGAACGCTGATGCTGTTC
7903	MS2-V5 FLAP-Y 14	TTTTCTTGAGTTGGGTACTG
7904	MS2-V5 FLAP-Y 15	TGATGCTGCATGGGGACATA
7905	MS2-V5 FLAP-Y 16	TTGGGGATGTATTCTTGGGG
7906	MS2-V5 FLAP-Y 17	TTGGTGCTCGGATGTGATTT
7907	MS2-V5 FLAP-Y 18	AAGAAACAACACTCCGAGCC
7908	MS2-V5 FLAP-Y 19	ATGGAGGGTTTGTCCAGTTG
7909	MS2-V5 FLAP-Y 20	TTTGTCTTGTTGGTGAGAGT
7910	MS2-V5 FLAP-Y 21	CTGATGCTGCTTCGAGAAGA
7911	MS2-V5 FLAP-Y 22	GTATGCTCGAGTGTTTCGAA
7912	MS2-V5 FLAP-Y 23	GATCGTCCACCCAAGAAATA
7913	MS2-V5 FLAP-Y 24	AATTCGTGAGAGCATGGGTG
7914	MS2-V5 FLAP-Y 25	TCGTATTGGACGTGGAACGA
7915	MS2-V5 FLAP-Y 26	TCGTGATCCCGAAAGGTAAG
7916	MS2-V5 FLAP-Y 27	ATCGTGCATGCTTGAATGTC
7917	MS2-V5 FLAP-Y 28	GTTGAGACTTGTGGAGCATG
7918	MS2-V5 FLAP-Y 29	TGAACCCATTTGGTAGTTTC
7919	MS2-V5 FLAP-Y 30	TTTGAGGTAGGAGTGGGTTC
7920	MS2-V5 FLAP-Y 31	TTGCCAGTTTTGTGGGAAGA
7921	MS2-V5 FLAP-Y 32	TTTGGTATGTTGGAATGGGC
7922	MS2-V5 FLAP-Y 33	GATGCTGTACCAGTAATTGT
7923	MS2-V5 FLAP-Y 34	TAGTAGTGAGAGATGTGGGC
7924	MS2-V5 FLAP-Y 35	TGCTGAACGGTTTGGTTTTT
7925	MS2-V5 FLAP-Y 36	TTGATTTTTCCGTGTGTACC
2726	MS2-V5 FLAP-Y 37	GTCTTTCGTATTTGTAAACC
2727	MS2-V5 FLAP-Y 38	TTGCGCTGGACGAAAGCGTG
2728	MS2-V5 FLAP-Y 39	CCGTCGGATGTTTTTCGTAA

2729	MS2-V5 FLAP-Y 40	GGTTGTAAGTTTGTGGGTTG
7930	OIP5-AS1 Flap-Y 01	CTGTTTCAAATGAAACTCAGCCTTCTGCTTG
7931	OIP5-AS1 Flap-Y 02	CAGTGGGGCTGTAGTACATTTGTTTCCAGGT
7932	OIP5-AS1 Flap-Y 03	GCTGCAAACAGGAAATGTAAATGGTCTGCAGA
7933	OIP5-AS1 Flap-Y 06	ACTAACCACAGATCTGTCAGTATTTTCAGGAC
7934	OIP5-AS1 Flap-Y 07	CTTTCTTCCTGGAGATGTTAGCAGGATAACTG
7935	OIP5-AS1 Flap-Y 08	GGACACATCTATATTAGACAAGGTTTCACACC
7936	OIP5-AS1 Flap-Y 09	TTTTCACTGATGGAATGACATATCACCAAGGG
7937	OIP5-AS1 Flap-Y 10	TCTCATTGATAAAGAATAAGCAGTCACAGGCC
7938	OIP5-AS1 Flap-Y 12	TCTTATTTGAGGTTTCTTTTCCACGATGACCC
7939	OIP5-AS1 Flap-Y 14	AGAGAAAGTCTACTCACTACGTAGCAACCTAG
7940	OIP5-AS1 Flap-Y 15	TAGGACTTCATGGAGAAATCTGTGATGTCTGG
7941	OIP5-AS1 Flap-Y 17	CAACATACTAGTGGAACACCATATATGTAGGC
7942	OIP5-AS1 Flap-Y 18	ATATGGAGGACAGCAAGTAAACACAGTGAGCA
7943	OIP5-AS1 Flap-Y 19	AAAGCTTTAGTCAAGAAATTGCAGTCTGGCCA
7944	OIP5-AS1 Flap-Y 20	CATGAGAATGTAAACTCCTGGAGGCCAAGAG
7945	OIP5-AS1 Flap-Y 21	GGAAAGGAAGGAAGCATGGCTACAAATCGACG
7946	OIP5-AS1 Flap-Y 22	ATAAAGCATATAGAGACGCTGAGTCTGATAGC
7947	OIP5-AS1 Flap-Y 23	GTTGGTCATTGGTAGTTGGTACCAGTATCATC
7948	OIP5-AS1 Flap-Y 24	TAATTGCTTAGAAATGCAGGTAGAAGCTCCAC
7949	OIP5-AS1 Flap-Y 25	GATGTGTTAAGTAAACACTGGGCAACTACAGC
7950	OIP5-AS1 Flap-Y 26	AAGATGGGTTTAACAGACACCTGGAAGGTGTC
7951	OIP5-AS1 Flap-Y 27	TCAATTCTGAACCACTGGAAAGTTACCATTCC
7952	OIP5-AS1 Flap-Y 29	TCAGAGGAATTACGGGACATAACAAGGATCCA
7953	OIP5-AS1 Flap-Y 30	AATGGGTCGGTTTTGTGATAAGTGCCAAGACT
7954	OIP5-AS1 Flap-Y 31	AGCAAAGACAGTTTTCTACGACAGTCTGTTCT
7955	OIP5-AS1 Flap-Y 32	GTGTCTTTTGTTTTCAAACATGCTTGGGGACT
7956	OIP5-AS1 Flap-Y 34	TTGTTTGTTTGAGGTGGAGTCTTACTGTCACC
7957	NORAD Flap-Y 01	AGGATGTCTAGCTCCAAGGGGTGGACTAA
7958	NORAD Flap-Y 02	AACAAATGCTTAGAGGGGTGTTACTTAAGCCC
7959	NORAD Flap-Y 03	AAGGTTGGGGTGGAGTTGAGAGCAGCTTTTTC
7960	NORAD Flap-Y 04	AAGCATTGGCAGAGTTCTGATTTAAATGCTGC
7961	NORAD Flap-Y 05	TATCGCTTCCAGAGGCCGTCTTAACAAC
7962	NORAD Flap-Y 06	ACCATCGTCTAGATATGGAGATACTAGCAAAG
7963	NORAD Flap-Y 07	CGCTGTAAACAGGATGGCATAGAGCTCTC

7964	NORAD Flap-Y 08	GTATCTTCTAAATGTGGCCATTTTGGCCTTG
7965	NORAD Flap-Y 09	AATGCAGACATAAGGGACAATGGTCAATGTGC
7966	NORAD Flap-Y 10	ATGGGTACTGCTCAGAGAATTGATCAATTTCC
7967	NORAD Flap-Y 11	CTTCTAAATACGAACATTCTGGTCTAGAACCC
7968	NORAD Flap-Y 12	ACATCGACTAGACGTAAATTAGGACTCCTATG
7969	NORAD Flap-Y 13	TTGAGTGTCTTCTAAATAGGAACATTCTGGCC
7970	NORAD Flap-Y 14	GACTAGATGTTGTCAATTAGGACTCGTCTGTC
7971	NORAD Flap-Y 15	ACAAGTACAGGCACGTGTAATGCTTAGAGATG
7972	NORAD Flap-Y 16	CATGTCATCTAACCCACTTAACAAACGTGGA
7973	NORAD Flap-Y 17	TTGACATATTCTTCTAGATCCTGTGTGTAGGC
7974	NORAD Flap-Y 18	ACTTAAGTTTGTCCGCTATATACACAGTAGGG
7975	NORAD Flap-Y 19	CCTCATAAACACACTGGACACTGGTAGAAAG
7976	NORAD Flap-Y 20	GGACATTCACAAAGGGACTGATTTACTACATC
7977	NORAD Flap-Y 21	GAGTATTACTCAACAGGTGATTTGGCCATTCC

Table 12.: Oligonucleotides used for HyPro experiments. Shown is the sequence which directly anneals to NEAT1, the digoxygenin is added at the 3'end of each oligonucleotide. ID = database entry number.

ID	**Name**	**Sequence (5' → 3')**
7770	hNEAT1-probe1	TCGTCTCATCTAACTCAGTC
7771	hNEAT1-probe2	GACCTAGTCTCCTTGCCAAG
7772	hNEAT1-probe3	TGCGGATATTTTCCATGCAG
7773	hNEAT1-probe4	ACAAGTTGAAGATTAGCCCT
7774	hNEAT1-probe5	AGCCCTTGGTCTGGAAAAAA
7775	hNEAT1-probe6	CGAGCTAAGTTCAGTTCCAC
7776	hNEAT1-probe7	GGCCGAGCGAAAATTACATA
7777	hNEAT1-probe8	TTAACTCCACATCACTCCTC
7778	hNEAT1-probe9	CTGTCAAACATGCTAGGTGC
7779	hNEAT1-probe10	AAGCGTTGGTCAATGTTGTC
7780	hNEAT1-probe11	TGCAGCATCTGAAAACCTTT
7781	hNEAT1-probe12	ACCGGAGGCTCAATTTAGAA
7782	hNEAT1-probe13	CAAGGTTCCAAGCACAAAAC
7783	hNEAT1-probe14	ACAGCTTAGGGATCTTCTTG

7784	hNEAT1-probe15	TGGCATCAACGTTAAAATGT
7785	hNEAT1-probe16	TCTACAAGGCATCAATCTGC
7786	hNEAT1-probe17	CAAACAGGTGGGTAGGTGAG
7787	hNEAT1-probe18	CTTCTCCGAGAAACGCACAA
7788	hNEAT1-probe19	CCAAGTTATTTCATCAGGCT
7789	hNEAT1-probe20	CACAACACAATGACACCCTT
7790	hNEAT1-probe21	AATGACCAACTTGTACCCTC
7791	hNEAT1-probe22	CAATGCCCAAACTAGACCTG
7792	hNEAT1-probe23	CTCCTAGTAATCTGCAATGC
7793	hNEAT1-probe24	AGCAAGAACAAAAGAGCACT
7794	hNEAT1-probe25	GGTCCTCTTACTAGAATGCC
7795	hNEAT1-probe26	ACTGTATCTCTAACCAACCC
7796	hNEAT1-probe27	CTAAGCAACTTCTCACTTCC
7797	hNEAT1-probe28	CTGTGTCACCTGTTTTCAGT
7798	hNEAT1-probe29	CCTTTGGTTCTCGGAAAACT
7799	hNEAT1-probe30	TGTGAGATGGCATCACACAC
7800	hNEAT1-probe31	ACAGTGACCACAAAAGGTTA
7801	hNEAT1-probe32	AGCAAAGGTACATGGATTCT
7802	hNEAT1-probe33	CATGTAGTAAAGGCACCTCG
7803	hNEAT1-probe34	CCCCAAGTCATTGGTTAAGA
7804	hNEAT1-probe35	TCCCAACGACAGTAATTGTT
7805	hNEAT1-probe36	CCCATACATGCGTGACTAAT
7806	hNEAT1-probe37	ATTTCACAACAGCATACCCG
7807	hNEAT1-probe38	CCAGTACTTTCAACCATCTA
7808	hNEAT1-probe39	AGTTCTTACCATACAGAGCA
7809	hNEAT1-probe40	TTGTGCTGTAAAGGGGAAGA

Table 13.: Primers encoding the gRNA sequence and complementary gRNA sequence (cgRNA) for CRISPR/Cas9 approaches. They were pairwise annealed and cloned into plasmid p2281 behind an U6 promoter, followed by a gRNA scaffold. Shown in red is the actual gRNA sequence. ID = database entry number.

ID	Name	Sequence (5' → 3')
7866	OIP5-AS1 cgRNA1a_F	AAACCACCCGGGTTCAAGCGATTCTC
7867	OIP5-AS1 gRNA1a_R	CACCGAGAATCGCTTGAACCCGGGT
7868	OIP5-AS1 cgRNA1b_F	AAACCGCTCAGCCTGATTTCTAGGC
7869	OIP5-AS1 gRNA1b_R	CACCGCCTAGAAATCAGGCTGAGCG
7870	OIP5-AS1 gRNA1c_F	CACCGAACTTGTGGTTGGGTCATCG
7871	OIP5-AS1 cgRNA1c_R	AAACCGATGACCCAACCACAAGTTC
7872	OIP5-AS1 cgRNA2a_F	AAACCCGATTATTCCACTAAGGCAC
7873	OIP5-AS1 gRNA2a_R	CACCGTGCCTTAGTGGAATAATCGG
7874	OIP5-AS1 cgRNA2b_F	AAACGGTGATATGTCATTCCATCC
7875	OIP5-AS1 gRNA2b_R	CACCGGATGGAATGACATATCACCA
7876	NORAD gRNA Ia_R	CACCGACGGCCGCCATTAGTCACCG
7877	NORAD cgRNA Ia_F	AAACCGGTGACTAATGGCGGCCGTC
7878	NORAD gRNA Ib_F	CACCGCCCACCTCGGTGACTAATGG
7879	NORAD cgRNA Ib_R	AAACCCATTAGTCACCGAGGTGGGC
7880	NORAD gRNA II_F	CACCGAGCGAAGTCCCGAACGACGA
7881	NORAD cgRNA II_R	AAACTCGTCGTTCGGGACTTCGCTC
7882	NORAD gRNA IIIa_F	CACCGTGTACAACGGCCTGTCGCAG
7883	NORAD cgRNA IIIa_R	AAACCTGCGACAGGCCGTTGTACAC
7884	NORAD gRNA IIIb_R	CACCGCTCAGCCTCTCGAGTAACTG
7885	NORAD cgRNA IIIb_F	AAACCAGTTACTCGAGAGGCTGAGC
7886	NORAD gRNA IVa_F	CACCGAGACATCCTGTACTTAGTCA
7887	NORAD cgRNA IVa_R	AAACTGACTAAGTACAGGATGTCTC
7888	NORAD gRNA IVb_F	CACCGCCTGTACTTAGTCACGGGGA
7889	NORAD cgRNA IVb_R	AAACTCCCCGTGACTAAGTACAGGC

6.7. Antibodies

Table 14.: Primary and secondary antibodies used for western blots, southern blots, and IF. AF = AlexaFluor.

Primary antibody	Species	Manufacturer, catalog number
α-Digoxygenin	sheep	Roche Deutschland Holding GmbH, 11093274910
α-FLAG	rat	Agilent, 200474
α-GFP, monoclonal	mouse	proteintech, 66002-1-Ig
α-GFP, polyclonal	chicken	abcam, ab13970
α-HA, monoclonal	rat	Roche, 1867423
α-hnRNP K	mouse	Santa Cruz Biotechnology, sc-28380
α-His	monoclonal	Thermo Fisher Scientific
α-Myc 9E10	mouse	Roche Deutschland Holding GmbH, 11667149001
α-PTBP1	mouse	Invitrogen, # 32-4800
Streptavidin-AF594		Life Tech ThermoScie, S11227
Streptavidin-HRP		Thermo Fisher Scientific 911 2112
α-V5	rabbit	cell signaling, 13202S

Secondary antibody	Species	Manufacturer
α-chicken-HRP	goat	abcam, ab97135
α-mouse HRP	rabbit	Dianova Jackson 515-035-003
α-rabbit-HRP	donkey	Amersham ECL, NA 934
α-rat-HRP	donkey	Rockland, 612-703-002

6.8. Kits

Table 15.: All commercial kits used during this work. The second column states the use of each kit.

Name of the Kit	Use	Manufacturer
QIAquick®PCR Purification	DNA purification after PCR	Qiagen GmbH
QIAGEN®Plasmid Midi Kit	isolation of plasmid DNA from *E.coli*	Qiagen GmbH
NucleoSpin Gel and PCR Clean-up	DNA purification after PCR and DNA extraction from agarose gel	Macherey-Nagel GmbH & Co.KG

NucleoSpin® Plasmid	isolation of plasmid DNA from *E.coli*	Macherey-Nagel GmbH & Co.KG
PCR DIG Probe Synthesis Kit	DIG labeled probe synthesis	Roche Deutschland Holding GmbH
DIG Oligonucleotide 3'-End Labeling Kit	3'-end labeling of oligonucleotides with DIG	Roche Deutschland Holding GmbH
Pierce™ ECL Western blotting substrate	chemiluminescene	Thermo Fisher Scientific GmbH
QIAquick® Gel Extraction	DNA extraction from agarose gel	Qiagen GmbH
4D-Nucleofector™X Kit S	nucleofection of HeLa cells	Lonza AG

7 Methods

7.1. Cell culture

HeLa WT, HeLa-11ht-EM2 as well as MEF WT and MEF β-actin MS2 cells were cultured in DMEM, supplemented with 10% FCS. Stable cells lines produced from HeLa-11ht-EM2 were grown in the same medium. All cell lines were cultivated without penicillin/streptomycin (P/S) for experiments, 1% P/S was added only to CRISPR clones and cells after FACS sorting. The cells were maintained at 37°C in a humidified incubator with an atmosphere of 5% CO_2.
Every third or forth day the cells were passaged in a 1:10 or 1:50 ratio, respectively. Therefore, cells were washed with 1xPBS and detached by Trypsin-EDTA treatment at 37°C. After approximately 5 min, cells were suspended in medium and seeded onto a new culture dish or flask. For experiments, cells were counted to determine the cell number before seeding. Therefore, 10 µL of cell suspension were mixed with 10 µL Trypan blue and a 10 µL sample was counted with the help of a cell counter.
For storing (created) cell lines, cells were trypsinized, spun down in a centrifuge for 5 min at 2000 rpm and resuspended in freezing medium. 1 mL of cell suspension was frozen in labeled cryo vials, slowly cooled down to -80°C in a Mr. Frosty^{TM} and finally stored in liquid nitrogen.

7.1.1. Transfection

24 h before a transfection cells were seeded, in most cases onto a 6-well-plate. On the next day cells were washed and the medium was replaced.
Lipofectamine^{TM} 2000 (Inivitrogen) and Opti-MEM^{TM} were used according to the manufac-

ture's instruction. Alternatively, FuGENE® HD Transfection Reagent (Promega) was used according to the manufacture's protocol together with Opti-MEM^{TM}. Cells were harvested (see section 7.1.6) or fixed for microscopy (see section 7.2.1) 20 to 24 h after the transfection. In case stable cell lines should be generated, it was continued with section 7.1.3

7.1.2. CRISPR/Cas9 mediated knock-in

For the design of gRNAs the free online tool from benchling.com was used. The on- and off-target score of a subset of handpicked gRNAs cutting in a desired region was compared with the predicted specificity and efficiency of Integrated DNA Technologies, Inc., (US). The best performing gRNAs were ordered as oligonucleotides (listed in Table 13) and cloned into plasmid p2281.
Each plasmid was transfected (see section 7.1.1 for protocol) and the gDNA was isolated after 48 h. For analysis of the *in cellulo* cutting efficiency, PCR over each cutting site was performed (see section 7.5.2 for protocol) and the resulting PCR product was send for sequencing. The chromatogram of the sequencing result was uploaded to tide.nki.nl and ice.synthego.com/. Both websites calculated the cutting efficiency, based on which the final decision for one gRNA was made.
For CRISPR/Cas9 mediated knock-in the OIP5-AS1 gRNA 2a (sequence: TGCCTTAGTG-GAATAATCGG) was ordered from Integrated DNA Technologies, Inc., (US) as chemically stabilized molecules (Alt-R ™CRISPR-Cas9 sgRNA). It was received as lyophilzed pellet, diluted with RNAse-free water to a final concentration of 100 µM and stored at -20°C.
The ribonucleoprotein particle (RNP) complex was prepared by mixing 2 µL recombinant Cas9 protein (61 µM), 2.8 µL (100 µM) gRNA, 1.2 µL 1x PBS and incubating everything for 20 min at RT. 2x10^5 cells were pelleted by centrifugation at 100xg for 10 min and resuspended in 20 µL Nucleofection mix (Lonza nucleofection supplement and Lonza Nucleofector solution in a 1:4.5 ratio). Prior to nucleofection, 0.5 to 2 µL RNP complex and 400 to 1000 ng of repair template were added to the cells and thoroughly mixed. The total volume was transferred into the Nucleocuvette Vessel, with a closed lid it was placed into the retainer of the 4D-Nucleofector X Unit and the program was started. After run completion, 80 µL pre-warmed medium was added and cells were seeded into 10 cm^2 dishes. As untransfected negative control, 2x10^5 cells were seeded into a second 10 cm^2 dish.
Alternatively, if a plasmid-based Cas9 enzyme was used, the above mentioned (see section 7.1.1) transfection protocol was followed, the gRNAs used are listed in Table 13.

7.1.3. Antibiotic selection and colony "picking" for monoclonal cell line generation

Firstly, the concentration of each antibiotic to kill WT cells was determined with an antibiotic killing curve. For that, the antibiotic concentration was tritrated up, until all cells were dead.

To generate stable cell lines, three days after transfection the medium was changed to antibiotic selection conditions and the selective pressure was kept for approximately seven days. Single colonies were then picked from the cell culture plate with a 200 µL pipette. The colonies were transferred into a 48-well-plate and expanded to bigger vessels over time.

7.1.4. Fluorescence activated cell sorting (FACS)

Cells were detached by Trypsin-EDTA, spun down and resuspended and thoroughly dissociated in FACS buffer. 3-4 mL of cell suspension with a concentration of 2.5 - 3.5x10^6 cells/mL were passed through the net of a FACS tube to remove cell aggregates. At the FACS Core Facility of the University Hospital Tübingen the cells were sorted and gated based on eGFP expression. Cells were collected in a 96-well-plate or tube and brought back in culture afterwards.

7.1.5. Biotinylation assay

Cells expressing a biotin ligase either transiently or stably were cultivated in medium containing either 50 µM biotin or dimethyl sulfoxide (DMSO) as a negative control. The incubation time was defined by the biotin ligase and the aim of the experiment. To stop the labeling reaction, all samples were placed on ice and and washed 5 times with cold (4°C) 1x PBS to remove excess biotin. Then cells were harvested and lysed as described in section 7.1.6.

7.1.6. Harvesting and lysis of cells

Cells were washed two times with cold 1x PBS and removed from the 6-well-plate in a volume of 500 µL 1x PBS with a cell scraper. The whole procedure was performed on ice. After spinning the cells down for 15 min at 1000 rpm and 4°C the supernatant was removed and the pellets were either stored at -20°C or lysed directly.
Lysis buffer was freshly prepared and 100 µL each were added to the cell pellets. After incubation for 15 min on ice the lysate was cleared by centrifugation for 15 min at 15000 g at 4°C. The supernatant was pipetted into fresh Eppendorf tubes and the protein amount quantified using Bradford assay.

7.2. Microscopy and imaging techniques

7.2.1. Single-molecule (inexpensive) fluorescence in-situ hybridization (sm(i)FISH)

In order to yield a homogenous probe set of equimolar concentration, all primary smiFISH probes for one target (listed in Table 11) containing the Flap-Y overhang and RNA specific sequence were mixed to yield a 100 µM stock solution. Next, a 20 µM working dilution was prepared. The fluorescent, 28 nt long FLAP-Y oligonucleotide is conjugated to two Cy3 or Cy5 moieties at the 5' and 3' termini. It is delivered lyophilized and upon arrival resuspended in TE buffer to a final concentration of 100 µM. For long-term storage, aliquots were stored at -20°C in the dark.
To anneal the primary probes to the fluorophor-labeled FLAP-Y oligonucleotide all RNase-free reagents were added to a PCR tube according to Table 16 and incubated in a thermocycler following the program in Table 17.

Table 16.: FLAP annealing reaction mix. The Flap-Y oligonucleotide was either linked to Cy5 or Cy3 moieties at each end and always added in excess to prevent non specific binding of the FLAP-Y linkers.

Component	Volume [µL]	Final concentration
10x NEB buffer r3.1	1	1x
primary probe set (20 µM)	2	4 µM
Flap-Y oligonucleotide (100 µM)	0.5	5 µM
nuclease-free H_2O	6.5	-

Table 17.: FLAP annealing program. The lid temperature was set to 110°C.

Step	Temperature	Time
Denaturation	85 °C	3 min
Annealing	65 °C	3 min
Annealing	25 °C	5 min

To image RNA in cells, Ø 12 mm microscopy glass cover slides were placed in a 24-well-plate and $5x10^4$ cells were seeded per well. The cells were cultivated for 24 h and from this point on, only RNase-free reagents were used.
To prepare cells for fluorescent staining, they were washed two times with 1x PBS and fixed with 4% paraformaldehyde freshly prepared in PBS for 20 min at RT under a fume hood. The fixation solution was removed by washing twice with 1x PBS and the cells were permeabilized

with 70% ethanol for at least 1 h up to and over night incubation at 4°C. Similarly, the ethanol was removed by washing with 1x PBS.
For the *in situ* hybridization, fixed and permeabilized cells were incubate in 15% formamide freshly prepared in 1x SSC buffer for 15 min at RT. Meanwhile, the smiFISH hybridization mix was prepared according to Table 18. The described volume of 100 µL is sufficient for two coverslips. 50 µL of the hybridization mix were dropped on top of a parafilm sheet placed in a humidified chamber. The coverslip was carefully layed on the drop with the cells facing downwards. The plate was sealed with Parafilm and incubated over night at 37°C protected from light.
The following day, the coverslides were washed two times with 15% formamide in 1x SSC for 30 min at 37°C. Next, the samples were incubated with 0.2 µg/mL of 4,6-diamidino-2-phenylindole (DAPI) in 1x PBS for 10 minutes. The counterstain was removed by three washing steps with 1xPBS. Finally, excess liquid was removed from the coverslides before mounting onto microscopy slides. For long-term storage, slides were stored at 4°C.
The majority of the protocol is adapted from Tsanov et al. (2016) [65].

Table 18.: smiFISH hybridization mix. All reagents were RNAse-free. The mixture was prepared on ice and mixed well by vortexing.

Component	Volume [µL]	Final concentration
20x SSC	5	1x
Formamide	15	15%
50 mg/mL BSA	9	4.5 mg/mL
20% dextran sulfate	53	10.6%
VRC (200 mM)	1	2 mM
tRNA (10 mg/mL)	4.3	0.425 mg/mL
RiboLock (40 U/µL)	1	0.4 U/µL
FLAP-annealed probe set	3	120 nM
nuclease-free H_2O	8.7	-

All samples were imaged using a Zeiss Observer Z1 inverted microscope.

7.2.2. Immunofluorescence (IF) and smFISH after *in vitro* biotinylation

Microscopy glass cover slides with ∅ 12 mm were placed in a 24-well-plate and 5x10^4 HeLa cells were seeded onto the slides per well. The cells were cultivated for 24 h to obtain 80-90% confluence and from this point on, only RNase-free reagents were used.
Before cells were fixed, they were washed with 1x PBS and then incubated with clear, freshly

prepared 0.5 mg/mL DSP solution for 30–45 min in PBS at RT. The fixation solution was removed by washing three times for 5 min each with 1x PBS supplemented with 20 mM Tris-HCl, pH 7.5. Then, cells were permeabilized with 70% ethanol for at least 1 h up to over night incubation at 4°C.
Before hybridization, fixed and permeabilized cells were rinsed with 10% formamide freshly prepared in 2x SSC buffer for 5 min at RT. Meanwhile, the digoxigenin-labeled oligonucleotides or unlabeled oligonucleotides for control experiments were mixed with fluorophore labeled oligonucleotides and diluted in hybridization buffer.

Table 19.: Hybridization buffer. All reagents were RNAse-free.

Component	Amount	Final concentration
20x SSC	1 mL	2x
Formamide	1 mL	10%
Dextran sulfate	1 g	10%
nuclease-free H_2O	up tp 10 mL	-

50 µL of the hybridization mix were dropped on top of a parafilm sheet placed in a humidified chamber. The coverslip with fixed and permeabilized cells was carefully laid onto the drop with cells facing downwards. The plate was sealed with Parafilm and incubated over night at 37°C. The next day, the coverslides were washed with 10% formamide in 2x SSC for 30 min at 37°C. Next, the samples were washed with 1xSSC for 15 min. To prevent non-specific binding, the cells were blocked with 0.8% (w/v) bovine serum albumin (BSA) in 4x SSC supplemented with 0.1 units/mL RNase inhibitor for 30 min at RT. For HyPro conjugation, the enzyme was diluted to 2.7 ng/mL in the blocking buffer described above, and incubated at RT for 1 h.
Before HyPro labeling, cells were washed once with 4x SSC, once with 4xSSC and 0.1% Triton X-100, and again with 4xSSC, 10 min each wash. Then the cells were incubated in 1x PBS (250 µL/well) and after 5 min the same volume of 1x PBS containing 1 mM biotin-phenol and 0.2 mM H_2O_2 was added and everything was incubated for 1 min at RT. As positive control, cells were incubated with 250 µL/well with 5.4 ng/mL HyPro protein (HyPro-infusion), then the same volume PBS supplemented with 1 mM biotin-phenol and 0.2 mM H_2O_2 was added for 1 min. The final labeling reaction occurred consequently with 0.5 mM biotin phenol and 0.1 mM H_2O_2. The labeling mixture was aspirated and 3 washes with Quencher solution for 30 s–1 min each wash followed immediately.

Table 20.: Quencher solution. All reagents were RNAse-free.

Component	Volume [mL]	Final concentration
sodium ascorbate (1 M)	0.5	10 mM
Trolox (500 mM)	0.5	5 mM
1x PBS	49	1x

To prepare the cells for the antibody staining, the coverslips were rinsed with 4x SSC. The streptavidin AlexaFluor594 conjugate was diluted 1:800 in 0.8% BSA in 4xSSC and 50 µL was placed on top of parafilm in a humidified chamber. The attached cells on the coverslide were carefully placed upside down onto the solution, the chamber was sealed and incubated for 1 h at RT protected from light. The coverslides were then washed with 4x SSC, 4xSSC and 0.1% Triton X-100 and again with 4xSSC, 10 min each. Next, cells were rinsed with 1xPBS, and 0.5 µg/mL of 4,6-diamidino-2-phenylindole (DAPI) in 1xPBS was incubated for 5 minutes. The counterstain was removed by one more washing step with 1x PBS. Finally, excess liquid was removed from the coverslides before mounting onto microscopy slides. For long-term storage, slides were stored at 4°C.

7.3. Expression and purification of recombinant proteins

Steps required for purification of the recombinant hybridization-proximity labeling (HyPro) enzyme were conducted by Orit Hermesh following the protocol published by Yap et al. (2022) [66]. The purified HyPro protein was aliquoted, snap-frozen in liquid nitrogen, and stored at -80°C.

7.4. Protein analysis

7.4.1. Bradford assay

To determine the total protein concentration, 1-10 µL of total protein lysate was diluted in 1 mL of 1x ROTI®Quant solution (Carl Roth GmbH + Co.KG). The same volume of lysis buffer was mixed with 1x ROTI®Quant solution, this samples served as blank. After 5 min incubation at RT, the concentration was determined measuring the absorbance at 595 nm in a spectrometer.

7.4.2. Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE)

For separation of proteins according to size sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used. Cell lysates or purified proteins were incubated at 95°C for 5 to 10 min with the appropriate amount of 5x or 1x Laemmli sample buffer for denaturation. Proteins were loaded on gels consisting of a 8 to 12% polyacrylamide resolving gel and a 6% stacking gel above. For size comparison 4 µL PageRuler Prestained Protein Ladder (Thermo Fisher Scientific) was loaded. Electrophoresis was performed in SDS-PAGE running buffer using a constant voltage of 80 V for approximately 20 min and of 120 V afterwards. Proteins were detected by Coomassie Blue staining (see section 7.4.3) or Western Blot (see section 7.4.4).

7.4.3. Coomassie Blue staining

After SDS-PAGE, gels were stained in Coomassie staining solution for 30 min. Gels were destained in Coomassie destaining solution until no background was distinguishable any more. A picutre was taken with the ChemiDoc imager.

7.4.4. Western blot

After performing SDS-PAGE, the stacking gel was removed and the remaining gel was incubated for ~ 3 min in 1x Towbin buffer. The cut PVDF (0.45 µm) membrane was activated in methanol. For blotting, the resolving gel was placed on three soaked Whatman® cellulose papers. The membrane was carefully placed on the gel and three more soaked Whatman® cellulose papers were added before the air bubbles were removed with a roller. On both sides plastic sponges were placed and everything was fixed by a plastic cassette. The cassette was inserted into a wet blotting chamber filled with 1x Towbin buffer and equipped with an ice pack. The proteins were transferred at 4°C for 80 min at 360 mA constant current.
After the transfer the membrane was incubated in blocking buffer for at least 30 min at RT. The primary antibody was diluted in 3% milk in TBS-T buffer and incubated with the membrane for at least one hour at room temperature or over night at 4°C. After three washing steps with 1x TBS-T buffer for 10 min each, the secondary antibody fused to horse radish peroxidase was diluted 1:1000 in 3% milk in TBS-T buffer and incubated with the membrane for at least one hour at RT or over night at 4°C. Finally, the membrane was washed three more times with 1x TBS-T buffer. For the chemiluminescent immunodetection of proteins, the membrane was incubated with a 1:1 mixture of component 1 and 2 of the ECL solution. The emerging chemoluminescence signal was detected with the ChemiDoc imager.

7.4.5. Ponceau S staining

To control the success of the Western Blot transfer and equal loading of protein amounts, a Ponceau S staining can be done. After the transfer, the membrane was incubated for 10 min in Ponceau S staining solution. Three washing steps with ddH_2O were performed before the membrane was imaged by the ChemiDoc imager.

7.5. Cloning methods

7.5.1. Agarose gel electrophoresis

The size of DNA fragments was analyzed by agarose gel electrophoresis. First, 1% (w/v) agarose was melted in TAE buffer. The solution was allowed to cool down and 0.01% (v/v) GelRed® was added. The gel was cast and cooled down to room temperature. DNA samples were mixed with 6x DNA loading dye and loaded onto the solidified gel. For size comparison 4 µL Gene Ruler^{TM} 1 kb Plus DNA Ladder (Thermo Fisher Scientific) were loaded. Electrophoresis was carried out in TAE buffer using a constant voltage between 80 to 130 V until the fragments were sufficiently separated. DNA was visualized by UV light using the iX 20 Imager (Intas Science Imgaging Instruments GmbH) with a wavelength of $\lambda = 254$ nm.

7.5.2. Polymerase chain reaction (PCR)

DNA fragment amplification via PCR was conducted using the following reaction setup. The amounts of DNA template varied from 100 pg for DIG-labeled probe synthesis, 50 ng for cloning to 100 ng for genotyping.

Table 21.: PCR reaction mixture. The mixture was prepared on ice and mixed well by vortexing.

Component	Volume [µL]	Final concentration
5x Herculase buffer	10	1x
Template	x	-
10 mM dNTPs	2	1 mM
10 µM Primer forward	1.25	25 nM
10 µM Primer reverse	1.25	25 nM
Herculase polymerase	0.5 µL	25 U/ml
nuclease-free H_2O	add to 50 µL	

In case the reaction was not successful in a first trial, 2 - 6 % DMSO and/ or 1 mM $MgCl_2$ were added to unfold super-coiled vectors and to increase specificity, respectively.
The PCR program was the following:

Table 22.: Standard PCR program. The lid temperature was set to 110°C.

Step	**Temperature** [°C]	**Time** [s]	
Initial denaturation	95	300	
Denaturation	95	30	30 cycles
Annealing	- [1]	30	
Elongation	72	30 s/ kb	
Final elongation	72	300	
Hold	8	∞	

[1] temperature depends on used primers

Success of the PCR was analysed by agarose gel electrophoresis (see section 7.5.1). If the PCR product had the appropriate size in comparison to the marker, it was purified with the QIAquick® PCR Purification Kit or bands were extracted from the agarose gel with the QIAquick® Gel Extraction Kit (both Qiagen GmbH). The procedure was performed according to the manufacturer's protocol. The DNA product was eluted and stored at -20°C.

7.5.3. Digoxygenin labeled probe synthesis for Southern blot analysis

The labeling of dsDNA probes with digoxygenin (DIG) was based on a commercial kit from Roche Deutschland Holding GmbH, (GE). The PCR reaction was prepared as described by the manufacturer and contained DIG-dUTP nucleotides. To generate MS2 probes, the primers #7234 and #7217, #7236 and #7237 as well as the plasmid templates p2314 and p2296 were used, respectively. Additionally, two plasmids for the lncRNA specific DIG probe generation were cloned, these are p2324 for NORAD and p2325for OIP5-AS1.
To verify the DIG labeling, labeled and unlabeled samples were run on an agarose gel. The successful integration of DIG-dUTP nucleotides into the probe was displayed by a size shift in the labeled sample (see Results Figure 11).

7.5.4. Restriction enzyme digestion

The digestion reaction to create sticky ends for ligation contained 2 µg circular plasmid or 1 µg purified insert from PCR reactions. To the DNA 1 µL of each restriction enzyme and 2 µL of the suitable 10x buffer were added. The reaction was filled up with nuclease-free water to an end volume of 20 µL, mixed well and incubated at 37°C for at least 30 min up to 2 hours.
Clean up of the PCR product was done with the QIAquick® PCR Purification Kit (Qiagen GmbH). The procedure was performed according to the manufacturer's protocol. The complete volume containing the digested vector was mixed with 6x DNA loading dye and agarose gel electrophoresis (see section 7.5.1) was performed. The linear vector was excised from the gel with a clean scalpel and isolated with the help of QIAquick® Gel Extraction Kit (Qiagen GmbH). The digested products were stored at -20°C.

7.5.5. Dephosphorylation of restriction digested vectors

In order to avoid self-ligation, the digested vector (see section 7.5.4) was mixed with the appropriate amount of 10x phosphatase buffer. To dephosphorylate 1 to 5 µg of plasmid DNA 1 µL of antarctic phospatase was added. The reaction was incubated at 37°C for 10 min. For inactivation the temperature was set to 65°C and the mixture was incubated for 5 min. The thereby dephosphorylated vector was ready to use for ligation.

7.5.6. Ligation

To ligate an insert into a plasmid backbone (both cut by matching restriction enzymes, see section 7.5.4) they were mixed with a molar ratio of 3:1. 2 µL 10x T4 DNA Ligase buffer and 1 µL T4 ligase were added. The reaction was filled up to 20 µL with nuclease-free water and

the mixture was incubated for a minimum of 2 h at room temperature or over night at 4°C. Afterwards, the T4 ligase was heat inactivated for 5 min at 80°C.
In the next step, 5 µL of ligated plasmid were transformed via heat shock in competent *E.coli* cells (see section 7.5.8).

7.5.7. Gibson assembly

For Gibson assembly the primer were designed according to the manufacturer's instruction (New England BioLabs). Two PCRs amplifying the insert and vector (following section 7.5.2) were completed and the products were purified.
To get rid of the vector template, the complete purified vector was subjected to DpnI digestion. Therefor, 1 µL of DpnI enzyme as well as the appropriate amount of 10x CutSmart buffer were incubated with the vector at 37°C for 30 min. For inactivation the mixture was placed at 80°C for 10 min.
From the purified PCR products 200 ng of linear vector and 500 ng of insert were mixed with 2x Gibson assembly master mix. The volume was adapted with nuclease-free water. The reaction was incubated at 50°C for 60 min.
5 µL of the Gibson assembly reaction were transformed into competent *E.coli* cells by heat shock (see section 7.5.8).

7.5.8. Bacterial transformation

Chemical competent DH5α or Top10 *E.coli* cells for DNA preparation were thawed on ice. The competent cells were carefully mixed with 100 ng plasmid DNA for propagation or with 5 µL of a ligation reaction mix. The transformation mix was incubated on ice for 10 min before the heat shock was applied for 90 s at 42°C. After another 10 min on ice, the complete cell suspension was added to 3 ml LB medium containing the corresponding antibiotic for selection and was incubated over night at 37°C while shaking. For cloning, the cell suspension was inoculated on LB agar plates containing the corresponding antibiotic and incubated over night at 37°C.

7.5.9. Low scale plasmid preparation (Mini Preparation)

Low scale DNA preparation was done with the NucleoSpin® Plasmid Kit. The procedure was performed according to the manufacturer's protocol with a 3 mL bacterial overnight culture. The LB medium was supplemented with antibiotics and inoculated with a single *E.coli* colony which was picked from an agar plate containing the same antibiotic. The test tube was shaken at 200 rpm over night at 37°C. The concentration of the isolated plasmid DNA was determined

with a NanoDrop photometer (see section 7.5.12) after elution and kept at -20°C for long term storage.

7.5.10. Large scale plasmid preparation (Midi Preparation)

A flask with 100 mL LB medium supplemented with the desired antibiotic for selection were inoculated with *E. coli* cells carrying the plasmid of interest. The bacteria suspension was grown overnight at 37°C while shaking at 200 rpm. To harvest the overnight culture, it was centrifuged for 15 min at 6000 g at 4°C. Isolation and purification of plasmid DNA was performed using the QIAGEN®Plasmid Midi Kit according to the manufacturer's instruction. The DNA was solved in nuclease-free water and the concentration was determined with a NanoDrop photometer (see section 7.5.12).

7.5.11. Plasmid restriction digestion

For test digestions after low scale plasmid preparation 3 µL plasmid, 1 µL 10x buffer and 0.3 µL of a suitable restriction enzyme were mixed. The reaction was filled up with water to a final volume of 10 µL and incubated for 1 h at 37°C. To analyze cloning efficiency, the fragments were analyzed on a 1% (w/v) agarose gel (see section 7.5.1).

7.5.12. DNA/RNA concentration measurement

For DNA/ RNA concentration measurement a NanoDrop photometer was used. As blank 1.5 µL of nuclease-free water or elution buffer were utilized and 1.5 µL of a DNA/RNA sample were measured at a wavelength of $\lambda = 260$ nm.

7.5.13. DNA sequencing

For sequencing, DNA (for amount see table below) with a final primer concentration of 2.5 µM were mixed in a test tube. The volume was filled up to 10 µL with nuclease-free water. The samples were labeled with barcode stickers and send to Eurofins Genomics (GE).

DNA type and size	**amount of DNA** [ng]
Plasmid DNA up to 15 kb	500 - 1000
PCR product with 150 - 300 bp	10

PCR product with 300 - 1000 bp	50
PCR product with 1000 - 3000 bp	100

7.6. DNA/RNA purification and analysis

7.6.1. Genomic DNA extraction with phenol/chloroform

A confluent T75 cm^2 flask of cells was harvested by trypsinization. The cell suspension was centrifuged at 2000 rpm for 5 min in a 15 mL tube. The supernatant was discarded and the cell pellet was washed two times by resuspending it with 5 mL of ice-cold PBS. Cells were lysed in a safe-seal microcentrifuge tubes with 1 mL of lysis buffer for gDNA isolation at 55°C over night while shaking (300 rpm). The lysate was mixed with 350 µL of phenol/chloroform/isoamyl alcohol (PCI) and rotated at RT for 60 min before centrifugation at 10000 rpm for 5 min. The upper aqueous layer, containing the gDNA, was transferred into a new tube and mixed with 1/10 volume of 3 M NaAc (pH 5.2) and three 3 volumes of EtOH. Inversion of the tube leads to DNA precipitation. The precipitated DNA was pelleted by centrifugation for 10 min at 10000 rpm at 4°C and the pellet was washed with 1 mL of 70% EtOH. The DNA pellet was air-dried at RT for 10 min and eventually solved in 100-500 µL 1x TE buffer. To facilitate solubilization, the samples can be shaken gently for 4 h at 55°C and stored overnight at 4°C. The concentration was determined by nanodrop measurements and DNA was keept at 4°C until further usage.

7.6.2. Quick genomic DNA extraction

For rapid DNA extraction, the QuickExtract™DNA Extraction Solution (Lucigen®) was used according to the manufacturer's instructions. Briefly, for genotyping of CRISPR clones by PCR, cells were seeded in a 96-well-plate, grown to confluency and washed in 1x PBS before lysis. 100 µL QuickExtract™solution was added per well, thoroughly resuspended and the whole volume was transferred to microcentrifuge tubes. The mixture was then incubated at 65°C for 10 min, shortly vortexed, followed by a 5 min incubation at 98°C. The lysates were then centrifuged for 10 min at 3200 g and the supernatant was used directly as template for a genotyping PCRs.

7.6.3. Southern blot

To verify the correct genomic integration of the MS2 aptamer into the endogenous gene, 10-15 µg gDNA was digested with restriction enzymes by overnight incubation at 37°C. As a negative control, gDNA prepared from parental cell lines was also digested. Digested and undigested gDNA of each samples was loaded on an agarose gel to confirm successful digestion. After digestion it might be necessary to concentrate the DNA by precipitation to be in a suitable volume to be loaded into the gel. The digested gDNA samples were then loaded next to a DIG labeled DNA ladder onto a GelRed®-free 0.6% agarose gel and separated at 70 V for 190 min. As positive control, 50 – 100 ng of a linearized/ digested plasmid containing the gene specific probe or the MS2 aptamer were loaded. After the run finished, the agarose gel was washed twice for 20 min in denaturation solution at RT. Next, the gel was rinsed in deionised water and washed two times in neutralization solution for 20 min each. The pH of the gel was measured and washing was continued until it reached a pH 8 or lower. The transfer was assembled on a glass plate over a tank of 10x SSC, that was connected through a wick to three soaked sheets of Whatman® paper. The gel was placed upside down on the Whatman® paper followed by a positively charged nylon membrane (0.45 µm) and three more soaked Whatman® papers. Air bubbles were removed and the gel was framed with Parafilm™M to prevent short circuits. Lastly, two sheets of dry Whatman® paper and a stack of paper towels were added on top. A second glass plate, carefully balanced to be horizontally, and a weight on top enabled capillary forces the DNA transfer over night.
The next day, everything was disassembed and the position of the gel wells were labeled with a pencil on the membrane. The DNA was cross-linked to the nylon membrane in a UV oven for 20-40 s. The membrane was rinsed with 2x SSC, inserted in the hybridization bottle, and incubated in 25 mL pre-warmed DIG Easy Hybridization buffer for 30 min in the hybridization oven at the calculated hybridization temperature. The hybridization temperature was calculated as follows:

$$T_m = 49.82 + 0.42(\%G + C) - \frac{600}{length\ of\ probe\ in\ bp}$$

$$T_{hyp} = T_m(-20\ to\ 25°C)$$

Meanwhile, 25 µL of DIG-labeled probe were diluted in 100 µL nuclease-free water and boiled at 95°C for 5-10 min. The tube was immediately placed on ice and centrifuged for 30 s at 2000 g. The denatured probe was diluted in 25 mL pre-warmed DIG Easy Hybridization buffer, which was used to replace the current buffer and the probe was left for hybridization over night while rotating.
On the next day, the membrane was washed twice for 5 min in wash buffer at RT on a shaker and two more times for 15 min at 65°C in pre-warmed high-stringency wash buffer in the hybridization oven while rotating. From this step until the development, the protocol was performed at RT. Another washing step in high stringency wash buffer was performed for 10 min,

followed by one for 5 min in DIG1 wash buffer. Blocking was performed for at least 30 min in DIG blocking buffer. The anti-digoxygenin-alkaline phosphatase conjugate was diluted 1:20000 in DIG blocking buffer and incubated with the membrane while shaking for 30 min. Three more wash steps for 20 min with DIG1 wash buffer and two steps for 2 min at RT in DIG3 buffer were followed by the incubation of the membrane in a 1:100 dilution of CDP-Star in DIG3 buffer for 5 min before developing with the ChemiDoc imager.
For stripping and reprobing the membrane with the second probe, the membrane was rinsed with deionized water and two washing steps for 15 min at 37°C with stripping buffer in the hybridization oven were carried out. Then it was washed twice with 2xSSC buffer for 5 min at RT. For the incubation of the second DIG-labeled probe, all steps were performed as described above starting from the pre-hybridization step.

7.6.4. Total RNA isolation

During the whole procedure, RNAse-free solutions and plastic ware was used. A confluent T75 cm^2 flask of cells was harvested by trypsinization. Subsequently, the cell suspension was centrifuged at 2000 rpm for 5 min. The supernatant was discarded and the cells were washed twice by resuspending it with 5 mL of ice-cold PBS. The cell pellet was lysed in a safe-seal microcentrifuge tubes with 1 mL of lysis buffer by incubation on ice for 30 min. 1 mL of acidic Phenol:Chloroform was added to the lysed cells and everything was centrifuged for 10 min at 16000 g to achieve phase separation. The upper aqueous phase was collected and RNA was precipitate with 0.7 mL of isopropanol and a 20 min incubation on ice. To pellet the RNA, the tube was centrifuged for 30 min at 4°C with 16000 g. The RNA pellet was washed twice with 70% ethanol. After the last wash, the pellet was air-dried for 1 min. To dissolve the RNA, 50 µL of RNAse-free water were added and after a short incubation at RT it was thoroughly resuspended. The concentration was determined by nanodrop measurements and the RNA was aliquoted and stored at -80°C until further usage.

7.6.5. DNase treatment of isolated RNA

To treat 1 µg of isolated RNA, 1 µL RQ1 DNase and 3 µL of 10x DNase buffer were mixed. The sample was filled up to a final volume of 29 µL with RNase-free water. After incubation in a thermocycler for 30 min at 37°C, 1 µL of Stop Solution was added, followed by an incubation for another 10 min at 65°C to inactivate the DNase enzyme.

7.6.6. cDNA synthesis from DNase-treated RNA

To synthesize complementary DNA (cDNA) from DNase-treated RNA, reverse transcription was performed. The sample contained 1200 ng of RNA, whereas the negative control lacking the reverse transcriptase contained 800 ng of RNA. The components of these two approaches are listed in Table 24 below.

Table 24.: cDNA synthesis mixture. Everything was prepared on ice and mixed well by vortexing.

Component	+ rev. Transcriptase	- rev. Transcriptase
DNase-treated RNA	1200 ng	800 ng
10x RT buffer	2.4 µL	1.6 µL
25 mM dNTPs	1.2 µL	0.8 µL
10x random Primer	2.4 µL	1.6 µL
40 U/ µL RiboLock	2 µL	1.6 µL
20x reverse transcriptase	1.2 µL	-
nuclease-free H_2O	add up to 20 µL	add up to 15 µL

The + reverse transcriptase and - reverse transcriptase sample were spun down shortly and incubated in a thermocycler for 10 min at 25°C, followed by 120 min at 37°C and a final step for 5 min at 85°C to denature the enzyme.

7.6.7. RNA Immunoprecipitation and RNA isolation from beads

Magnetic Protein G Dynabeads were resuspended and an aliquot was taken, which was washed three times with NT2 buffer. The beads were then blocked with 5% BSA in NT2 buffer by overnight incubation at 4°C with rotation. After blocking, the beads were washed three times with NT2 buffer. After the final wash, the beads were divided into separate tubes based on the number of controls and samples required. Antibody was added to all samples, except the bead control and incubated for 6 h to overnight at 4°C while rotating. The next day, beads were washed three more times with NT2 buffer, and MEF cell lysate was added to all samples except the antibody control. The samples were adjusted to equal volumes with NT2 buffer and incubated overnight at 4°C with rotation. Following incubation, 10% of the supernatant was collected from each sample and mixed with 5× Laemmli buffer. The beads were then washed three times with NT2 buffer, and 10% of the first and last wash fractions were also collected and mixed with 5× Laemmli buffer for Western blot analysis.
Then, beads were split into two equal parts, one for RNA extraction and one for Western blot

analysis.
For protein elution, beads were resuspended in 1× Laemmli buffer. Additionally, a 10% input sample was prepared from crude MEF cell lysate with 5× Laemmli buffer. All samples were boiled for 10 min at 95°C to denature proteins. The remaining other half of bead fractions were resuspended in 100 µL in NT2 buffer, and 2 µL of 20 mg/mL Proteinase K and 1 µL of 10% SDS were added. Samples were incubated at 55°C for 45 min, followed by the addition of another 100 µL NT2 buffer. RNA was extracted by adding 200 µL acidic phenol:chloroform, and samples were centrifuged at 16,000 RCF at 4°C for 5 min. The aqueous phase containing the RNA was collected. For RNA precipitation, 700 µL of 100% ethanol, 90 µL of 2.68 M sodium acetate, and 5 µL of GlycoBlue were added, followed by gentle inversion. Samples were incubated overnight at -20°C to achieve maximal precipitation. The next day, samples were centrifuged at 16,000 RCF at 4°C for 15 min, and the resulting blue RNA pellet was washed once with 800 µL of 70% ethanol. After centrifugation at 16,000 RCF at 4°C for 10 min, the pellet was air-dried for 1 min and resuspended in 30 µL of RNase-free water.
In case GFP-coupled agarose beads were used, the procedure was the same as described above, except that the antibody coupling step was skipped.

7.6.8. Real-time quantitative PCR (RT-qPCR)

RT-qPCR was performed in 10 µL reaction volume, containing 5 µL of Fast SYBR™Green Mastermix, 4 µL cDNA as template and 0.5 µL each of a 10 mM forward and reverse primer pair. For each primer pair, one water control sample was run without cDNA template. Each run included at least technical duplicates for all samples that were reverse transcribed with the enzyme. Amplification was performed on a Real-time PCR device using the program described in the table below.

Table 25.: Standard RT-qPCR program. The lid temperature was set to 105°C.

Step	**Temperature** [°C]	**Time** [s]	
Initial denaturation	95	20	
Denaturation	95	3	} 40 cycles
Annealing and Elongation	60	30	
Melting point determination	90	15	
	60	60	
	60 to 95	0.3 °C/s	
Hold	8	∞	

After 40 cycles, the melting point of the generated double-stranded amplicons were determined by a gradual increase in temperature (see second half of Table 25) Analysis of the melting point

curves and the multicomponent plots of the RT-qPCR results were used to confirm that only one specific product was amplified. If the melting point and the multicomponent plot were unsuitable, the c_t value was excluded from further calculations.

From the detected average c_t values of the RT-qPCR samples, firstly the c_t value of the target RNA was normalized to the c_t value of the reference RNA, for both the control and test condition with the following equations:

$$\Delta c_t(control) = c_t(target,\ control) - c_t(reference,\ control)$$
$$\Delta c_t(test) = c_t(target,\ test) - c_t(reference,\ test)$$

Next, the Δc_t value of the test condition was normalized to the Δc_t value of the control:

$$\Delta\Delta c_t = \Delta c_t(test) - \Delta c_t(control)$$

From the $\Delta\Delta c_t$ value, the fold enrichment was calculated and then plotted.

$$fold\ enrichment = 2^{(-\Delta\Delta c_t)}$$

Part III.

Research

8

Establishing an *in vivo* proximity labeling assay for lncRNA

8.1. Introduction

8.1.1. The RNA subclass of lncRNAs

The eukaryotic genome is pervasively transcribed, but only 2% of all genomic sequences are protein-coding genes, comprising approximately 25,000 genes. The vast majority of transcribed sequences correspond to non-(protein-)coding RNA (ncRNA) genes, whose biological and pathological functions have been increasingly studied since the early 2000s [10].
Among ncRNAs, long non-coding RNAs (lncRNAs) form the largest group, characterized by transcripts exceeding 200 nucleotides in length, though some researchers proposed a threshold of 500 nucleotides [10]. LncRNA genes can be classified based on their genomic context, as shown in Figure 3: Antisense lncRNA genes overlap with protein-coding genes, intronic lncRNAs are located within introns of protein-coding genes, and lincRNAs genes are found in intergenic regions. Notably, certain genomic loci encode both protein-coding and lncRNA genes, which needs to be considered when experimentally disturbing these loci to study the biological function of either the mRNA/protein or the lncRNA. While this classification provides a structural framework and nomenclature, it offers limited insights into the functions or mechanisms of these non-coding transcripts [67].
Antisense (AS) lncRNAs comprise the largest class, encompassing 51% of all annotated lncRNAs in the human genome. Sense (S) lncRNAs, which overlap with a protein-coding gene and are transcribed in the same direction, make up 24% of lncRNAs, with 11% of lncRNAs falling into both Sense and Antisense categories (S & AS). Intergenic lncRNAs include 36% of all lncRNAs and thus constitute the second largest lncRNA group [68].

A significant proportion of lncRNA genes (over 60% in mammals) are transcribed from pro-

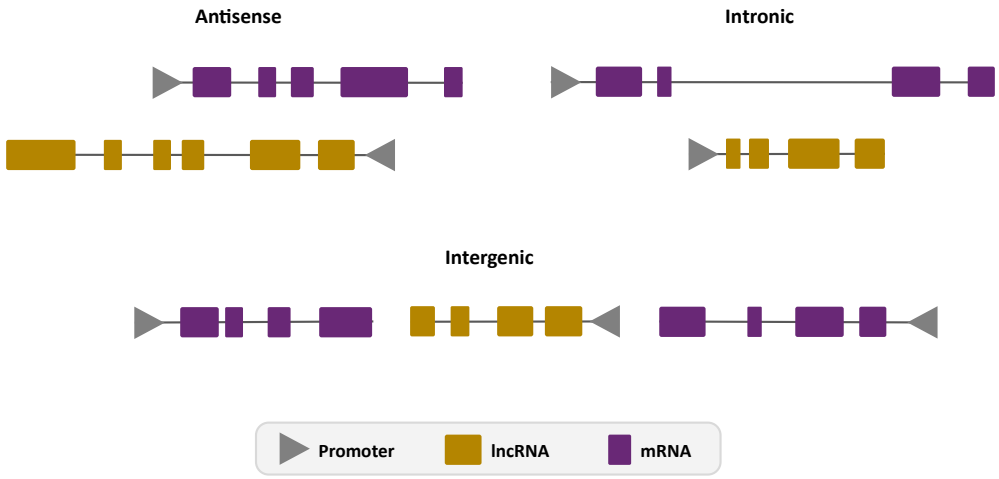


Figure 3.: Classification of lncRNAs, based on their position in the genome. Schematic representation of different lncRNA types based on their genomic position relative to protein-coding genes. The yellow/ purple boxes represent single exons. lncRNAs can be classified in (1) antisense, RNA genes that overlap with a protein-coding locus on the opposite strand; (2) intronic, RNA genes that are positioned within an intron of a coding gene on the same strand; (3) intergenic (lincRNA), transcripts positioned between two genes [67].

moters shared with protein-coding genes. This suggests potential co-regulation mediated by shared transcriptional and epigenetic factors, although the degree of this phenomenon is still debated [69].
LncRNAs generally exhibit low sequence conservation, which led to the early assumption that they are simply transcriptional noise. However, the relatively rapid evolution of lncRNA sequences can be explained by their lack of codon-preservation constraints. Interestingly, lncRNA promoters and splice junctions show a similar level of conservation as mRNAs. The secondary and tertiary structure of lncRNAs is also often highly conserved, which is likely more relevant for their biological function than the sequence itself [10], [70], [71]. On top of that, it was proposed that homologous lncRNAs may exist in species with large evolutionary distances, yet because of their low sequence similarity, they are not recognized as such [72].

The group of lncRNAs is diverse and heterogeneous, so despite advances in annotation, predictions about their molecular mechanisms cannot be made from the sequence alone. They localize to various cellular compartments, including the nucleus (e.g. XIST) and cytoplasm (e.g. NORAD), and may even shuttle between the two, performing distinct functions in each. While some are rather diffusely distributed, others localize to specific subcellular compartments such as e.g. NEAT1 which localizes exclusively to paraspeckles (elaborated in subsection 9.1.1) [67]. The cellular localization of lncRNAs is influenced by sequence motifs that mediate interactions with RNA-binding proteins (RBPs) or chromatin, directing nuclear retention or cytoplasmic export [73].
LncRNAs are predominantly transcribed by RNA polymerase II (Pol II), which is also responsible for mRNA transcription. This results in lncRNA transcripts that are 5'-capped, spliced, and polyadenylated, similar to mRNAs. Some lncRNAs exhibit less efficient splicing compared

to mRNA, leading to extensive alternative splicing. This generates multiple isoforms that (may) differ in subcellular localization and functional significance [73].
Functionally, lncRNAs operate at multiple regulatory levels: (1) serving as 3D macromolecular scaffolds for protein recruitment, (2) acting as molecular sponges for sequestering RNAs or proteins, and (3) functioning as regulatory elements modulating transcription [74]. The molecular mechanisms of lncRNA action is mostly not well understood. In the nucleus, they often regulate transcription in cis or in trans by forming structural and regulatory complexes with specific RNA-binding proteins, as exemplified by XIST-mediated gene/ chromosome silencing. The Xist lncRNA is one of the best-characterized lncRNAs and is responsible for the X-chromosome inactivation (XCI) in female cells. To achieve this major task, Xist functions as a scaffold and recruits repressive complexes such as the Polycomb Repressive Complex 2 (PRC2), to establish and maintain an inactive state. During this process, Xist was found to interact with many proteins involved in chromatin modification and RNA remodeling pathways. It was also found that inhibition of proteins of the Xist interactome leads to disturbance of the XCI, further emphasizing the essential functional coordination between RNAs and proteins [75], [76], [77]. More recent research has shown that some lncRNAs form triplex structures with DNA or even RNA–DNA hybrids to, for example, recruit chromatin modifiers to specific loci in the genome [10]. In the cytoplasm, lncRNAs seem to primarily control mRNA stability and translation like lincRNA-p21, or are involved in regulating protein stability like the lncRNA HOTAIR which was found to interact with two ubiquitin ligases. However, the functional roles and interactions of cytoplasmic lncRNAs are less understood compared to nuclear lncRNAs [69], [71].
One of the hallmark features of lncRNAs is their tissue or cell lineage-specific expression, often at very low levels. This combination poses a major challenge for studying their localization, function, as well as interactions with proteins or nucleic acids [13], [78]. Unlike mRNAs, point mutations/ deletions may have no phenotypic consequence as the lncRNA may still be able to fold properly, the effect may also be context-dependent and not evident under homeostatic conditions [10]. The importance of better understanding the roles of lncRNAs is obvious when looking at the aberrant lncRNA expression in various cancers. They contribute to tumor initiation and progression, underscoring their clinical potential as biomarkers and therapeutic targets [79].
Despite the latest advances, the relationship between lncRNA sequence, structure, and function remains poorly understood. Unlike mRNAs, which encode proteins with identifiable domains, the functional elements or motifs of lncRNAs are not well characterized, limiting our ability to infer their functions or group them based on shared sequence features. As many aspects of their function remain to be elucidated, identifying binding partners and understanding the regulatory networks in which a lncRNA is involved adds valuable insights into the biological role and function. While lncRNAs hold immense biological and clinical significance, their full mechanistic understanding remains an open challenge and more research in necessary to elucidate the molecular functions and interactions of lncRNAs.

8.1.1.1. The lncRNA NORAD

NORAD is a highly abundant, ubiquitously expressed lncRNA. In the human genome, NORAD is an unspliced lincRNA of 5.3 kb, which does not overlap with other genes (see Figure 4). It is transcribed from an RNA polymerase II transcription unit, characterized by a single strong promoter and a canonical polyadenylation signal at its 3' end. The expression is induced following DNA damage, earning it the name "non-coding RNA activated by DNA damage" (NORAD) [80], [81]. Despite its induction upon DNA damage, NORAD predominantly localizes to the cytoplasm and does not redistribute to the nucleus under such conditions [82].

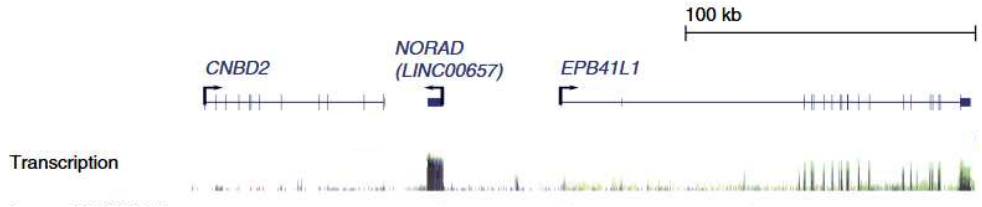


Figure 4.: The genomic position of the long non-coding RNA NORAD. NORAD is a lincRNA transcribed from chromosome 20. The gene has a single strong promoter and is unspliced. Edited from [81].

Sequence analysis revealed remarkable conservation across mammalian species, particularly in the central ∼3.5 kb region of NORAD, which contains 14 repeating units of ∼300 nucleotides each. These units are separated by spacer sequences of similar lengths [83].
Functionally, NORAD acts as a molecular decoy for PUMILIO proteins. They normally bind the 3' untranslated regions (UTRs) of target mRNAs and promote the deadenylation and decapping of the mRNAs, leading to their degradation and repression. There are two human and mouse PUF proteins, PUMILIO1 (PUM1) and PUMILIO2 (PUM2) (exhibiting 91% similarity in their RNA-binding domains). They bind specific target sequences known as Pumilio response elements (PREs), which consist of an eight-nucleotide motif (UGUANAUA or UGUANAUN). NORAD contains multiple conserved PREs distributed throughout its sequence, allowing it to effectively sequester Pumilio proteins [81]. More recent research found that upon DNA damage-induced NORAD up-regulation, NORAD, as well as Pumilio seem to form cytoplasmic granules. These RNP granules are proposed to form via liquid-liquid phase separation (LLPS), with NORAD serving as a scaffold and promoting interactions between intrinsically disordered regions (IDRs) of PUM proteins.
Loss of NORAD results in PUMILIO hyperactivity and drastic downregulation of target mRNAs that are critical for genome stability, mitosis, DNA repair and DNA replication. In consequence, NORAD-depleted human cells exhibit chromosomal instability and aneuploidy [80], [84].
Detailed structural analyses of NORAD revealed conserved secondary structures within the NORAD repeating units, located immediately downstream of PREs. These structures appear to mediate specific protein interactions. For instance, the RNA-binding protein SAM68 associates with A/G-rich regions, and its integrity is required for NORAD-mediated repression of PUMILIO activity [85]. Although the main interactors of NORAD seem to be the PUM proteins, proteomic high-throughput methods searching for NORAD-interacting partners showed

enrichment of proteins involved in DNA damage response, mitotic cell cycle regulation, and the minichromosome maintenance (MCM) complex [86] which further completes the picture of NORAD as "defender of the genome" [87]. To gain a better understanding of these interactions and the additional roles of NORAD in this context would be of great interest.
Intriguingly, NORAD-related pseudogenes are found throughout mammalian genomes. In humans, four pseudogenes share >90% homology with NORAD over >4 kb, though most show minimal transcriptional activity. A notable exception is HCG11, a lincRNA expressed at significantly lower levels than NORAD and observed in multiple tissues [80], [81].
Studies in mice further underscore the critical biological role of NORAD. While NORAD$^{-/-}$ mice are viable and fertile, by the age of six months they develop a progressive multi-system degenerative phenotype resembling premature aging. This phenotype is characterized by increased chromosomal instability, extensive aneuploidy, and mitochondrial dysfunction, highlighting the lncRNA's role in diverse cellular functions as well as organismal health and the necessity to understand it better [88].

8.1.1.2. The lncRNA OIP5-AS1

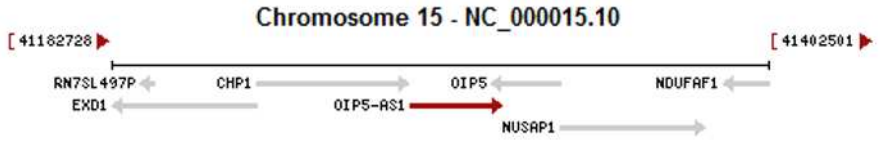


Figure 5.: The position in the genome of long non-coding RNA OIP5-AS1. The long non-coding ribonucleic acid OIP5-AS1 is transcribed from chromosome 15 in antisense direction to the gene that encodes Opa-interacting protein 5 (OIP5). From [89].

While NORAD was discovered by a screen for induced genes after DNA damage, the lncRNA OIP5-AS1 was identified by Ulitsky et al. (2011), who sifted through chromatin marks, poly(A)-site mapping and RNA-Seq data of zebrafish to map lincRNAs. In a second step, the authors searched for putative mammalian orthologs among the identified transcripts. They found one that possesses a highly conserved 67-nt sequence shared between 45 vertebrate genomes, harboring a miRNA binding site for miR-7 and named this lincRNA cyrano in zebrafish. In mammals, it is called OIP5-AS1 due to its transcription in the opposite direction to its nearest neighbor, the oncogene OIP5 (see Figure 5) [90]. Further analysis revealed 37 more motifs conserved throughout mammals [72]. The full-length OIP5-AS1 transcript is 15.6 kb long, with a mature main isoform of 8.8 kb consisting of three short and one long exon. It has at least six splicing variants, some of which have distinct subcellular localizations and functions [91].
Cyrano is highly expressed in the nervous system and notochord of zebrafish embryos. Knockdown experiments in zebrafish resulted in developmental defects, demonstrating its role in embryogenesis, with mouse or human orthologs rescuing these defects [90]. However, subsequent studies found cyrano to be dispensable for embryogenesis and viability. Contrary to

the earlier zebrafish studies, phenotypes were deemed non-specific, and cyrano or its conserved miR-7-binding site was shown to be not necessary for embryonic development or pluripotency [92].
Cyrano was also shown to be expressed in mouse embryonic stem cells (ESCs). While earlier studies suggested cyrano supports ESC self-renewal and colony maintenance by regulating cell adhesion [93], more recent findings using different genetic models revealed that it is non-essential in murine and human induced pluripotent stem cells (iPSCs) and ESCs. RNA sequencing showed minimal global transcriptome changes in cyrano-depleted cells, which retained their potential to differentiate into the three germ layers. These findings challenge previous conclusions regarding its role in ESC pluripotency [94].
OIP5-AS1 localizes predominantly to the cytoplasm [95] with several interactors postulated. It was found to bind the RNA-binding protein HuR, a member of the ELAV (embryonic lethal abnormal vision) RBP family that regulates cell proliferation. This interaction stabilizes OIP5-AS1, but competition from miRNA miR-424 can release HuR from OIP5-AS1, altering HuR's availability for target mRNAs [96]. Another interactor of OIP5-AS1 in HeLa cells was discovered to be GAK mRNA, which encodes a cyclin G-associated kinase important for mitotic progression. Binding reduced GAK mRNA stability and GAK protein abundance. This regulation helps to maintain proper chromosome segregation during mitosis by maintaining low GAK levels [97]. A major issue with the two previous studies is their claim that the mature OIP5-AS1 transcript is 1.9 kb long, whereas newer annotations suggest that this isoform does not exist.
Work from Naemura et al. (2018) found a role of OIP5-AS1 in promoting cell proliferation. In their experiments, silencing OIP5-AS1 caused G2/M phase cell-cycle arrest but did not induce apoptosis. Importantly, it was emphasized that each splice variant of OIP5-AS1 may have distinct target genes or functions [98]. The same point was stressed by more recent results, finding that one OIP5-AS1 splicing variant (ENST00000501665.2) was retained in the nucleus of HEK293 cells, where it associated with chromatin and influenced the expression of its sense partner, OIP5 [91].
While OIP5-AS1 and cyrano have been linked to multiple cellular and developmental processes, the precise role remains under investigation, with many proposed functions and mechanisms yet to be confirmed. Additionally, OIP5-AS1 has been implicated in miRNA binding and regulation in diverse cancer types by various publications, functioning either as an oncogene or tumor suppressor, though these roles remain largely unconfirmed (reviewed in [99]). This once again stresses the functional relevance of OIP5-AS1, but more research and a better understanding of its interactors is needed to advance the current, partially contradicting, level of knowledge.

8.1.2. RNA tagging and the CRISPR/Cas9 system

RNA aptamers as tags

Tagging RNA has become a useful tool over the last few years, especially since the discovery of CRISPR/Cas9. This method provides a versatile platform for studying RNA biology. Genomic incorporation of aptamer-coding sequences into a gene is particularly useful for dissecting the

functional roles of RNAs, such as determining how lncRNAs influence gene expression, chromatin organization, or signal transduction. The integrated aptamers are recognized by specific RNA-binding proteins or protein domains, which are usually fused to tags like GFP or affinity tags. Tagging an RNA of interest then enables e.g. the direct observation of RNA dynamics through live-cell imaging to study RNA localization, transport, and decay in real-time. It also facilitates pull-down assays, for isolating a specific RNA to identify and analyze their associated proteins or other biomolecules, shedding light on how specific RBPs regulate RNA stability, processing, and function within ribonucleoprotein complexes.

Several aptamer systems are currently in use, including the related MS2 and R17 stem-loops and the BoxB hairpin, which are recognized by the MS2 coat protein (MCP), R17 coat protein, and λN peptide, respectively. They all originate from different bacteriophages [100]. A further variant is the PP7 stem-loop and the corresponding PP7 coat protein from bacteriophage PP7. Although all RNA secondary structures can be generally characterized as stem-loops, the size, shape and sequence of the loop account for the differences in binding specificity [101].
Among the RNA tagging systems, the MS2-MCP pair is most used due to its good balance of specificity and affinity. Primarily, the MS2 hairpin loop (MS2 loop) [102] bound to its complementary protein MCP [103] have been successfully applied as an imaging tool [104]. To visualize RNA, the aptamer sequence is usually fused to the 3'UTR of a target mRNA and then bound by the exogenously expressed MCP fused to a fluorescent protein. Thereby, the entire RNA becomes visible in live-cell imaging microscopy [105]. After the initial experiment, the first optimization was the increase of MS2 loop repeats to enhance signal intensity and raise the signal-to-noise ratio [106]. However, concerns arose about potential impacts on RNA stability or functional disturbances when multiple aptamer repeats are used [59], [107]. A study in 2015 showed that 24 MS2 stem-loops bound by MCP are sufficient to block Xrn1 and prevent RNA degradation in yeast [59]. More recent improvements of the system promise a visualization of the complete life cycle of even unstable mRNAs in yeast. Therefore, a single nucleotide in the MS2 loop sequence was changed which reduced the MCP affinity so it no longer affected the decay. They also showed that reducing the number of stem loops from 24 to 12 helped to increase the degradation of the RNA [61]. As the MS2-MCP system allows co-expression of the tagged RNA and the MCP *in cellulo*, which facilitates the assembly of ribonucleoprotein (RNP) complexes under physiological conditions prior to cell lysis and purification, this system was also used for pull-down approaches. The high affinity between the MS2 loops and the MCP protein permits the use of stringent washing conditions, reducing background contamination [60], [108], [109]. Studies using MS2 aptamers for affinity purification also demonstrated improved pull-down efficiency with up to ten repeats, but showed no further increase in the purity of the final product between 10x and 24x aptamers [110].
The utility of RNA tagging extends beyond imaging and pull-down assays to more sophisticated methods such as RaPID and RNA-BioID. Here, aptamers provide the essential link for tethering biotin ligases or peroxidases to RNA *in vivo*. This approach allows the biotinylation of RNA-proximal proteins, enabling the identification of RNA-protein interactomes under near-native conditions (for details, see section 4.3). In RNA-BioID, the MS2-MCP aptamer system is key to anchoring the enzymatic tool to the RNA of interest.

Using CRISPR/Cas9 to tag RNAs

A key challenge in RNA tagging, for example with MS2 aptamers, is the integration of the aptamer sequence into the genome of target cells, particularly within endogenous genes encoding lncRNAs. This requires precise editing of genomic loci, a task that has been revolutionized by the CRISPR/Cas9 system, which allows systematic genome engineering in eukaryotic cells with high specificity. This technology is now widely used for genome modification due to its ease of design, high specificity, and ability to target a variety of cell types and organisms.
The clustered regularly interspaced short palindromic repeats system, or in short, the CRISPR system originates from the adaptive immune defense of bacteria and archaea. The second important player within this immune system is Cas9, a programmable RNA-guided endonuclease. It uses short RNA sequences as guide to locate the target DNA sequence with the help of Watson-Crick base pairing. After finding the target DNA locus, Cas9 introduces a DNA double-strand break (DSB) [111].
The most commonly employed CRISPR-Cas system is derived from *Streptococcus pyogenes* (SpCas9). It requires a 20-nucleotide guide RNA (gRNA) sequence, which hybridizes with the target DNA. As a prerequisite to cutting the DNA, it must immediately be preceded by a 5'-NGG protospacer adjacent motif (PAM). Upon binding to the target DNA, SpCas9 introduces a DSB three base pairs upstream of the PAM (see upper part of Figure 6). The PAM requirement is not a severe limitation, as in the human genome such sites can be found approximately every 8–12 bp [60], [112].
Efficient tagging also requires a repair template to direct modifications at the DSB site. Following cleavage, the cell's DNA repair machinery engages one of two major pathways to restore genomic integrity: non-homologous end-joining (NHEJ) or homology-directed repair (HDR) (see lower part of Figure 6). A third option is microhomology-mediated end-joining (MMEJ). Each pathway has distinct features and applications in genome editing.
NHEJ is the predominant repair mechanism in most cell types and happens also in the absence of a repair template. It is an error-prone process that directly ligates the broken DNA ends, often resulting in insertions or deletions (indels) at the break site [111], [112]. In contrast, HDR is a high-fidelity repair pathway that needs an exogenous repair template with homology arms flanking the cutting site to guide precise DNA repair. However, HDR occurs at lower frequency, as it is active predominantly in the late S/ G2 phases of the cell cycle. Its efficiency varies widely depending on factors such as cell type, genomic locus, and the design of the repair template. Repair templates can be in the form of double-stranded DNA constructs or single-stranded sequences, with longer homology arms generally improving integration efficiency [112], [113].
MMEJ is an alternative repair pathway that uses microhomologous arms (ranging from 5–100 bp) flanking the DSB for error-prone repair. Unlike HDR, MMEJ is active during the G1 and early S phases of the cell cycle, broadening its utility in non-dividing or slowly dividing cells with low HDR activity. MMEJ has been adapted for efficient gene knock-ins, both in cultured cells and animals. Studies have demonstrated that PCR-generated repair templates with short homology arms can effectively mediate targeted integrations, with increasing homology arm length improving efficiency [114], [115].

While the CRISPR/Cas9 system has transformed genomic editing, its application to long noncoding RNAs poses huge challenges due to the complex genomic context of many lncRNA loci.

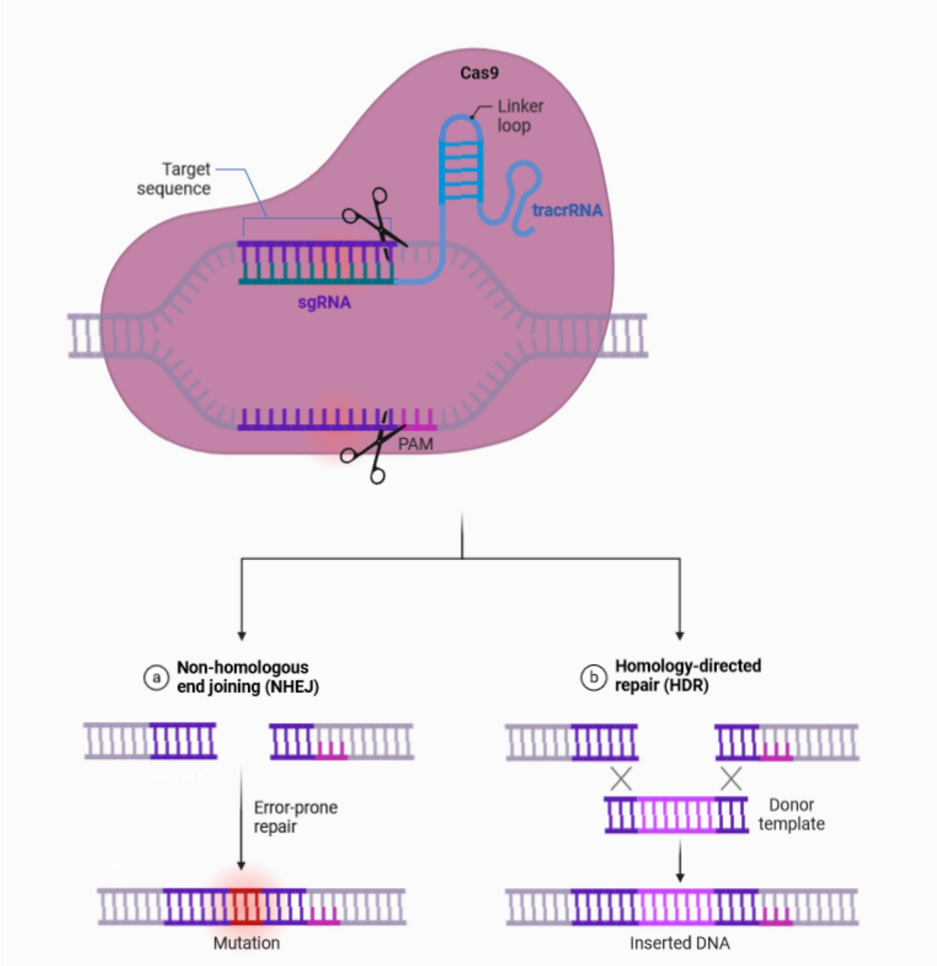


Figure 6.: Scheme for CRISPR/Cas9-mediated targeted insertion. Schematic Cas9 enzyme, bound to its single guide RNA (sgRNA). After cleavage, two main repair pathways seal the DNA break: (a) non-homologous end-joining (NHEJ) or (b) homology-directed repair (HDR).

Before CRISPR editing can be performed, the target genomic locus must be carefully analyzed. Many lncRNAs overlap with protein-coding genes or are derived from bidirectional promoters (see Figure 3). Editing of these lncRNAs by CRISPR may disrupt the function of neighboring or overlapping genes. Therefore, it is essential to ensure that targeted modifications alter only the intended lncRNA gene without disrupting or influencing the nearby genes or regulatory regions [13].

A systematic analysis of lncRNA loci by Goyal et al. (2017) revealed that only a minority of lncRNA genome loci are safely amenable to CRISPR editing. In their genome-wide study of 15,929 loci, only 38% of targets were susceptible without risking deregulation of neighboring genes [68]. To address these challenges, the authors formulated a set of rules for lncRNA 'CRISPRability', categorizing lncRNAs based on their genomic context. If lncRNAs, no matter whether they are antisense or intergenic, are transcribed from bidirectional promoters, defined by the presence of another promoter within 2000 base pairs upstream or downstream of the

lncRNA start site, or the start site of the lncRNA was located closer than 2000 bp to the start of the neighboring gene, they were considered 'non-CRISPRable'. LncRNAs were also deemed 'non-CRISPRable' when transcribed from internal promoters situated within the length of another gene. These rules highlight the limitations posed by the genomic organization of lncRNAs on the safe and specific application of CRISPR/Cas9. Additionally, many of the regulatory elements within the genome remain poorly characterized and annotated, further complicating the usage of CRISPR in these areas [68].

8.2. Results

The primary research objective of this project was to identify RNA-binding proteins (RBPs) that interact with lncRNAs, focusing on NORAD and OIP5-AS1 as candidates (introduced above in section 8.1.1.1 and 8.1.1.2), by establishing a proximity labeling assay for lncRNAs. This class of RNA is generally little understood, and identifying binding partners adds valuable insights into their function. While for NORAD the main interactors are identified (PUM1 and PUM2) and should serve as valuable positive control, the lncRNA OIP5-AS1 is more cryptic and the assay should contribute to elucidating its cellular function.
To achieve this, the RNA-BioID assay, originally introduced by Mukherjee et al. (2019), was adapted and advanced. In the original publication, mouse embryonic fibroblast (MEF) cells tagged with 24xMS2 loops in the 3' UTR of the β-actin gene were used. This cell line was obtained from genetically modified mice generated and published by Lionnet et al. (2011) [58]. To establish the RNA-BioID assay, Mukherjee et al. stably transfected these cells with the 2xMCP-eGFP-BirA* fusion protein, which consists of three functional components: 2xMCP to bind and tether the construct to the MS2 loops of the RNA, GFP for visualization, and BirA*, a biotin ligase with slow labeling kinetics (∼16 h). Additionally, the fusion protein includes a nuclear localization signal (NLS), ensuring that free biotin ligase is concentrated in the nucleus. This design helps to minimize cytoplasmic background biotinylation.
Building on this groundwork, the experimental plan of this project involved two key steps: (1) stable integration of either 2xMCP-eGFP-TurboID or 2xMCP-eGFP-microID into cells, followed by evaluation to determine the more suitable option for biotin labeling (detailed in section 8.2.1); and (2) tagging the endogenous loci of lncRNA OIP5-AS1 and NORAD with 24xMS2 loops using CRISPR-Cas9 (see section 8.2.2). After this setup, optimization of biotinylation and streptavidin pull-down will be followed by mass spectrometry analysis of the enriched proteins.
Together, these steps aimed to establish an lncRNA-BioID assay and enable the identification of the interactomes of these lncRNA candidates.

8.2.1. Stable introduction of a biotin ligase

Both TurboID and microID provide valuable improvements over traditional PL enzymes, with TurboID excelling in labeling speed and microID offering a highly compact alternative with strong activity. Each enzyme comes with advantages and potential drawbacks, making their selection context-dependent based on experimental needs.

The first step to achieve stable integration of either 2xMCP-eGFP-TurboID or 2xMCP-eGFP-microID into cells, was cloning the plasmid p2263 containing 2xMCP-eGFP-TurboID. To validate the functionality of the newly constructed plasmid, a series of transient transfection experiments were conducted in HeLa cells (data not shown) prior to proceeding with stable cell line generation. These tests aimed to assess both the full-length expression of the constructs

and their biotinylation activity.
For the TurboID construct (p2263, CMV:NLS-HA-2xMCP-eGFP-TurboID), transient transfections of HeLa cells were performed using a DNA-to-FuGENE ratio of 1:3 (w/v). Expression was controlled using live-cell microscopy, revealing robust nuclear localization with some minor cytoplasmic GFP signal. A biotinylation assay followed by Western blot analysis confirmed successful full-length protein expression. The biotinylation activity was tested by supplementing cells with either 50 µM or 100 µM biotin. No significant differences in biotinylation efficiency were observed between the two biotin concentrations. Time-course experiments with biotin incubation for 1, 3, and 6 hours showed a strong biotinylation signal after just 1 hour, with only a mild increase after longer incubation times. However, control transfections without additional biotin exhibited some background biotinylation.
The microID construct with plasmid number p2250 (CMV:NLS-HA-2xMCP-eGFP-microID) was similarly tested. Biotinylation activity was assessed using again 50 µM and 100 µM biotin, with no difference in the outcome between the two concentrations. Time-course experiments with incubation times of 15, 60, and 180 minutes showed weak biotinylation signals at the shortest time point. By 1 hour, biotinylation signals were clearly detectable, with minimal background signal in transfected cells without additional biotin compared to the TurboID sample. After 3 hours, biotinylation signals were strong, confirming the functionality of the construct.
These results demonstrate that both TurboID and microID constructs are functional in transient transfection settings. Based on these findings, the next step is the stable integration of the constructs into target cells for further experiments.

To establish a robust system for studying lncRNA interactions, a stable HeLa cell line expressing MCP-eGFP-TurboID or MCP-eGFP-microID fusion constructs was generated. HeLa cells were chosen for their ease of culture, robust growth, and the expression of the lncRNA candidates under investigation. Specifically, the HeLa-EM2-11ht cell line was used, which was developed by Weidenfeld et al. (2009) and is engineered to express the tetracycline-controlled transcription activator rtTA2-M2 uniformly across the cell population. This enables the expression of introduced constructs under a tetracycline/ doxycycline inducible promoter, a useful feature to timely control the transcription and translation of exogenous proteins. Additionally, this cell line contains a genomic locus flanked by heterospecific wild-type (F) and mutated (F3) FRT sites, allowing for directed cassette exchange via Flp recombinase-mediated cassette exchange (RMCE) [116]. This feature facilitates the precise integration of constructs into a defined genomic locus (shown in Figure 7(A)).
To integrate the desired constructs into HeLa-EM2-11ht cells, 2xMCP-eGFP-microID and 2xMCP-eGFP-TurboID were cloned into the pSF3 plasmid backbone containing F and F3 sites (yielding constructs p2279 and p2280, respectively). These plasmids were co-transfected with a Flp recombinase expression vector (p2042) into HeLa-EM2-11ht cells. The Flp recombinase facilitated recombination and cassette exchange, resulting in the stable integration of the constructs. The transfected cells were grown under antibiotic selection pressure as described by Weidenfeld et al. (2009) and monoclonal cell lines were picked. Following clonal expansion, the full-length expression and functionality of the integrated constructs were analyzed using live-cell microscopy, flow cytometry, and biotinylation assays. Several monoclonal cell lines

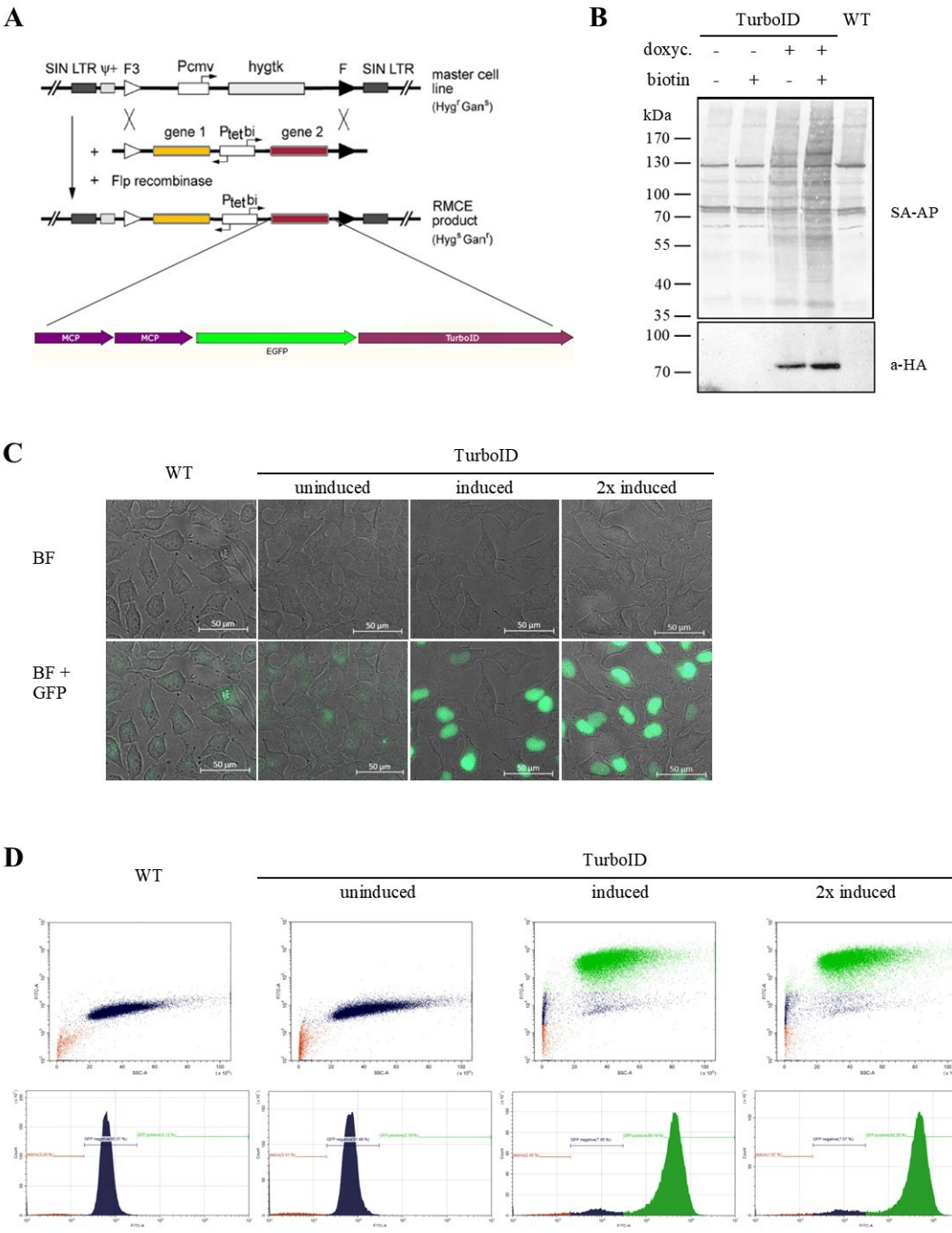


Figure 7.: Stable integration of the 2xMCP-eGFP-TurboID fusion construct into HeLa-EM2-11ht cells. (A) The HeLa cell line produced by Weidenfeld et al., 2009 called HeLa-EM2-11ht contains a retargetable genomic locus and is shown here as master cell line [116]. Integration of constructs flanked by a F3 and F site can be achieved by co-transfecting a Flp recombinase, resulting in doxycycline inducible expression of, in this case, the 2xMCP-eGFP-TurboID fusion construct. RMCE = recombinase-mediated cassette exchange. **(B, C, D)** To control the successful introduction of the fusion construct 2xMCP-eGFP-TurboID, it's doxycycline inducible expression and correct function, several assay were performed.

Figure 7.: (Continued from previous page.) **(B)** A biotinylation assay to control for the function of TurboID as biotin ligase was performed and analyzed by Western blot. The lower blot shows expression of the fusion construct upon addition of doxycycline, with a band at the expected size of 88 kDa. The upper blot shows a clear biotinylation signal after the cells have been incubated with exogenous biotin for 10 min. However, also without the addition of exogenous biotin, a biotinylation of proteins can be seen when TurboID is expressed. **(C)** Live cell microscopy to control for eGFP expression and function after the induction with doxycycline was done. While the uninduced cells show no GFP signal, induced cells or 2x induced cells supplemented with the double amount of doxycycline show clear nuclear GFP signal (green). The white bar represent 50 µm. **(D)** For quantification of induction, the GFP expression of cells after the addition of doxycycline was measured by flow cytometry. Again, wild type and uninuced cells show no GFP expression, while induced cells or 2x induced cells supplemented with the double amount of doxycycline show GFP expression in $\sim$ 90% of cells. Scatter blot shown in the upper row, cell count below. The red cell population represents debris (2-5%), the purple population in not expression GFP, while the GFP expressing cells are in green. WT = HeLa-EM2-11ht cells.

were tested and compared for each construct. As the performance was very similar between the individual clones, only results of one clonal cell line are depicted here.

To confirm the functionality of TurboID as biotin ligase, a biotinylation assay was performed (shown in Figure 7(B)). Cells were seeded and 200 ng/ml doxycycline were added in the evening to induce expression of the fusion construct, followed by incubation with 50 µM biotin for 10 minutes the next day. Western blot analysis demonstrated clear induction of the fusion protein upon doxycycline treatment, with a band observed at the expected size of 88 kDa when probed with an anti-HA antibody ((B), lower blot). Biotinylation activity was confirmed by probing with a streptavidin-alkaline phosphatase conjugate where strong biotinylation signals were detected following biotin treatment ((B), upper blot). In the absence of exogenous biotin, background biotinylation was observed when TurboID was expressed. This is probably due to the strong affinity of TurboID for biotin and the presence of low concentrations of biotin in the medium as supplement. The induction and localization of the fusion construct was assessed using live-cell microscopy (see Figure 7(C)). Following overnight doxycycline treatment, GFP fluorescence was detected predominantly in the nucleus of induced cells, while cells treated with double the doxycycline dose showed no difference. Uninduced and wild type cells displayed no detectable GFP signal. For quantification, the GFP expression of cells was measured by flow cytometry after an over night induction with doxycycline (Figure 7(D)). The analysis confirmed the microscopy data. Wild type and uninuced cells show no GFP expression, while induced cells or 2x induced cells supplemented with the double amount of doxycycline show GFP expression in $\sim$90% of cells. The scatter plot ((D), upper row) and histogram ((D), lower row) show the GFP-expressing population in green, with debris (2–5%) and non-expressing cells represented in red and purple, respectively.
These results validate the successful integration and functionality of the 2xMCP-eGFP-TurboID construct in HeLa-EM2-11ht cells. The construct demonstrated robust induction, proper nuclear localization, and biotinylation activity. However, the high background biotinylation observed with TurboID needs further optimization to minimize unintended labeling.

The same analyses as described above were performed with the monoclonal microID cell lines to validate the successful integration and functionality of the 2xMCP-eGFP-microID fusion

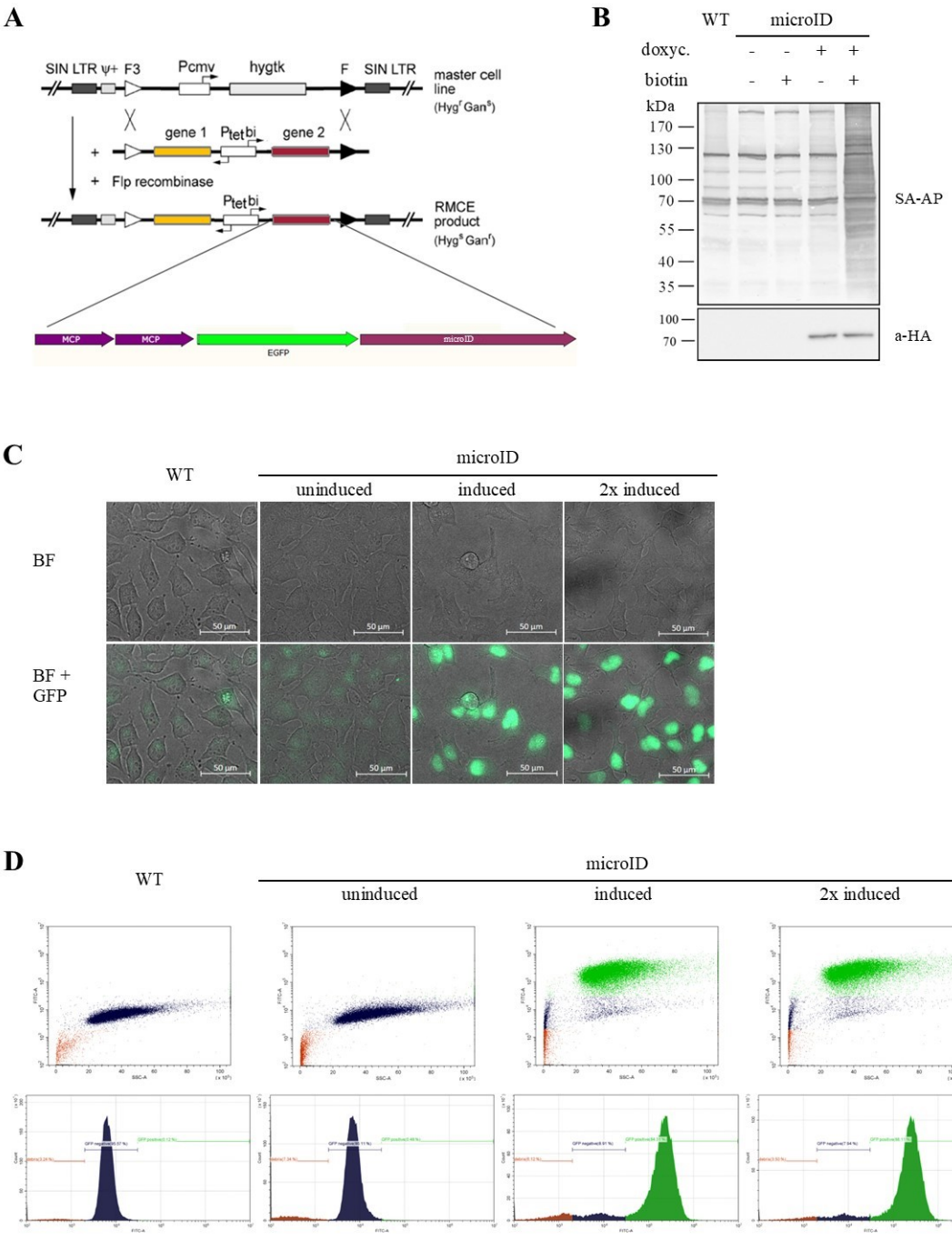


Figure 8.: Stable integration of the 2xMCP-eGFP-microID fusion construct into HeLa-EM2-11ht cells. (A) The HeLa cell line produced by Weidenfeld et al., 2009 called HeLa-EM2-11ht contains a retargetable genomic locus and is shown here as master cell line [116]. Integration of constructs flanked by a F3 and F site can be done by co-transfecting a Flp recombinase, resulting in doxycycline inducible expression of, in this case, the 2xMCP-eGFP-microID fusion construct. RMCE = recombinase-mediated cassette exchange. **(B, C, D)** To control the successful introduction of the fusion construct 2xMCP-eGFP-microID and it's doxycycline inducible expression and correct function, several assay were performed.

Figure 8.: (Continued from previous page.) **(B)** A biotinylation assay to control for the function of microID as biotin ligase was performed and analyzed by Western blot. The lower blot shows expression of the fusion construct upon addition of doxycycline, with a band at the expected size of 76 kDa. The upper blot shows a clear biotinylation signal when the fusion construct is expressed and after the cells have been incubated with exogenous biotin for 2 h (last lane). **(C)** Live cell microscopy to control for eGFP expression and function after the induction with doxycycline was done. While the uninduced cells show no GFP signal, induced cells or 2x induced cells supplemented with the double amount of doxycycline show clear nuclear GFP signal (green). The white bar represent 50 µm. **(D)** For quantification of induction, the GFP expression of cells after the addition of doxycycline was measured by flow cytometry. Again, wild type and uninuced cells show no GFP expression, while induced cells or 2x induced cells supplemented with the double amount of doxycycline show GFP expression in $\sim$ 86% of cells. Scatter blot shown in the upper row, cell count below. The red cell population represents debris (2-7%), the purple population in not expressing GFP, while the GFP expressing cells are in green. WT = HeLa-EM2-11ht cells.

construct. Shown are the results of one monoclonal cell line.
A biotinylation assay confirmed the activity of microID as a biotin ligase (see Figure 8(B). Cells were treated with 200 ng/mL doxycycline overnight, followed by incubation with 50 µM biotin for 2 hours. Western blot analysis showed a clear band at the expected size of 76 kDa upon doxycycline induction, as detected with an anti-HA antibody (see Figure 8(B), lower blot). Biotinylation activity was confirmed by streptavidin-AP probing, with a strong biotinylation signal observed in doxycycline-induced cells treated with biotin (upper blot).
Live-cell microscopy demonstrated nuclear GFP localization in cells treated with doxycycline, while uninduced cells showed no GFP signal (Figure 8(C)). Increased doxycycline concentrations did not result in visibly higher GFP intensity.
Flow cytometry quantification further validated these observations. No GFP expression was detected in wild-type or uninduced cells, while $\sim$ 86% of doxycycline-treated cells exhibited GFP fluorescence. Scatter plots (Figure 8(D), upper row) and histogram data (lower row) show the GFP-positive population in green, with debris (red, 2–7%) as minor population and non-expressing cells (purple). The results aligned with microscopy data, confirming efficient inducible expression of the fusion construct.
These results confirmed the successful generation of a functional microID cell line with robust induction, correct nuclear localization, and effective biotinylation activity.

In direct comparison, TurboID shows higher biotinylation activity as microID and faster labeling kinetics. TurboID achieves strong biotinylation signals within a 10-minute labeling time, while a similar labeling intensity is reached after 2 hours with microID. This shorter labeling time is biologically more interesting as it enables to capture a snapshot of interactions. However, TurboID has such a high activity, which is also described in literature [41], that even without adding exogenous biotin to cells, biotinylation can be observed (Figure 7(B), lane 3). Therefore, the expression level of the TurboID construct was optimized with the help of a time-course induction experiment.
TurboID expression was induced with doxycycline at five different time points, and all samples were supplemented with biotin for 10 minutes prior to lysis. Western blot analysis was conducted including three control samples and WT lysate as negative control with anti-HA antibody and SA-AP, the blots are shown in Figure 9. The HA blot confirmed an increasing

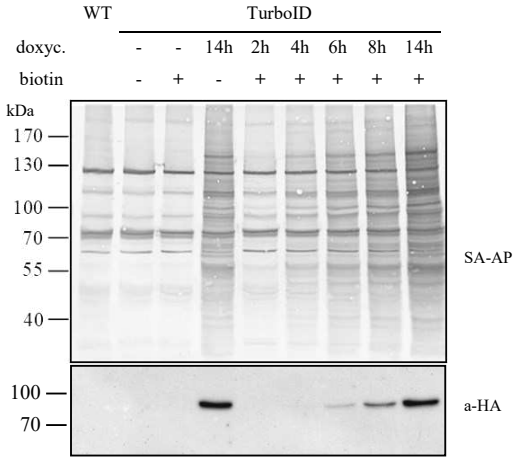


Figure 9.: Biotinylation assay to investigate the optimal induction time of TurboID. The expression of TurboID was induced at different time points with doxycycline. Biotin was added to all samples for 10 min. The result was analyzed by Western blot. The lower blot shows expression of the fusion construct, with the band getting more prominent with longer induction time. The upper blot shows the respective biotinylation pattern getting stronger when the fusion construct is expressed for a longer time span. WT = HeLa-EM2-11ht cells.

expression level of the fusion construct over time, with a faint band detectable after 4 hours of induction with longer blot exposure. The SA-AP blot showed corresponding biotinylation activity, with a signal emerging after 4 hours and reaching a robust level at 8 hours of induction. At shorter induction times (2 hours), biotinylation activity was indistinguishable from wild type or uninduced controls. Based on these results, an 8-hour induction period was selected for optimal TurboID expression and activity. Live-cell imaging confirmed these findings, showing that after 4 hours of doxycycline induction, some cells exhibited bright GFP signals while others displayed weaker fluorescence, reflecting heterogeneous expression (data not shown). These results established the optimal induction duration for TurboID expression, balancing efficient biotinylation activity with manageable background levels.

In conclusion, both constructs (2xMCP-eGFP-TurboID and -microID) were successfully stably integrated into HeLa-EM2-11ht cells. Both fusion proteins are expressed and functional and show a very good doxycycline inducibility, while the promoter is not "leaky" and does not show expression without doxycycline addition.

8.2.2. CRISPR/Cas9-mediated tagging of endogenous lncRNA genes

The second step in order to establish a proximity labeling assay for lncRNA is the precise tagging of lncRNAs with 24xMS2 loops. The CRISPR/Cas9 system was chosen to insert the MS2 loops into the endogenous gene. This elegant tagging directly at the endogenous locus provides a method with minimal interference in the genome. However, as described in subsection 8.1.2, lncRNA 'CRISPRability' is rare, so careful choice of lncRNA candidates depending on the genomic locus was given. OIP5-AS1 and NORAD lncRNAs seemed suitable candidates, as both are lincRNAs with no direct overlapping protein-coding gene.

CRISPR attempt based on non-homologous end joining

To achieve tagging of OIP5-AS1 and NORAD with 24xMS2 loops, a CRISPR/Cas9-mediated approach was applied to the HeLa-EM2-11ht cells established in section 8.2.1 above, which express the 2xMCP-eGFP-TurboID. As the first target, the lncRNA OIP5-AS1 (characterized in section 8.1.1.2) was chosen. The CRISPR approach selected follows the protocol described by Cheng and Qiu, published in 2017 [117]. Their method describes the targeted insertion of tags into genomic loci through the non-homologous end joining (NHEJ) pathway (see subsection 8.1.2 for elaboration) and was used by the authors for lncRNA tagging and expression manipulation in mammalian cells. Their strategy is based on the simultaneous introduction of two plasmids, the first one expresses a Cas9 enzyme from *S. pyogenes* (SpCas9) and two sgRNAs driven by a U6 promoter each (defined as Cas9-2sgRNA system). The second plasmid is called donor plasmid and contains the tag as well as a puromycin resistance gene under a CMV promoter, which serves as a selection marker for cells with successful targeted insertion. Once brought into cells, the Cas9-2sgRNA system expresses the Cas9 nuclease alongside the two guide RNAs, one targeting the genomic locus of the lncRNA candidate and the other targeting the donor plasmid. This leads to DNA double-strand breaks (DSBs) at both the genomic target site and within the donor plasmid, which linearizes the repair template. Upon linearization, the donor plasmid containing the 24xMS2 loops is integrated into the target locus via NHEJ, circumventing the need for homologous sequences. To further streamline the process, a Cre recombinase encoding plasmid is subsequently transfected to excise unnecessary parts of the integrated donor construct, ensuring only the minimal desired tag (24xMS2 loops) remains (see Figure 10 for visualization).

To implement this CRISPR/Cas9-based tagging strategy, plasmids were obtained from the authors of the publication [117] and modified before use. The plasmid p2281, which contains the Cas9-2sgRNA system, was used as starting point for generating the CRISPR construct. Five different guide RNAs targeting OIP5-AS1 were designed (listed in Table 13) and cloned into the BbsI restriction site of plasmid p2281 following the protocol described in 7.1.2. The five created constructs (p2404-p2408, listed in Table 9) are capable of expressing the Cas9 enzyme and two gRNAs with one targeting the donor plasmid repair template and the other different positions in the gene of the lncRNA.
To assess the *in cellulo* efficiency of the designed gRNAs, transient transfections were performed

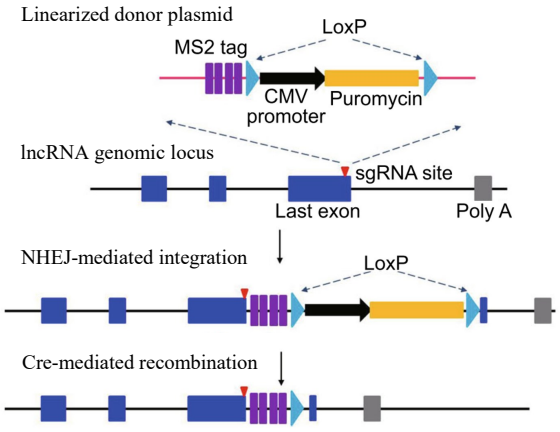


Figure 10.: Schematic workflow for tagging a lncRNA with MS2 loops by non-homologous end joining. This CRISPR approach was proposed by Cheng and Qiu in 2017 for introducing MS2 loops into the genome of lncRNAs. It includes the removal of the plasmid backbone containing the CMV promoter and puromycin resistance gene by Cre-mediated recombination. Figure adapted from [117].

with p2404 to p2408. Following transfection, genomic DNA was isolated, and the targeted region was amplified via PCR. The PCR products were sequenced, and the efficiency of each gRNA was evaluated based on the extent of genomic editing at the intended site. The most efficient gRNA (2a) was selected for further experiments. In parallel, a donor plasmid (p2313) was constructed to serve as repair template. This plasmid, CMV-puro-loxP-24xMS2 V5-loxP, was designed to contain the 24xMS2 loop sequence flanked by loxP sites.
Finally, both the CRISPR construct containing the selected gRNA (p2407) and the donor plasmid were co-transfected into the TurboID-expressing HeLa-EM2-11ht cell line (see Figure 13(A) for workflow). Following transfection, cells were subjected to antibiotic selection to enrich for successful integration events. However, despite following the published protocol of Cheng and Qiu (2017) [117], all cells died, indicating that the integration attempt was unsuccessful. This first approach, relying on plasmid-encoded gRNAs and a donor plasmid, proved to be too inefficient for achieving stable integration of the tag including the antibiotic resistance gene. Consequently, alternative strategies need to be explored to optimize the CRISPR-based tagging process.

To improve the integration efficiency of the MS2 loops into the OIP5-AS1 locus, an alternative CRISPR approach was employed using recombinant Cas9 protein pre-loaded with a chemically stabilized gRNA (see Figure 12(A)). This method aimed to enhance the CRISPR efficiency and quicken the process itself by co-delivering an already linearized repair plasmid (p2313) with the pre-made RNP complex of rCas9 and gRNA 2a. Multiple ratios (v/w in [µL]/[ng]) of the gRNA/Cas9 complex to the linear repair template were tested, including 0.5:400, 1:400, 1.5:400, 0.5:800, and 1:800. Following puromycin selection, eight colonies were obtained, with seven of the eight colonies arising from transfections using 800 ng of linear repair template. These colonies were picked and expanded to monoclonal cell lines for further analysis.

The validation of these clones required a series of thorough tests to ensure successful and precise knock-in of the MS2 loops. The following aspects should be tested: (1) Correctness of integration locus: Following isolation of genomic DNA (gDNA) the integration site can be analyzed via PCR and sequencing using out-in, in-out and primer sets on the repair template itself. (2) Expression level: The expression of the tagged RNA can be assessed by RT-qPCR, Northern blotting, and single-molecule FISH (smFISH). (3) Zygosity: The allelic state of the integration (heterozygous or homozygous) can be determined using Southern blotting and smFISH. (4) Off-target effects: Southern blotting and smFISH can be used to exclude off-target integration of the repair template. (5) Genomic damage: Ultimately, whole-genome sequencing can be performed to detect potential (off-target) genomic damage caused by CRISPR activity.

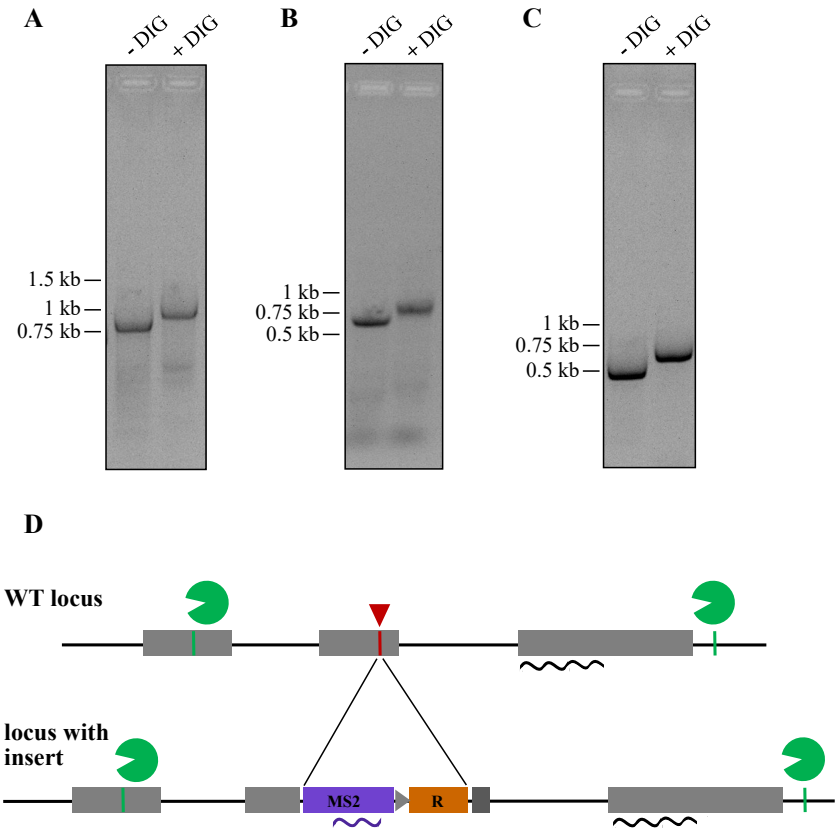


Figure 11.: Digoxygenin labeled probe synthesis for Southern blot analysis. To verify the Digoxygenin (DIG) labeling, labeled and unlabeled samples were run on an agarose gel. The successful integration of DIG-dUTP nucleotides into the probe was displayed by an upward size shift in the labeled sample. **(A)** Generation of DIG-labeled MS2 probe, expected size: 1080 bp. **(B)** Generation of DIG-labeled OIP5-AS1 probe, expected size: 788 bp. **(C)** Generation of DIG-labeled NORAD probe for later use, expected size: 704 bp. **(D)** General outline of a targeted genome locus before (WT) and after CRISPR/Cas9 mediated insertion of a cassette containing MS2 loops and a selection marker (orange). For analysis by Southern blot, WT and CRISPR clone gDNA is digested with a restriction enzyme (shown in green), separated by size with an agarose gel and detected with DIG-labeled probes annealing to the insert (in purple) and a genome specific site (in black). Due to the insertion, the fragment size is increased after successful tagging.

As a start, genotyping of clones was attempted. For gDNA extraction, the QuickExtract™ solution was used according to the manufacturer's instructions. PCR amplification targeting the MS2 loops as well as junctions between the loops and the genome failed to yield any detectable products. Solely primers amplifying the puromycin resistance cassette resulted in products with the expected size in the agarose gel.
Southern blotting was performed to confirm these findings and provide insights into zygosity and off-target effects (for schematic outline see Figure 11(D)). Two Digoxygenin-labeled probes were generated for that, using the PCR DIG Probe Synthesis Kit. PCRs with plasmid p2325 containing an OIP5-AS1 fragment and with the repair template, containing the MS2 V5 array were performed (see section 7.5.3 for protocol). The successful incorporation of DIG-dUTP was controlled by agarose gel electrophoresis and comparison to a negative control without DIG-dUTP (see Figure 11(A and B)).

Genomic DNA from the parental cell line served as a negative control, while the linearized or digested plasmids used as template to synthesize the probes served as positive control. Restriction digestion of 14 µg gDNA per sample with the BlpI enzyme and subsequent agarose gel electrophoresis of digested gDNA with controls is shown in an exemplary gel picture in Figure 12(B), upper image. Southern blotting and developing with the DIG-labeled probes showed one band each at the expected size for the positive control, an example blot probed with the MS2 DIG probe can be seen in Figure 12(B), lower picture. Probing of the membrane with OIP5-AS1 probe and stripping and re-probing with MS2 probe revealed no bands corresponding to the expected fragment size with MS2 loops (13,178 bp), neither with the OIP5-AS1 probe, nor with the MS2 probe (Figure 12(C)). Instead, only fragments corresponding to the unmodified locus (8,186 bp) were observed (left blot). Further evidence was obtained using smiFISH, which showed no detectable signal for the MS2 loops in any of the clones. These results indicate that the integration of the MS2 loops had failed.
A likely explanation for this failure is that exonucleases degraded the linearized repair template once it was within the cells. In consequence, only parts of the repair template were integrated, but upon adding antibiotic for selection only cells that contained the intact puromycin resistance cassette including the promoter were left. This hypothesis is supported by the PCR results, which managed to amplify the puromycin gene from gDNA of the clones, but nothing else. Southern blotting and smiFISH confirmed the absence of MS2 loops in the OIP5-AS1 locus (or somewhere else). Consequently, all clones generated from this CRISPR attempt were discarded. These findings necessitated the development of a new CRISPR strategy. The next approach, based on the protocol described by Chen et al. (2020) [118], was identified as a promising alternative (see page 93).

In parallel with the CRISPR/Cas9 tagging attempts for OIP5-AS1, the same strategy was applied to the second chosen lncRNA candidate, NORAD (characterized in section 8.1.1.1). Five gRNAs specific to the NORAD locus were designed (listed in Table 13) and cloned into the BbsI restriction site of plasmid p2281, following the protocol detailed in the Methods section 7.1.2, yielding plasmids p2409-p2413. To evaluate the *in cellulo* efficiency of the designed gRNAs, transient transfections were performed as described for OIP5-AS1. The same donor plasmid (p2313), containing CMV-puro-loxP-24xMS2 V5-loxP, was used as a repair template.

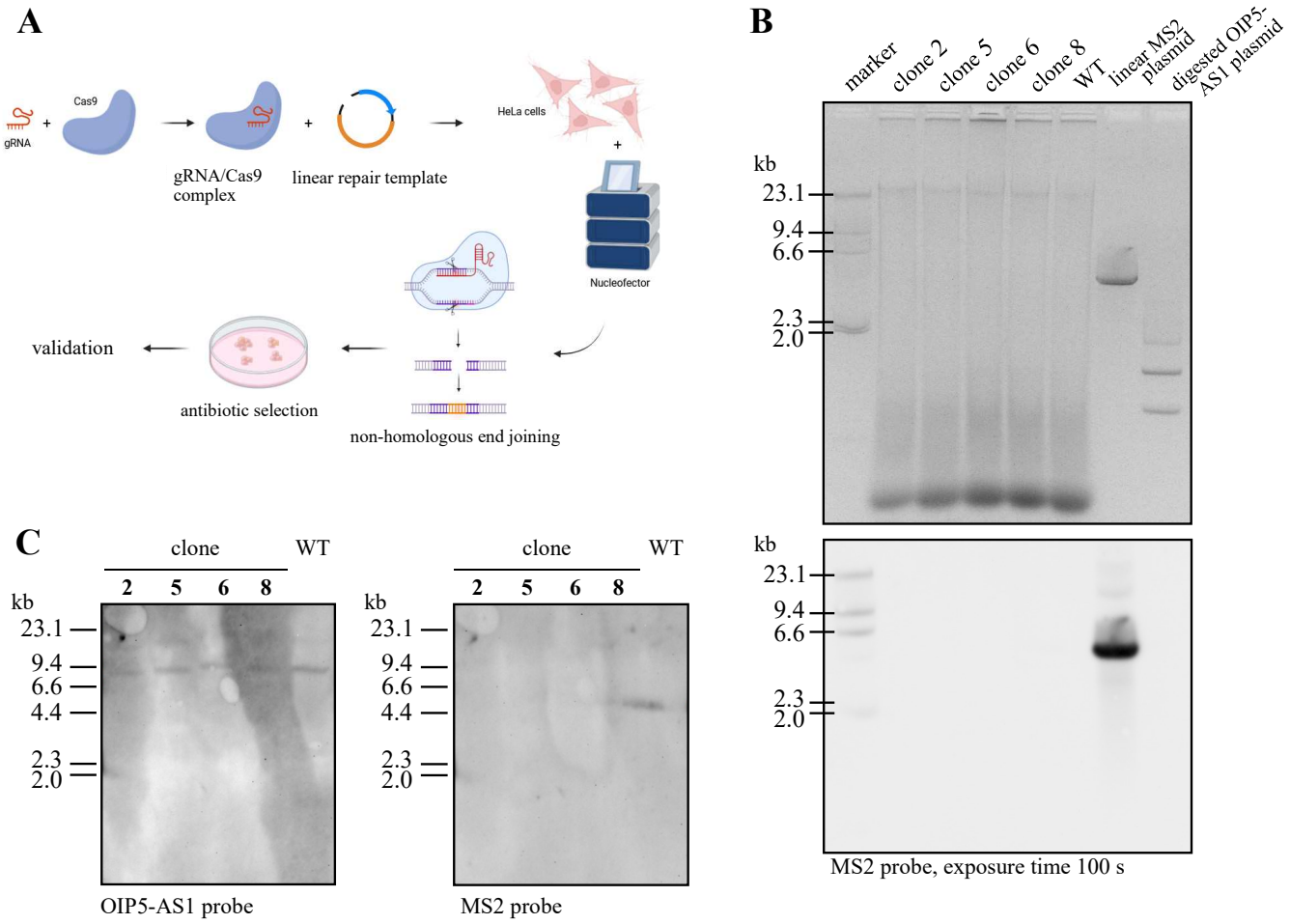


Figure 12.: CRISPR/Cas9-mediated targeted insertion of 24xMS2 loops into the lncRNA OIP5-AS1 via non-homologous end joining. (A) The schematic workflow preformed to tag the lncRNA OIP5-AS1 with 24x MS2 loops. The recombinant Cas9 protein and chemically stabilized sgRNA were joined and together with a linearized repair plasmid transfected into HeLa cells with the help of a nucleofector. Three days afterwards, puromycin was added as selective pressure and the growing monoclonal cell lines were isolated and thoroughly analyzed. **(B), (C)** Southern blot results of CRISPR clone 2, 5, 6, and 8. **(B)** Genomic DNA from four different CRISPR clones and from the parental cell line as a negative control were digested with the restriction enzyme BlpI. The linearized or digested plasmids used as template for probe synthesis served as positive control. Subsequent agarose gel electrophoresis is shown in the gel picture, upper image. Southern blotting showed one band each at the expected size for the positive controls, an example blot incubated with the MS2 DIG probe can be seen in the lower picture with a short exposure time of 100 s. **(C)** Probing of the membrane after Southern blot and stripping and re-probing revealed no bands corresponding to the expected fragment size with MS2 loops (13,178 bp), neither with the OIP5-AS1 probe, nor with the MS2 probe. Only fragments corresponding to the unmodified locus (8,186 bp) were observed (left blot). Both blots were imaged with an exposure time of 2122 s. WT = HeLa-EM2-11ht cells.

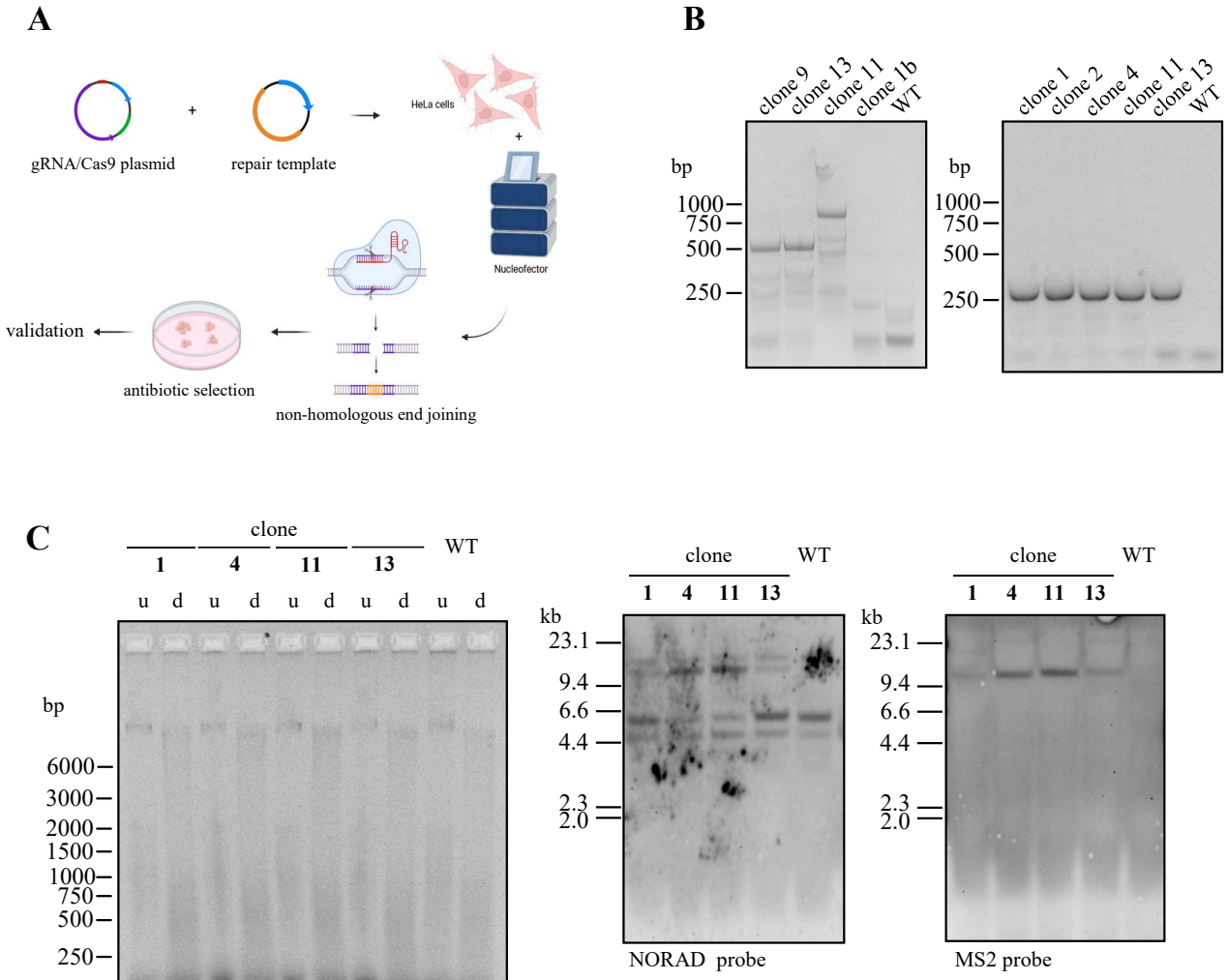


Figure 13.: CRISPR/Cas9-mediated targeted insertion of 24xMS2 loops into the lncRNA NORAD via non-homologous end joining. (A) Schematic workflow to tag the lncRNA NORAD with 24x MS2 loops. The first plasmid expresses a Cas9 enzyme and two different gRNAs, the second plasmid is the donor plasmid and serves as repair template encoding 24xMS2 loops. Both are transfected into HeLa cells with the help of a nucleofector. Inside the cells, simultaneous DSBs are induced in the genome by one Cas9-sgRNA complex, in the repair plasmid for linearization by the vector-targeting Cas9-sgRNA complex. Then, the linearized vector can be integrated into targeting site via the NHEJ mechanism. Three days afterwards, puromycin was added as selective pressure and the growing monoclonal cell lines were isolated and thoroughly analyzed. **(B)** Genomic DNA from different CRISPR clones and from the parental cell line as a negative control were isolated and analyzed by PCR. Left side: The downstream junction between repair template and genome was amplified and yielded a band in the agarose gel at the expected size of 688 bp for clone 9 and 13, but not for the WT negative control. Right side: The upstream junction between the MS2 loops and NORAD gene was successfully amplified from all clones tested here and gave the expected product size pf 367 bp. This implicates that the full length construct was integrated in the NORAD genome locus.

Figure 13.: (Continued from previous page.) **(C)** After gDNA isolation, it was digested with the enzyme AlfI. To confirm complete digestion, all samples were compared to their undigested gDNA as shown in the exemplary gel picture on the left. "u" corresponding to undigested gDNA, "d" to digested samples recognizable by the smear of DNA. For the Southern blot, the restriction digested gDNA was separated by size with an agarose gel and blotted to a nylon membrane. The linearized or digested plasmids used as template for probe synthesis served as positive control (not shown). Probing of the membrane, stripping and re-probing revealed band with the expected size with and without 24xMS2 loops: 6060 and 11052 bp, respectively with the NORAD probe (left blot) and 11052 bp with the MS2 probe (right). The NORAD and MS2 blots were imaged with an exposure time of 1413 s and 1255 s, respectively. WT = HeLa-EM2-11ht cells.

The CRISPR construct containing the most efficient gRNA (Ia) and the donor plasmid were co-transfected into the TurboID-expressing HeLa-EM2-11ht cell line (see Figure 13(A) for workflow). Following antibiotic selection with puromycin, 14 colonies were obtained. These clones were expanded to monoclonal cell lines and analyzed for successful integration of the MS2 loops. To confirm the correct integration of the construct at the NORAD genome locus, gDNA was isolated, and junction PCRs were performed (see Figure 13 (B)). Two primer sets were used to validate the integration. Primer 6376 and 7268 were used to amplify the downstream junction (puromycin gene to genome) and yielded the expected 688 bp fragment for clone 9 and 13 (left gel picture). The upstream junction PCR was done with primer pair 6203 and 7292 (genome to MS2 loops). The PCR yielded the expected 367 bp fragment for all clones (right gel picture). The WT negative control did not show any fragment, confirming specificity of the primer pairs. The presence of these fragments confirmed that the full-length construct, including the 24xMS2 loops, was successfully integrated into the NORAD locus.
To further validate the integration, Southern blot was performed. gDNA was digested with AlfI, and complete digestion was confirmed by comparing digested samples to undigested gDNA on an agarose gel (shown in the exemplary gel picture in Figure 13 (C), on the left). For the Southern blot, the fully digested gDNA was separated by gel electrophoresis, transferred to a nylon membrane, and analyzed using DIG-labeled probes specific to NORAD and the MS2 loops. The linearized or digested plasmids used as a template for probe synthesis served as a positive control (not shown). The NORAD blot revealed bands at the expected sizes of 6060 bp (without MS2 loops) and 11052 bp (with MS2 loops), confirming heterozygosity of the integration (left blot). However, two additional bands were observed, suggesting nonspecific binding of the NORAD probe. A BLAST analysis revealed that the probe had significant sequence similarity to pseudogenes on chromosome 12 (annotated as OVCH1-AS1) and on chromosome X, which matched the sizes of the observed bands (13413 bp and 5126 bp, respectively). Stripping and re-probing with the MS2-specific probe detected a single band at 11052 bp for all clones but not the WT, consistent with successful integration of the MS2 loops.
Despite this, smiFISH microscopy analysis showed no detectable expression of the MS2 loops. This suggests that while the MS2 loops were correctly inserted into the NORAD locus, they are not transcribed as part of the lncRNA.
In conclusion, the presence of MS2 loops at the targeted NORAD genome locus was confirmed by both PCR and Southern blotting. However, the lack of expression indicates a functional issue, possibly due to interference of the sequences in the donor plasmid, including the puromycin resistance cassette.

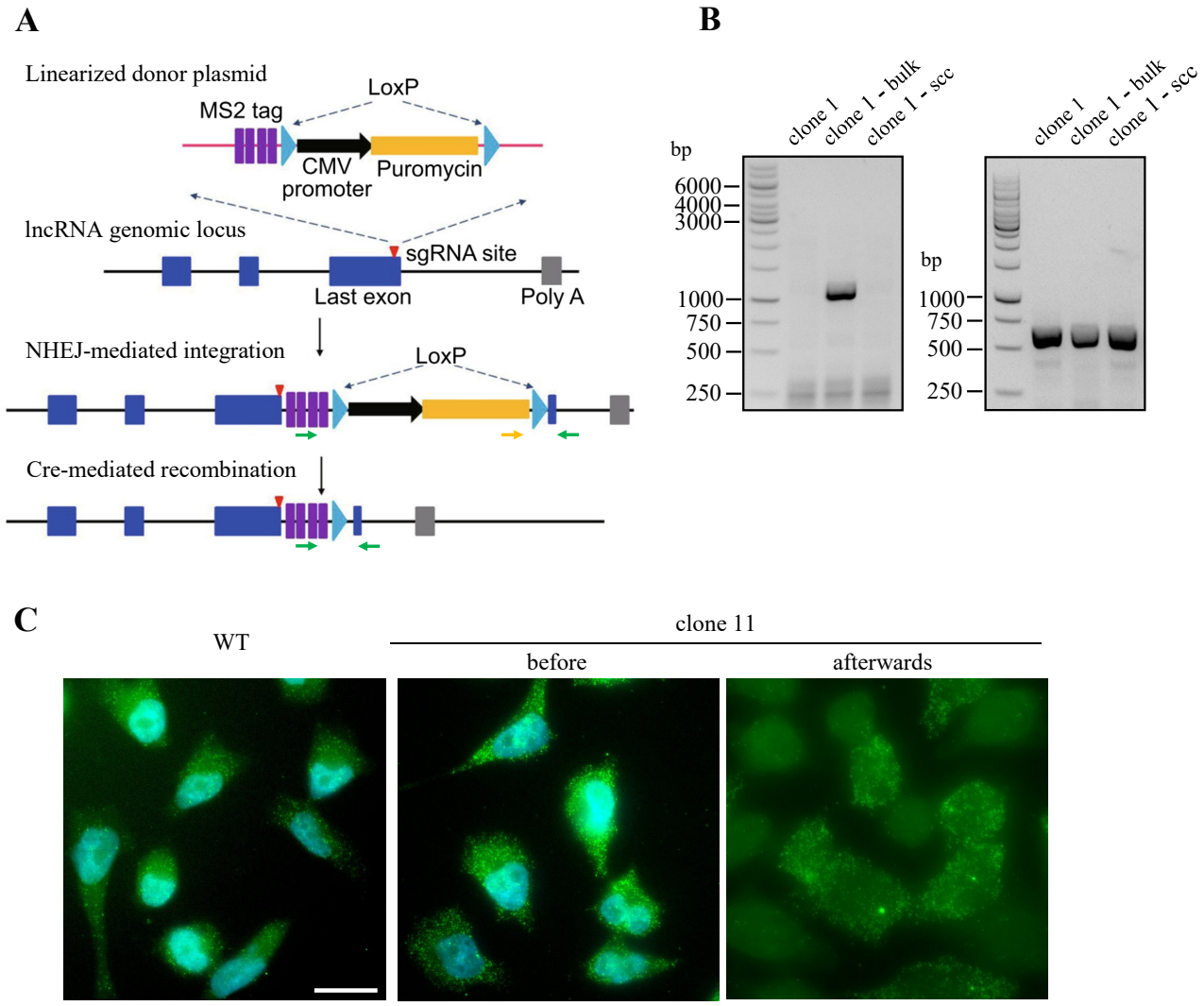


Figure 14.: Cre-mediated recombination in the lncRNA NORAD genome locus tagged with 24xMS2 loops. (A) Schematic workflow for tagging lncRNAs with MS2 loops by non-homologous end joining and removing the backbone of the plasmid containing the CMV promoter and puromycin resistance gene by Cre-mediated recombination by taking advantage of the loxP sites. Adapted from [117]. **(B)** Genomic DNA from CRISPR clones after transfection of the Cre recombinase and from the parental cell line as a negative control were isolated and analyzed by PCR, here shown for clone 1. Left side: Recombination was confirmed via junction PCR, the expected fragment size before Cre-mediated removing of the resistance gene is 4577 bp, after successful recombination 1157 bp. A band corresponding to this size can only be seen after the bulk sort (middle lane), but not in the single cell clone (clone 1 - ssc). Right side: The downstream junction between the puromycin gene and NORAD gene was amplified and should yield a fragment of 688 bp before recombination, while afterwards no product should be produced as the froward primer anneals in the puromycin gene. Interestingly, all samples show this band. **(C)** smiFISH microscopy images, WT cells and clone 11 before recombination show normal and equal expression of NORAD (green signal) distributed over the whole cell as expected. Cre mediated recombination produced a heterogeneous cell population, with some cells showing barely any NORAD expression, while some show a normal level. DNA in cell nuclei was stained with DAPI (blue signals). All images are maximum intensity z-stacks. Scale bar = 20 µm. WT = HeLa-EM2-11ht cells.

To potentially restore expression of the integrated MS2 loops, a Cre recombinase plasmid (p2275) was transiently transfected. Cre recombinase, encoded as a GFP fusion protein, facilitates recombination between loxP sites, excising most of the repair template and leaving only the MS2 loops. This step was also suggested by the authors of the original study [117] (see Figure 14(A)) and will determine whether removing the additional plasmid elements restores proper transcription of the tagged lncRNA.
The Cre-GFP fusion protein expression enables fluorescence-activated cell sorting (FACS) to enrich for successfully transfected cells. GFP-positive cells were sorted into 96-well plates to obtain single-cell clones (scc). However, the survival rate was very low, probably due to stress during transfection and the single-cell sorting process. Sorting GFP-positive cells from NORAD clones 1 and 11 and WT cells as control into separate 96-well plates yielded in survival of one colony from the wild type and one colony from clone 1. These surviving cells were expanded for further analysis. Given the poor outcome from single-cell sorting, another FACS attempt involved bulk sorting GFP-positive cells of clones 1 and 11 into T25 flasks each. While many cells died during this process as well, some survived and were expanded.
To evaluate the success of Cre-mediated recombination (CmR), three assays were performed. As CmR should remove the puromycin resistance gene, this should render these cells sensitive to puromycin. However, the NORAD clone 1 which survived after single-cell sorting (clone 1- scc), also survived puromycin treatment in contrast to WT cells which served as negative control and died. Interestingly, the bulk-sorted populations showed partial sensitivity to puromycin treatment, with many cells initially dying upon addition of puromycin. Some, however, survived and eventually overgrew the plate when maintained in culture for extended periods. This indicates that recombination occurred only in a subset of cells.
Genomic DNA was isolated from the bulk-sorted and single-cell population and analyzed by PCR using primers designed to detect the recombination event. gDNA of clone 1 prior to CmR was used as control. For the upstream junction, recombination was confirmed by PCR (Figure 14(A), green primer pair), where the expected fragment size before Cre-mediated excision was 4577 bp, and after successful recombination was 1157 bp (see Figure 14(B), left gel). A band of 1157 bp was observed only in the bulk-sorted population from clone 1 (middle lane), but not in the single-cell clone (clone 1- scc). For the downstream junction, the expected fragment size was 688 bp before recombination, with no product expected after recombination (Figure 14(A), yellow and green primer). However, all samples still produced a 688 bp band, suggesting incomplete recombination or a mixed population of cells (Figure 14(B), right gel).
smiFISH was performed to assess MS2 expression and NORAD RNA levels. No MS2 expression was detected in any population. Interestingly, approximately half of the cells displayed a drastic reduction in NORAD RNA spots (green signal), indicating that Cre-mediated recombination likely disrupted the NORAD locus in some cells, reducing its expression (see Figure 14(C).
These results suggest that Cre-mediated recombination occurred in a subset of cells, but the recombination process appears to have disrupted the integrity of the NORAD locus, leading to reduced expression of the lncRNA. Furthermore, the mixed population observed after bulk sorting highlight the inefficiency and heterogeneity of the recombination process. In sum, it is unlikely that these cells can be salvaged for further experiments. A new CRISPR approach may be necessary, potentially using a different gRNA that targets a more optimal region of the NORAD locus and will not disrupt RNA expression. The current gRNA Ia targets a site

shortly after the transcription start site of NORAD, which may be suboptimal for maintaining its expression after tagging.

CRISPR attempt based on microhomology-mediated end joining

Both NHEJ-based CRISPR approaches to integrate MS2 loops into the OIP5-AS1 or NORAD gene failed, though for different reasons. To address these issues, the MMEJ repair pathway was explored as an alternative. MMEJ uses short microhomology arms (MHAs) flanking the DSB introduced by Cas9 to direct and mediate precise DNA repair. For example, Chen et al. (2020) employed CRISPR/Cas9-induced MMEJ to insert 24xMS2 loops into the well-studied lncRNA NEAT1 [118]. They used primers which, with one half of their sequence, anneal to their repair template containing the tag and a selection marker. The other half of the primers consists of ∼20 bp overhangs constituting the MHAs, with the 5' end being thiophosphorylated. By amplifying the repair construct by PCR from a plasmid, they yielded a degradation-protected, linear DNA fragment with MHAs from the primer overhangs, significantly improving integration efficiency. Earlier reports similarly highlighted the role of phosphorothioate (PS) bonds in stabilizing linear donor DNA for successful insertions [114] [113].

Building on these findings, the CRISPR/Cas9 system was employed in conjunction with MMEJ to attempt the integration of MS2 loops into the OIP5-AS1 locus. A donor template containing MS2 loops and a blasticidin resistance gene was generated by PCR amplification from the p2321 plasmid. Primers 7754 and 7755 were modified with PS bonds at their last three 5' nucleotides to enhance the stability of the linear DNA fragment and included 20 bp overhangs serving as MHAs flanking the integration site. The same modified gRNA 2a as before and recombinant Cas9 enzyme were co-transfected with the linear donor template into TurboID-expressing HeLa-EM2-11ht cells. After antibiotic selection with blasticidin, 16 clones survived and were expanded to monoclonal cell lines (see Figure 15(A) for a schematic workflow).
For validation of the 24xMS2 loop knock-in, firstly genotyping PCRs were performed. Amplification using primers targeting the MS2 loop sequence and the blasticidin resistance gene was successful for some clones, confirming that the majority of the repair template had been integrated (see Figure 15 (B)). However, junction PCRs targeting the upstream and downstream regions of the intended integration site were unsuccessful, suggesting that the insertion might have occurred at an undesired location.
To further confirm the location of the integration, Southern blotting was performed using DIG-labeled probes against the endogenous OIP5-AS1 locus and the MS2 loops (see Figure 15 (C)). Restriction digestion of genomic DNA with SspI and subsequent blotting revealed unexpected bands, indicating that the MS2 loops had likely been inserted at off-target sites. The expected fragment sizes for correct integration were 5104 bp (with MS2 loops) and 2718 bp (without MS2 loops), but additional unspecific bands were detected, suggesting misintegration. Additionally, clone 8 and 11 showed two bands in the MS2 blot, indicating multiple integration events. Why the OIP5-AS1 blot shows two bands below 2 kb including the WT sample is unclear. A BLAST analysis of the OIP5-AS1 probe revealed a potential off-target hybridization site on chromosome 1, corresponding to a 1969 bp fragment, which explains one of the bands.
Microscopy analysis did not reveal evidence of MS2 loop expression (see Figure 15 (D)) but only

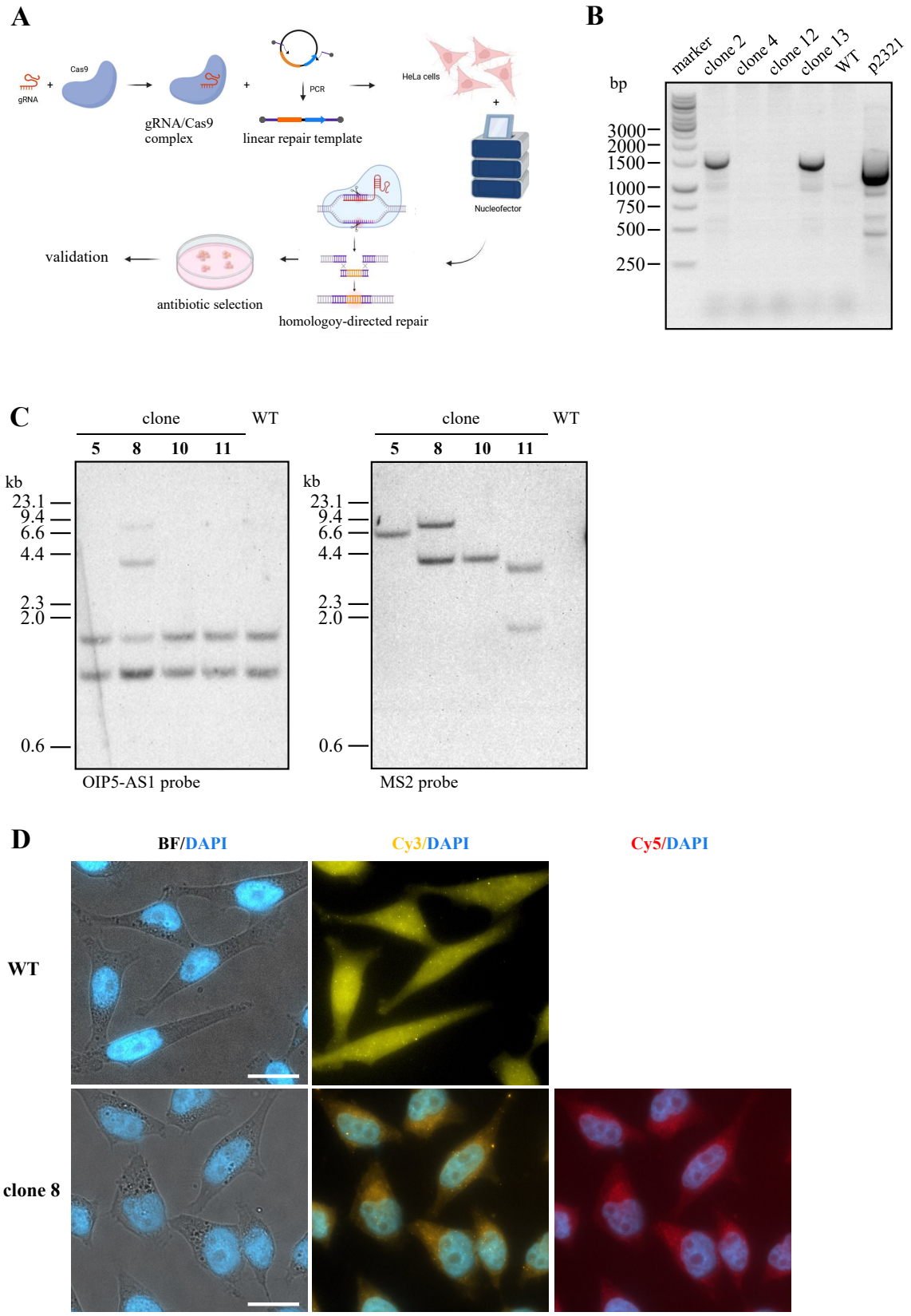


Figure 15.: The CRISPR/Cas9 system based on microhomology-mediated end joining was used to attempt the integration of MS2 loops into the OIP5-AS1 locus.

Figure 15.: (Continued from previous page.) **(A)** A donor template containing MS2 loops (orange) and a blasticidin resistance gene (blue) was generated by PCR amplification from a plasmid. The primers used include 20 bp MHAs flanking the integration site, and were modified with PS bonds (gray) at their 5' end to enhance the stability of the linear DNA fragment. The modified gRNA and recombinant Cas9 enzyme were co-transfected with the donor template into TurboID-expressing HeLa-EM2-11ht cells by nucleofection. After integration, antibiotic were added to select for successful CRISPR events. **(B)** Genomic DNA from different CRISPR clones was isolated and analyzed by PCR spanning the whole repair template. Primer 7291 and 7354 should produce a fragment of 1624 bp, which can be seen in clone 2 and 13, as well as in the positive control (p2321), while the negative control (WT) shows no band as expected. **(C)** 12 µg of gDNA was digested by restriction enzyme SspI and analyzed by Southern blot. The left side should show one band with 2718 bp in all samples and after correct integration, clones should show a band with 5104 bp. The right blot was expected to show only one band with 5 kb in CRISPR clones. However, the repair template seemed to be integrated in several loci, as for clone 8 and 11 two bands can be seen. Exposure time for both blots was 3600 s. **(D)** smiFISH microscopy images, WT images for comparison showing equal distribution of OIP5-AS1 in the cell (yellow signal). Microscopy analysis of CRISPR clones, here shown for clone 8, did not reveal evidence of MS2 loop expression (red signal). DNA in cell nuclei was stained with DAPI (blue signals). All images are maximum intensity z-stacks. Scale bar = 20 µm. WT = HeLa-EM2-11ht cells.

background signal in the Cy5 channel, further supporting the conclusion that the integration was unsuccessful at the intended locus.
In sum, the presence of MS2 loops in the genome was confirmed by PCR and Southern blotting, but the results indicate that the loops were not integrated at the desired region of the OIP5-AS1 locus. This likely explains the failure of junction PCRs from MS2 loops to the genome.

The MMEJ pathway was also employed to integrate 24xMS2 loops into the lncRNA NORAD locus, using a revised CRISPR approach designed to avoid disrupting RNA expression. Two newly designed gRNAs (IVa and IVb), targeting a more optimal region of the NORAD gene, were encoded on plasmids along with the Cas9 enzyme (p2414 and p2415). The linear repair template, generated by PCR from plasmid p2341, included microhomology arms (MHAs) specific to the NORAD genomic locus. The primers used for amplification were protected from degradation by phosphorothioate (PS) modifications at their 5' ends. The CRISPR plasmid and repair template were nucleofected into TurboID-expressing HeLa-EM2-11ht cells (Figure 16(A)). After antibiotic selection with blasticidin, 14 clones were obtained and subjected to a series of validations, including genotyping PCR, Southern blot analysis, and smFISH.
First, clones were screened by genotyping PCR. PCR amplification targeting the MS2 loop sequence and the blasticidin resistance gene confirmed the presence of the complete repair template in the genome, indicating successful integration in most clones. However, junction PCR results were less promising. The upstream junction amplification was successfully in some clones, but downstream junction PCR frequently failed. These results suggested incomplete or imprecise integration in most cases (data not shown).
To verify MS2 loop integration and assess the accuracy of the insertion site, genomic DNA was digested with HindIII and analyzed by Southern blot using DIG-labeled probes specific to the NORAD locus and the MS2 loops (see Figure 16(B)). The expected fragment sizes were 5736 bp for the NORAD locus with MS2 loops and 3379 bp for the wild type NORAD locus. Southern blot results confirmed the presence of MS2 loops in several clones. However, the detected fragments had incorrect sizes or showed multiple bands, indicating off-target integration

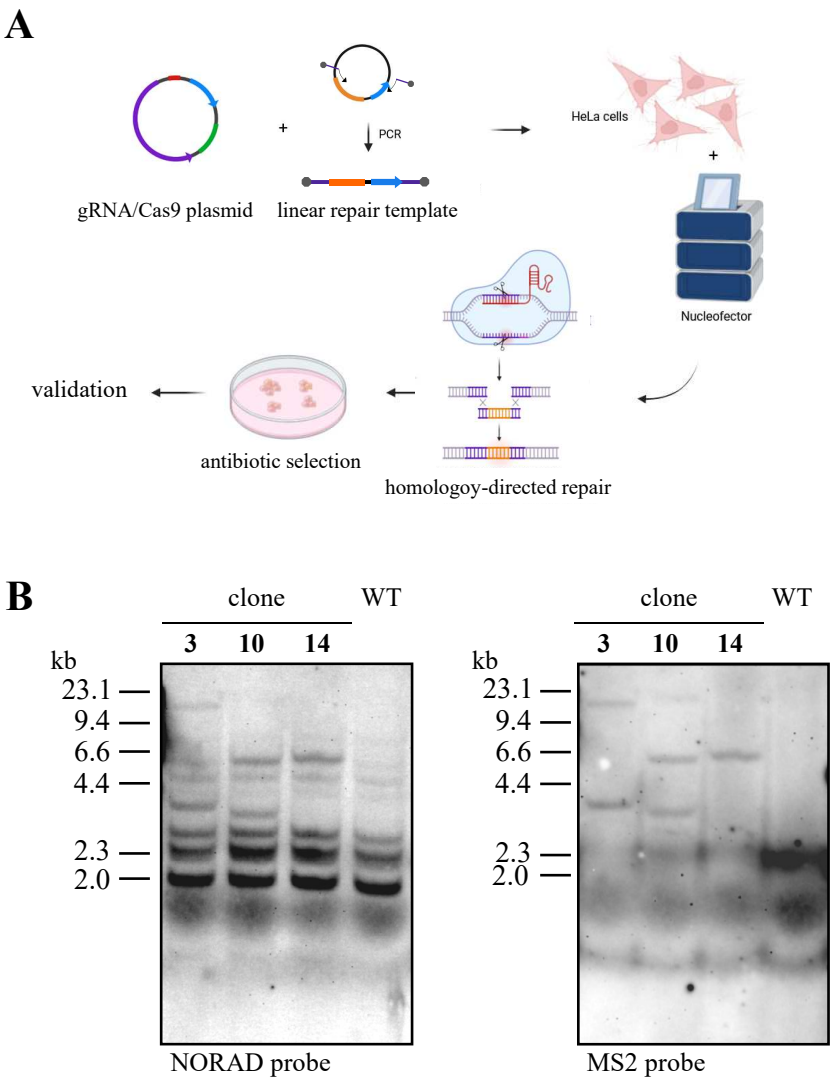


Figure 16.: CRISPR/Cas9 system based on microhomology-mediated end joining was used to attempt the integration of MS2 loops into the NORAD locus. (A) A donor template containing MS2 loops (orange) and a blasticidin resistance gene (blue) was generated by PCR amplification from a plasmid. The primers used include MHAs flanking the integration site, and were modified with PS bonds (gray dots) at their 5' end to enhance the stability of the linear DNA fragment. A plasmid encoding the NORAD targeting gRNA and a Cas9 enzyme were co-transfected with the donor template into TurboID-expressing HeLa-EM2-11ht cells by nucleofection. After integration, antibiotic was added to select for successful CRISPR events. **(B)** gDNA was digested by restriction enzyme HindIII and analyzed by Southern blot. The blot on the left side should show one band with 3379 bp in all samples and after correct integration, clones but not the WT negative control should show a second band with 5736 bp. Unfortunately the NORAD probe seems to bind to several targets. The right blot was expected to show only one band with 5736 bp in CRISPR clones. However, the repair template seemed to be integrated in several loci. Exposure time for both blots was 2400 s. WT = HeLa-EM2-11ht cells.

or genomic rearrangements. The NORAD-specific probe hybridized with several unexpected bands, likely due to the presence of pseudogenes or homologous sequences in the HeLa genome. A BLAST analysis of the NORAD DIG probe revealed similarities of the probe to several genomic loci, explaining most additional bands. Notably, a 2 kb band was observed on the blot that could not be explained by the expected digestion patterns. To improve the NORAD blot result showing multiple bands, as a first trial the hybridization conditions were optimized to reduce nonspecific binding, although without initial success. Alternatively, using different restriction enzymes for digestion could be considered, but with the previously used enzyme (see Figure 13) also two unspecific bands appeared, indicating that due to the nature of NORAD it appears to be difficult to achieve highly specific binding.
smFISH was used to detect MS2 loop expression in NORAD transcripts. Despite the confirmed integration of MS2 loops in some clones, no MS2 signal was detected, indicating that the tagged NORAD RNA was either not being expressed or the knock-in of MS2 loops was unspecific and happened in a silent locus.
While the MMEJ approach enabled the integration of MS2 loops into the genome, the results were not satisfactory. Most clones failed junction PCR validation, and Southern blot analysis revealed significant off-target integration and potential genomic rearrangements. Furthermore, no MS2 loop expression was detected in smFISH, rendering the generated clones unsuitable for further experiments. To resolve these issues, a new CRISPR approach with improved specificity is necessary. To increase the likelihood of successful and precise integration longer homology arms could help guide the repair after DSB.

CRISPR attempt based on homology-directed repair

As a final attempt to achieve successful MS2 loop integration at the correct position, a strategy leveraging the homology-directed repair (HDR) pathway was employed. HDR, a less error-prone mechanism compared to NHEJ or MMEJ, was expected to facilitate precise insertion of the 24xMS2 cassette into the OIP5-AS1 locus. A plasmid consisting of two ∼800 bp homology arms was suggested by Spille et al. (2019) [60]. Following this suggestion, two repair plasmids, p2340 and p2341, were created with almost 900 bp homology arms flanking a selectable blasticidin resistance gene and the 24xMS2 loops. Both plasmids were similar, except that p2341 featured loxP sites instead of gRNAs flanking the resistance gene and its promoter to allow subsequent removal via transfection of Cre recombinase.
Two transfection strategies were tested (see Figure 17(A)). In the first, two plasmids were co-delivered: one encoding the OIP5-AS1-targeting gRNA 2a and the Cas9 enzyme, while the other one represents the circular repair template. In the second strategy, a recombinant Cas9 enzyme pre-loaded with the modified gRNA 2a was co-transfected with the same circular donor template. Following nucleofection into TurboID-expressing HeLa-EM2-11ht cells, the gRNA-Cas9 complex introduces a DSB at the genomic locus of OIP5-AS1 as well as in the circular repair template, which harbors the same sequence and is thereby linearized. After HDR, antibiotic selection with blasticidin yielded 16 clones in total. From the colonies that survived, 3 clones arose from transfection using p2340 and 13 from those using p2341.

To validate the integration of the MS2 cassette, genomic DNA from the clones was analyzed

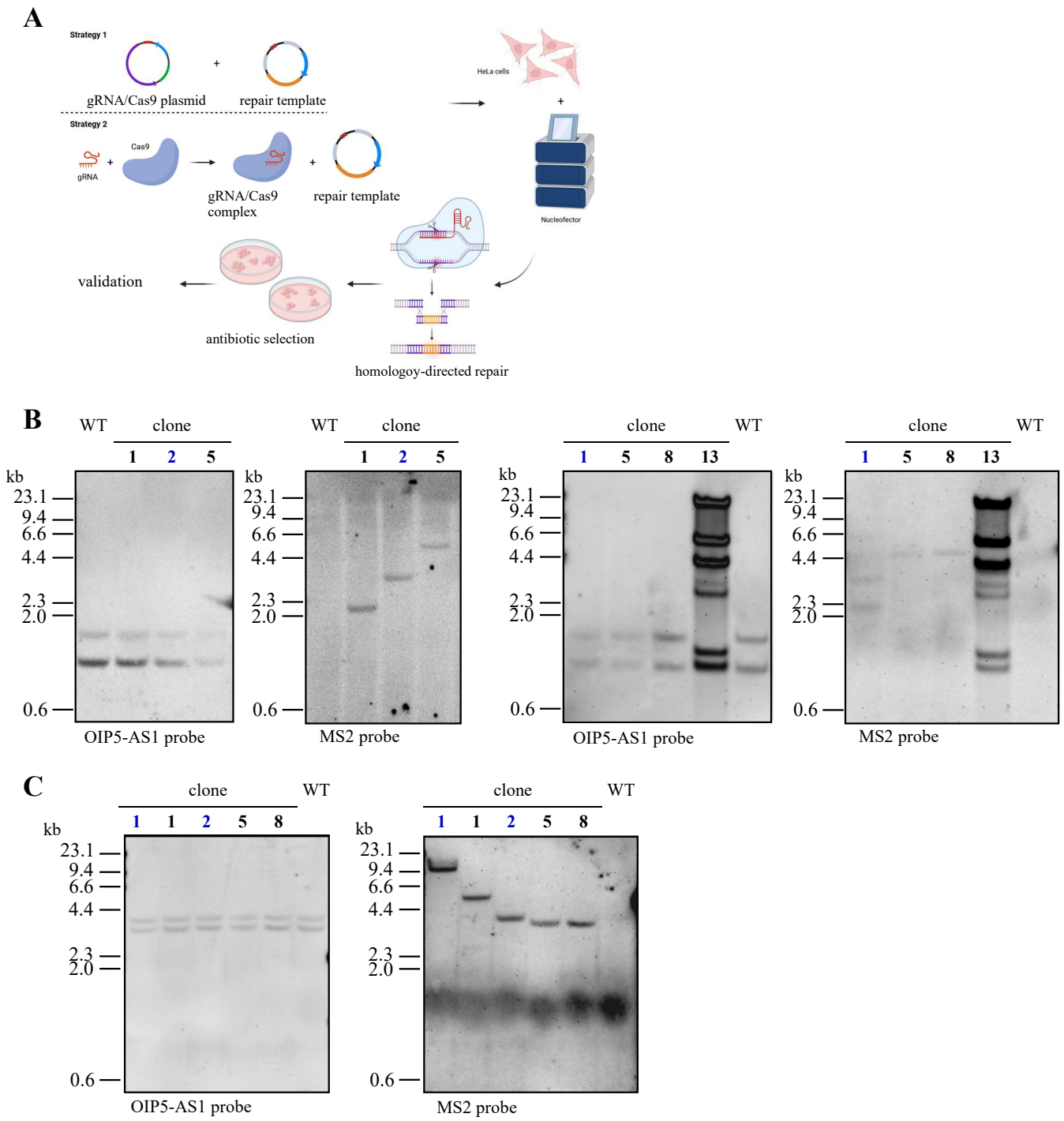


Figure 17.: CRISPR/Cas9 system based on homology-directed repair was used to attempt the integration of MS2 loops into the OIP5-AS1 locus. (A) Two different strategies were used, the first one implies two plasmids, one encoding the OIP5-AS1 targeting gRNA and a Cas9 enzyme, the other represents the donor template containing MS2 loops (orange) and a blasticidin resistance gene (blue) surrounded by homology arms (gray). The second strategy consists of a modified gRNA and a recombinant Cas9 enzyme that were co-transfected with the same donor template as in strategy 1. After nucleofection into TurboID-expressing HeLa-EM2-11ht cells antibiotic selection was performed to select for successful CRISPR events.

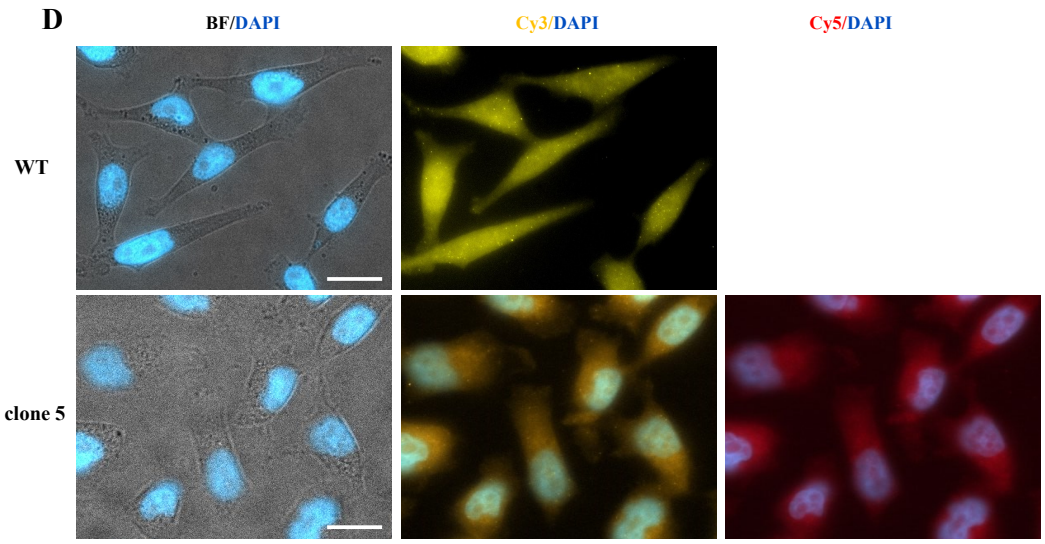


Figure 17.: (Continued from previous page.) **(B)** gDNA was digested by restriction enzyme SspI and analyzed by Southern blot. Six different clones were examined on two different gels, the expected fragment sizes is 5012 bp with MS2 loops and 2718 bp without MS2 loops. Clones with a blue number contain repair template 2340, black numbers indicate repair plasmid p2341. Exposure time for both blots on the left was 1800 s, on the right 1875 s. The MS2 blot shows bands at various sizes, with only clone 5 and 8 potentially at the expected size of 5 kb. As both OIP5-AS1 blots show a double band even in the WT sample, the Southern blot assay was repeated with another restriction enzyme. **(C)** gDNA was digested by two restriction enzymes (NdeI and SacI) and analyzed by Southern blot. Clones with a blue number contain repair template 2340, black numbers indicate repair plasmid p2341. Exposure time for the OIP5-AS1 blot was 720 s and for the MS2 blot 2400 s. The left side should show one band with 5545 bp in all samples and after correct integration, clones should show a second band with 7839 bp. However, a double band below 4.4 kb can be seen again. The right blot was expected to show a band with 5736 bp in CRISPR clones. Nonetheless, none of the bands matched this size. **(D)** smiFISH microscopy images, WT images show equal distribution of OIP5-AS1 in the cytoplasm (yellow signal). Exemplary shown here for clone 5, microscopy analysis of CRISPR clones did not show any MS2 signal (red), but lower OIP5-AS1 expression compared to WT. DNA in cell nuclei was stained with DAPI (blue signals). All images are maximum intensity z-stacks. Scale bar = 20 µm. WT = HeLa-EM2-11ht cells.

using PCR. Amplification targeting the repair template indicated successful integration of the full-length donor construct in most clones. However, junction PCRs designed to test the correct integration over the homology arms were less successful. Only a few clones showed correct upstream junctions, and only two clones exhibited correct integration at both junctions (data not shown).
Southern blot analysis was performed to further investigate the integration and confirm its accuracy (see Figure 17(B)). Using the restriction enzyme SspI and a DIG-labeled probe specific to OIP5-AS1, the expected fragment sizes were 5012 bp (with MS2 loops) and 2718 bp (without MS2 loops). However, two bands were detected in all clones, including the WT negative control, and the detected sizes did not match the predictions. This result is similar to the one seen in Figure 15(C). As BLAST analysis of the OIP5-AS1 probe could only explain one of the bands (potential off-target hybridization site on chromosome 1, with a 1969 bp fragment) the second band could refer to OIP5-AS1, although the expected size is 2.7 kb and not ∼1 kb. To address the inconsistent results, an alternative enzyme combination (NdeI and SacI) was tested (see Figure 17(C)). With this approach, the expected fragment sizes were 7839 bp (with MS2 loops) and 5545 bp (without MS2 loops). Again, Southern blot results revealed two bands with the OIP5-AS1 probe and no evidence of correct integration. Probing with an MS2-specific DIG-labeled probe detected bands at various sizes, but none matched the expected size, including for clones 5 and 8, which initially appeared promising.
To explain the unexpected double band observed in both WT cells and CRISPR clones when probed with the DIG-labeled OIP5-AS1 probe (Figure 15 and Figure 17), one hypothesis was that the restriction enzyme has an additional cleavage site in the OIP5-AS1 locus between probe hybridization site and gRNA region used for CRISPR. This could potentially give rise to a shorter fragment than expected on the membrane. To rule this out, PCR from wild type HeLa gDNA was performed, covering the region in question. Sequencing of the PCR result showed no additional cleavage site. Therefore, it was hypothesized that the observed effect is due to incorrect genome annotation or a cell line-specific copy-number effect.
smiFISH was performed to evaluate the expression of OIP5-AS1 and to detect MS2-tagged transcripts (see Figure 17(D). Several clones exhibited low expression levels of OIP5-AS1 compared to the WT, especially the cell lines with multiple integration events according to the Southern blot analysis. While with normal expression levels of OIP5-AS1 up to 40 molecules per cell can be detected (own observation and [93]) CRISPR clones only showed $\sim$ 10 RNA spots. No evidence for MS2 loop expression were detected in any clone (red signal), indicating either silencing or misintegration in a non-expressed locus of the MS2 cassette.
Despite using the HDR pathway and donor templates with homology arms, the integration of the 24xMS2 cassette into the OIP5-AS1 locus was unsuccessful. While the MS2 array was integrated into the genome of all analyzed clones, Southern blotting and genotyping PCRs indicated that the integration event did not happen at the intended locus. The digestion patterns revealed unexpected bands in the Southern blot, and smFISH showed that the MS2 loops did not appear to be expressed as part of the OIP5-AS1 transcript.
These results highlight the challenges of achieving precise lncRNA tagging and suggest that further optimization of the CRISPR approach is needed, potentially involving alternative gRNA designs or repair templates. While CRISPR/Cas9 editing approaches have been widely used, the application to lncRNAs remains challenging due to the unique properties of these tran-

scripts, including their unknown and often complex secondary structures, and genomic context. It is possible that the available techniques for precise lncRNA tagging are not yet fully optimized. Additionally, the selected lncRNA candidates may have been suboptimal for such modifications. Given that various promising strategies, including different repair pathways, were tested without any success, continuing with this approach no longer seemed feasible.

9 Establishing an *in vitro* proximity labeling assay for lncRNA

9.1. Introduction

9.1.1. The lncRNA NEAT1

Nuclear-Enriched Abundant Transcript 1 (NEAT1) is a highly abundant, ubiquitously expressed, and nuclear-localized lncRNA that is conserved across placental mammals. It is essential for the formation and function of paraspeckles, a type of mammal-specific, membrane-less nuclear body. The NEAT1 gene produces two isoforms through alternative transcription termination, the short transcript is 3.7 kb in length and called NEAT1_1, the longer one 23 kb and called NEAT1_2. While both isoforms localize to paraspeckles, NEAT1_2 plays an architectural role in their assembly, and its depletion leads to the disintegration of these structures [118], [119]. Paraspeckles contribute to gene expression regulation, RNA quality control, and miRNA processing. The number and size of paraspeckles are dynamically changed depending on cellular conditions [118].
NEAT1 RNA interacts with multiple paraspeckle components, including NONO, SFPQ, and PSPC1. In addition to its role in paraspeckle formation, NEAT1 regulates miRNA levels by sequestering RNAs through sponge-like activities and influences epigenetic gene expression by directly interacting with chromatin. As confirmed by affinity purification followed by mass spectrometry, it also interacts with RNA splicing factors [73], [86], [119].
As this lncRNA is fairly well understood, Yap et al. (2021) used NEAT1 (among others) as a model lncRNA to establish their assay called HyPro (see for subsection 9.1.3 for details). The authors revealed that paraspeckles also influence the accumulation of incompletely processed, adenosine-to-inosine-edited transcripts at the interface of their encoding chromosomal regions and the nuclear body. Furthermore, they observed that paraspeckles can establish specific contacts with genomic regions, and concluded that NEAT1 might be involved in higher-order

nuclear organization [64].
Despite its known role in nuclear architecture, NEAT1 is not essential for normal development in mice, indicating that its functions are more critical in specific cellular or stress-related contexts [10].

9.1.2. Proximity labeling with biotin peroxidases

As described in chapter 4, besides ligases, peroxidases can be used in PL assays to covalently tether biotin to nearby molecules. There is a smaller variety of promiscuous biotin peroxidases available compared to biotin ligases, with the main players being APEX [27], [28] and its improved version APEX2 [49].
APEX (ascorbate peroxidase) is a 28 kDa monomeric enzyme originally derived from a cytosolic plant peroxidase [28]. To enhance its limited activity, APEX2 was engineered through yeast-display evolution, resulting in significantly improved catalytic efficiency. APEX2 exhibits very fast labeling kinetics, enabling biotinylation within less than one minute. This rapid labeling allows capturing interactomes at high temporal resolution, making APEX2 particularly useful for studying dynamic intracellular interactions [24], [49].
A significant discovery was that APEX2 can directly biotinylate RNA, allowing the identification of RNAs associated with organelles or specific cellular structures. This has expanded the application range of proximity labeling, and the purification and subsequent sequencing of subcellular RNA populations led to the first comprehensive atlas of RNA distribution within cells [51], [52], [120]. Additionally, dynamic RNA-protein complexes, like the translation initiation complex, have been investigated using this approach [51].
Despite its advantages, peroxidase-based PL has certain limitations. Not all cell types readily take up the biotin-phenol substrate, and the requirement for hydrogen peroxide restricts its use *in cellulo* and hinders *in vivo* applications, due to its cytotoxicity and unstable nature [121]. Additionally, the application of peroxidase-based PL in plants is difficult because of the background activity from endogenous plant peroxidases which can obscure the specific labeling.

9.1.3. *In vitro* proximity labeling

Most proximity labeling assays are based on the expression of a recombinant biotinylation enzyme in living cells. This poses significant challenges for poorly transfectable cell types, non-model organisms, and clinical samples as PL workflows typically require complex and time-consuming CRISPR-based genetic modifications. Additionally, fusing a PL enzyme to a bait protein may disrupt its function or disturb its correct localization. Besides, careful experimental design and optimization of fusion protein expression levels are required to minimize background labeling.

RNA-centric PL approaches, such as those presented by Mukherjee et al. (2019) and Ramanathan et al. (2018), come with another major challenge. These strategies, further detailed in section 4.3, include a tagged RNA and an expressed labeling enzyme, fused to an affinity moiety that tethers it to the RNA. This results in substantial amounts of unbound or mislocalized biotinylation enzymes, leading to nonspecific background labeling. Moreover, these methods require either lengthy CRISPR/ Cas9 tagging or the simultaneous overexpression of multiple components, making them impractical for many cell lines, organisms, and clinical applications.

As an alternative, *in vitro* hybridization-based PL has emerged as a promising solution [64], [122], [123], [124]. These methods rely on a fixation and permeabilization protocol that preserves subcellular organization during probe hybridization and labeling, but enables efficient isolation of biotinylated proteins and RNAs afterwards. Compared to live-cell PL, *in vitro* approaches avoid problems with cytotoxicity, as labeling occurs in fixed and permeabilized cells rather than in living systems. Although fixation may slightly affect protein complex formation or localization, it eliminates the need for labor intensive genetic perturbations, thereby broadening the applicability of PL to a wider range of biological systems. These features make *in vitro* PL a valuable alternative for studying protein and RNA interactions in otherwise challenging biological contexts.

The first *in vitro* RNA PL assay was presented by Yap et al. in 2021 [64]. The authors designed, expressed and purified the 'HyPro' enzyme, short for 'hybridization-proximity labeling' consisting of a bacterial codon-optimized APEX2 version fused by a linker to the DIG10.3 domain. The DIG10.3 domain was designed *in silico* and binds digoxigenin (DIG) with a subnanomolar affinity [125]. Yap and colleagues next took advantage of digoxigenin-labeled antisense probes, which they hybridized against different RNA molecules of interest in chemically fixed and permeabilized cells. The HyPro enzyme is then attached to the RNA via binding of the DIG domain to the DIG-labeled probes. To avoid background labeling, unbound HyPro is removed in a washing step. Upon addition of the APEX substrates, biotin-phenol and hydrogen peroxide, proteins and RNAs in physical proximity to the RNA bait are labeled *in situ* (see Figure 18(A)). Following crosslink reversal, biotinylated proteins and RNAs can be isolated and enriched with streptavidin beads and analyzed by mass spectrometry or RNA sequencing, respectively (Figure 18(B)) [64], [66].
A similar approach was published in 2024 by Tsue et al., though their method differs in the tethering strategy of the labeling enzyme and the enzyme itself [124]. They used HRP instead of APEX, which was directly fused to an oligonucleotide. The sequence of the oligonucleotide was complementary to the overhang sequence attached to the primary probe targeted to the RNA of interest. The overall workflow, however, was very similar to the one previously described (see Figure 18(B)).

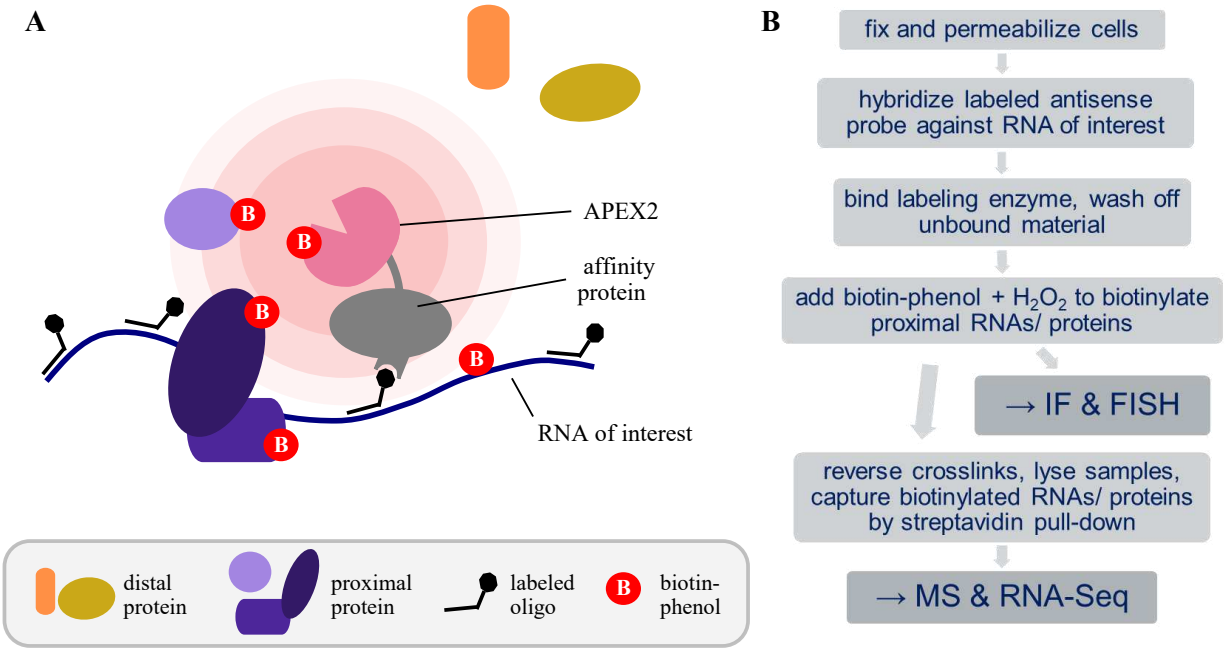


Figure 18.: Simplified workflow of the *in vitro* RNA proximity labeling assay to identify interactors. (A) Schematic of the key principle of an *in vitro* RNA PL assay. First, the labeling enzyme (APEX2 in this case) fused to an affinity domain is expressed and purified. Then, labeled oligonucleotides are hybridized to an RNA of interest in fixed and permeabilized cells. Next, the APEX2 fusion construct is bound to the oligonucleotides by the affinity protein, biotin-phenol and H_2O_2 are added and the proteins/ RNAs proximal to the RNA of interest (depicted in blue/ purple) are covalently labeled. Afterwards, the cross-linking is reversed, cells are lysed, the biotinylated material is enriched by a streptavidin pull-down followed by mass spectrometry or RNA sequencing. **(B)** The workflow presented as flow chart. As quality control step, IF and FISH can be done after biotinylation.

9.2. Results

The primary objective of this study was to identify RNA-binding proteins that interact with lncRNAs by establishing an *in vitro* proximity labeling assay for lncRNAs, as this class of RNA remains relatively understudied. Identifying their binding partners provides valuable details to better understand their function.

To achieve this, in a first trial the RNA-BioID methods was adapted. However, as this was unsuccessful (see previous chapter 8.2), an alternative method is needed. The assay described by Yap et al. (2021) was employed because they only used nuclear localized ncRNA and the aim hereof is to apply this assay to cytoplasmic lncRNAs like NORAD and OIP5-AS1. As an initial step, the method was established using NEAT1 as a benchmark, a well-characterized lncRNA (introduced above in section 9.1.1) that was also used in the original study as a proof of principle. The published protocol from Yap et al. was followed [66], beginning with the expression and purification of the HyPro enzyme in *E. coli*. The authors argued that the compact design of the enzyme with 47.8 kDa allows better access to dense molecular environments in fixed samples and

provides a smaller proximity-labeling radius compared to bulkier antibody-based biotinylation systems, such as those described by Bar et al. (2018) [122]. Next, the purified enzyme was validated *in vitro* by assessing its DIG-binding capacity and peroxidase activity. This was followed by verification of probe specificity and enzyme localization, with immunofluorescence-based biotinylation serving as a functional readout. Once the system is successfully established, the streptavidin pull-down will be optimized, enabling mass spectrometry-based identification of enriched proteins.
Together, these steps aimed to establish the *in vitro* lncRNA-BioID assay and facilitate the identification of the interactomes of the selected lncRNA candidates.

9.2.1. Expression, purification and validation of the labeling enzyme

As first step to establish the *in vitro* proximity labeling assay, the recombinant HyPro was expressed in *E. coli* in a soluble form, and its peroxidase and digoxigenin-binding activities were validated *in vitro*.
The expression of the protein was done by Orit Hermesh (following [66]), and purification was carried out using a HisTrap column followed by size-exclusion chromatography. SDS-PAGE analysis of the peak fractions from this final purification step confirmed a clean and soluble end product at the expected molecular weight, with no visible aggregates or degradation products (see Figure 19(A).
To assess the full enzymatic function of the purified HyPro protein, first its APEX2 peroxidase activity was evaluated *in vitro*. Therefore, purified HyPro was mixed with reconstituted enhanced chemiluminescence (ECL) reagent. This was incubated for one minute at room temperature, followed by spotting onto a filter paper for immediate imaging. A parallel assay was conducted using bovine serum albumin (BSA) as a negative control. The results demonstrated a clear peroxidase activity, as chemiluminescence was detected in the HyPro samples but not in the BSA control. Furthermore, increasing amounts of HyPro correlated with stronger signal intensity, confirming that the enzyme has its expected catalytic function Figure 19(B). This assay was performed by Orit Hermesh.
Next, the digoxigenin-binding activity of HyPro was assessed using another spot assay. Oligonucleotides tiling the lncRNA NEAT1 were labeled at their 3'-end with digoxigenin using the DIG Oligonucleotide 3'-End Labeling Kit. A dilution series of the labeled probes were spotted onto a nitrocellulose membrane. As negative control, unlabeled NEAT1 oligonucleotides were included, while pre-labeled probes provided by the kit and control-labeled oligonucleotides, that were labeled in parallel with the NEAT1 probes, served as positive controls. The probes were cross-linked to the membrane using a UV crosslinker, followed by blocking with BSA and incubation with HyPro (diluted 1:1000) for one hour. After washing, the membrane was developed using an ECL substrate. For comparison, a parallel experiment was conducted using an anti-DIG and a secondary HRP-coupled antibody, instead of HyPro. The results showed that HyPro specifically binds to digoxigenin-labeled oligonucleotides while retaining its peroxidase activity (Figure 19(C)). Notably, self-labeled NEAT1 probes (upper row) produced the strongest signal, with a clear dilution effect visible at least on the membrane analyzed with the anti-DIG

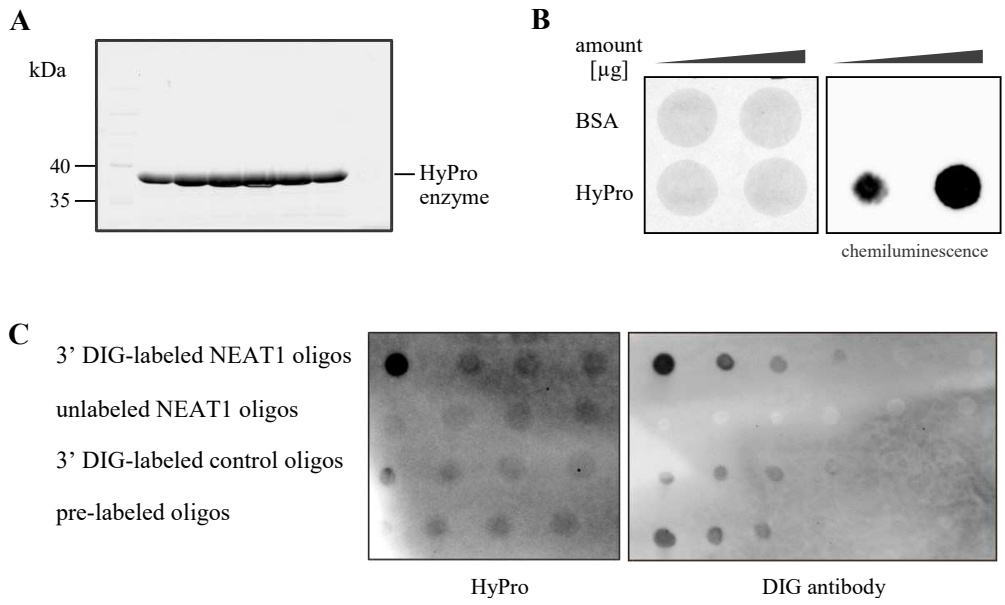


Figure 19.: HyPro protein expression, purification and functionality tests. (A) The six main peak protein fractions after size-exclusion chromatography showed clean single bands with the expected size on the SDS-PAGE gel stained with Coomassie blue. Assay performed by Orit Hermesh. **(B)** The APEX2 peroxidase activity of the purified protein was validated and showed clear chemiluminescence signal (lower row), BSA was used as negative control. Assay performed by Orit Hermesh. **(C)** A spot assay was used to demonstrate that the HyPro enzyme can bind digoxigenin-labeled oligonucleotides and simultaneously keep its function as peroxidase. Unlabeled oligonucleotides were used as negative control, pre-labeled and control-labeled oligonucleotides served as positive control.

antibody, while unlabeled probes exhibited almost no background.
In summary, HyPro was successfully expressed and purified, demonstrating both peroxidase and digoxigenin-binding activity *in vitro*.

9.2.2. Fluorescence microscopy analysis of HyPro-FISH/ IF

In order to establish an *in vitro* proximity labeling assay, besides a functional enzyme, specific probes are important to guide the enzyme to the target RNA of interest.
First, HeLa cells were treated with the crosslinking reagent dithio-bis(succinimidyl propionate) (DSP) and subsequently permeabilized with 70% ethanol (EtOH). This fixation and permeabilization protocol was suggested by the authors of the publication [66] and effectively preserves cellular morphology, as evidenced by the normal appearance of DAPI-stained nuclei and the well-maintained compartmental organization.
To ensure correct localization of the digoxigenin (DIG)-labeled probes targeting the lncRNA NEAT1 (listed in table Table 12), they were hybridized with the cells, followed by immun-

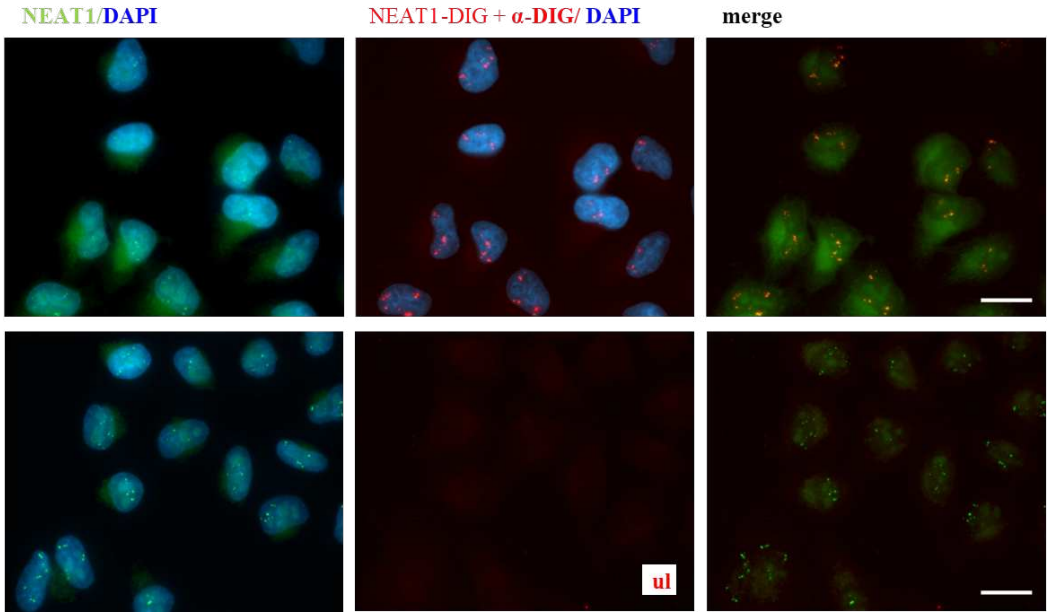


Figure 20.: Functionality and specificity test of 3' DIG-labeled NEAT1 probes. To validate the 3' DIG-labeled NEAT1 probes, they were mixed with directly labeled NEAT1 Atto488 probes (green signals) and hybridized with fixed and permeabilized HeLa cells. NEAT1-DIG probes were then detected by an anti-DIG antibody, using anti-mouse Alexa594 as secondary antibody (red signals). The lower panel was treated similarly, except that unlabeled probes (ul) were used as negative control. NEAT1-DIG probes show specific co-localization with NEAT1-Atto488 probes (upper panel). DNA in cell nuclei was stained with DAPI (blue signals). All images are maximum intensity z-stacks. Scale bar = 20 µm.

odetection using an anti-DIG primary antibody and a fluorophore-labeled secondary antibody. To assess probe specificity, the same cells were co-stained with previously validated NEAT1 probes which have the same sequence as NEAT1-DIG probes, but are conjugated to the fluorophore Atto488 (probes were validated by and obtained from Daniel Hofacker, Stafforst group at IFIB). Microscopy analysis revealed complete co-localization of the DIG-labeled probes with the Atto488-labeled probes within nuclear paraspeckles (see Figure 20), confirming that the DIG probes correctly recognize NEAT1 transcripts. As a negative control, unlabeled NEAT1 probes were incubated with the anti-DIG antibody and its fluorophore-conjugated secondary antibody (see Figure 20, lower panel). These samples showed only weak background signal, indicating that the antibody does not bind non-specifically to cellular structures.
In conclusion, the DIG-labeled probes demonstrated specificity for NEAT1, as evidenced by their precise localization within paraspeckles and their complete co-localization with previously validated probes. This confirms that the DIG probes are suitable for further experiments.

After confirming the correct localization of the DIG-labeled probes, the next step was to determine whether the HyPro enzyme binds to the DIG moiety of the labeled NEAT1 probes once they are hybridized to the NEAT1 transcript. To assess correct enzyme localization, three different concentrations of NEAT1-DIG probes were tested, and the results are shown in Figure 21 (A). At the lowest probe concentration, HyPro, detected using an anti-His antibody,

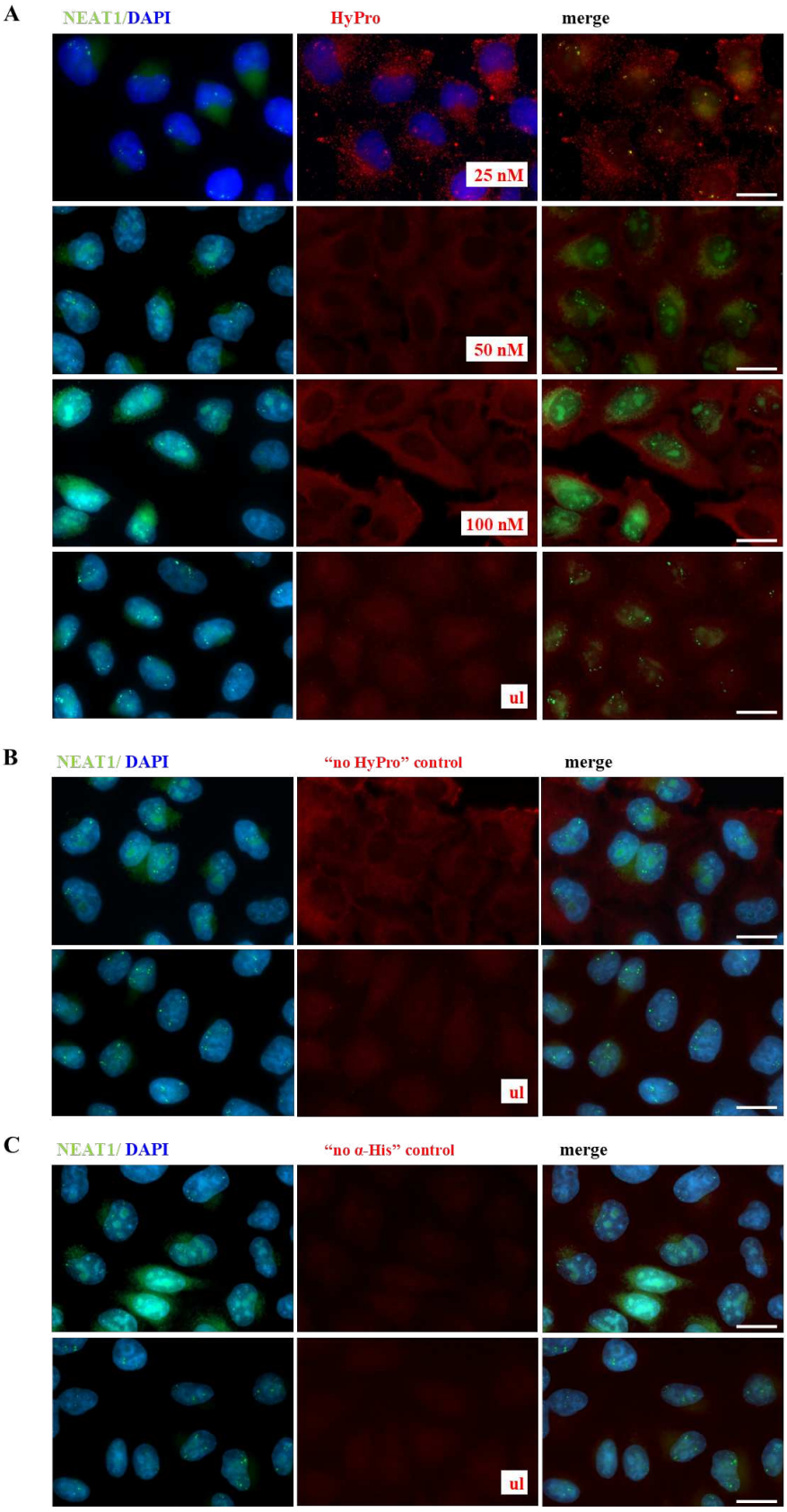


Figure 21.: Microscopy analysis of HyPro enzyme localization.

Figure 21.: (Continued from previous page.) **(A)** Three different NEAT1-DIG probe concentration were tested, starting with the lowest (25 nM) to the highest (100 nM). The last panel shows the negative control with unlabeled NEAT1 oligonucleotides (ul), treated with the same antibodies. Only the lowest NEAT1-DIG concentration showed co-localization with NEAT1-Atto488 probes (green signals). **(B)** A "no HyPro" control was conducted and cells were stained with NEAT1-DIG probes, anti-His and anti-mouse Alexa594 (upper panel) or NEAT1 unlabeled probes (ul) anti-His and anti-mouse Alexa594 as negative control. With or without the HyPro enzyme, NEAT1-DIG probes with the anti-His antibody seem to result in a pattern in the cytoplasm. **(C)** For the "no anti-His" control, cells were incubated with NEAT1-DIG probes, HyPro and anti-mouse Alexa594 or unlabeled NEAT1 probes (ul, lower panel) for the negative control. Without the anti-His antibody no spots or pattern appear, and only a distributed unspecific background from secondary antibody can be seen. DNA in cell nuclei was stained with DAPI (blue signals). All images are maximum intensity z-stacks. Scale bar = 20 µm.

co-localized with directly fluorophore-labeled NEAT1-Atto488 probes in nuclear paraspeckles, indicating correct localization. However, a noticeable background signal in the form of speckles was also present in the cytoplasm. In contrast, at medium and high probe concentrations, the anti-His signal was predominantly cytoplasmic, with complete nuclear exclusion of the detected signal. The reason for this altered distribution remains unclear. A negative control using unlabeled NEAT1 probes showed no nuclear signal, while a diffuse, spotty background pattern of the anti-His/anti-mouse Alexa594 staining was observed throughout the cell (Figure 21(A), lowest panel).

To investigate the specificity of the anti-His antibody, a "No HyPro" control experiment was performed (see Figure 21(B)). In these samples, cells were stained with NEAT1-DIG probes and anti-His antibody, but no HyPro enzyme was bound to the probes. The resulting images revealed nuclear exclusion of the signal and a cytoplasmic pattern similar to that seen at medium and high NEAT1-DIG probe concentrations in (A). Likewise, when unlabeled NEAT1 probes were used without HyPro, but anti-His antibody, a diffuse, spotty signal was detected across the cell (lower panel), closely resembling the background seen in the previously described negative control. These results suggest that, regardless of HyPro presence, the combination of NEAT1-DIG probes with the anti-His antibody produces a cytoplasmic pattern, indicating potential non-specific binding.

To determine whether the secondary antibody (anti-mouse Alexa594) contributed to the observed background, a "No anti-His" control was conducted (see Figure 21(C)). Here, cells were stained without the primary anti-His antibody. In the absence of anti-His, no abnormal cytoplasmic pattern was observed, neither with NEAT1-DIG probes nor with unlabeled probes (lower panel). Some background fluorescence from the secondary antibody was present, but was evenly distributed across the cells, confirming that the anti-His antibody was responsible for the non-specific cytoplasmic signal.

Despite the high background, the analysis at 25 nM NEAT1-DIG probe concentration demonstrated that the HyPro enzyme, under this condition, correctly localizes to nuclear paraspeckles. These results suggest that while non-specific binding of the anti-His antibody affects overall signal interpretation, the enzyme still follows the probes under optimized conditions.

To provide a comparison and additional controls, two alternative enzymes were included in this study. SNAP- and Halo-APEX2 each utilize a different binding chemistry than the HyPro

enzyme [126] [127], (reviewed in [128],) but share APEX2 as the labeling enzyme. They were expressed and purified by Daniel Hofacker (Stafforst group at IFIB). First, their localization was assessed to determine whether they correctly and specifically follow the NEAT1 probes to the nuclear paraspeckles.

As before, HeLa cells were fixed and permeabilized. The localization of SNAP-APEX2 was tested using an anti-His antibody, together with a co-staining with NEAT1-Atto488 probes (see Figure Figure 22(A)). Clear co-localization of SNAP-APEX2 with the NEAT1-Atto488 signal was observed, indicating correct localization within paraspeckles. As a negative control, unlabeled NEAT1 probes were used (lower panel), which resulted in no detectable nuclear signal. However, some background staining and a spotty cytoplasmic pattern were present, likely due to non-specific antibody interactions.
Similarly, Halo-APEX2 localization was tested (see Figure 22(B)). Detection with the anti-His antibody revealed partial co-localization with NEAT1-Atto488 probes, though background staining and scattered spots were also visible in the red channel. In the negative control, where unlabeled NEAT1 probes were used, no co-localization was observed, but a diffuse, spotty background signal was again distributed throughout the cell.
Overall, both SNAP-APEX2 and Halo-APEX2 demonstrated correct localization to nuclear paraspeckles. Given these results, they will be used as control enzymes in subsequent experiments.

Having confirmed that the HyPro enzyme localizes correctly to nuclear paraspeckles at 25 nM probe concentration, the next step is to assess whether *in vitro* biotinylation works and occurs at the expected cellular site. To this end, DSP/EtOH-treated HeLa cells were hybridized with DIG-labeled NEAT1 probes, incubated with HyPro, washed, and subsequently exposed to the APEX2 substrates biotin-phenol and hydrogen peroxide. Biotinylation efficiency and localization were then evaluated by staining the cells with fluorescently labeled streptavidin (Figure 23).
Several biotinylation conditions were tested. Besides the original described conditions [66], biotinylation was performed with high HyPro/ high biotin-phenol/ high NEAT1-DIG probe concentration. However, the results remained consistent across all conditions: strong biotinylation signals were detected in bright speckles predominantly located outside the nucleus, with no apparent co-localization with the NEAT1-Atto488 signal (Figure 23(A)).
To further investigate this unexpected biotinylation pattern, multiple controls were performed (Figure 23(B)). As positive control, HeLa cells were infused with HyPro, and without a washing step it was proceeded with the biotinylation reaction. This condition resulted in bright red speckles distributed throughout the cytoplasm (without overlap with the NEAT1 signal). These findings indicate that HyPro-mediated biotinylation occurs and the enzyme is active. Additional controls were included to assess background signal and antibody specificity. In a "no HyPro" control, the enzyme was omitted, and NEAT1-DIG probes were incubated with the streptavidin-Alexa594 antibody. This resulted in low background staining but no detectable speckles, confirming that the observed signals in the experimental conditions were dependent on the presence of HyPro. Finally, a negative control using unlabeled NEAT1 probes (lowest panel) showed some background signal and scattered speckles, similar to previous observations

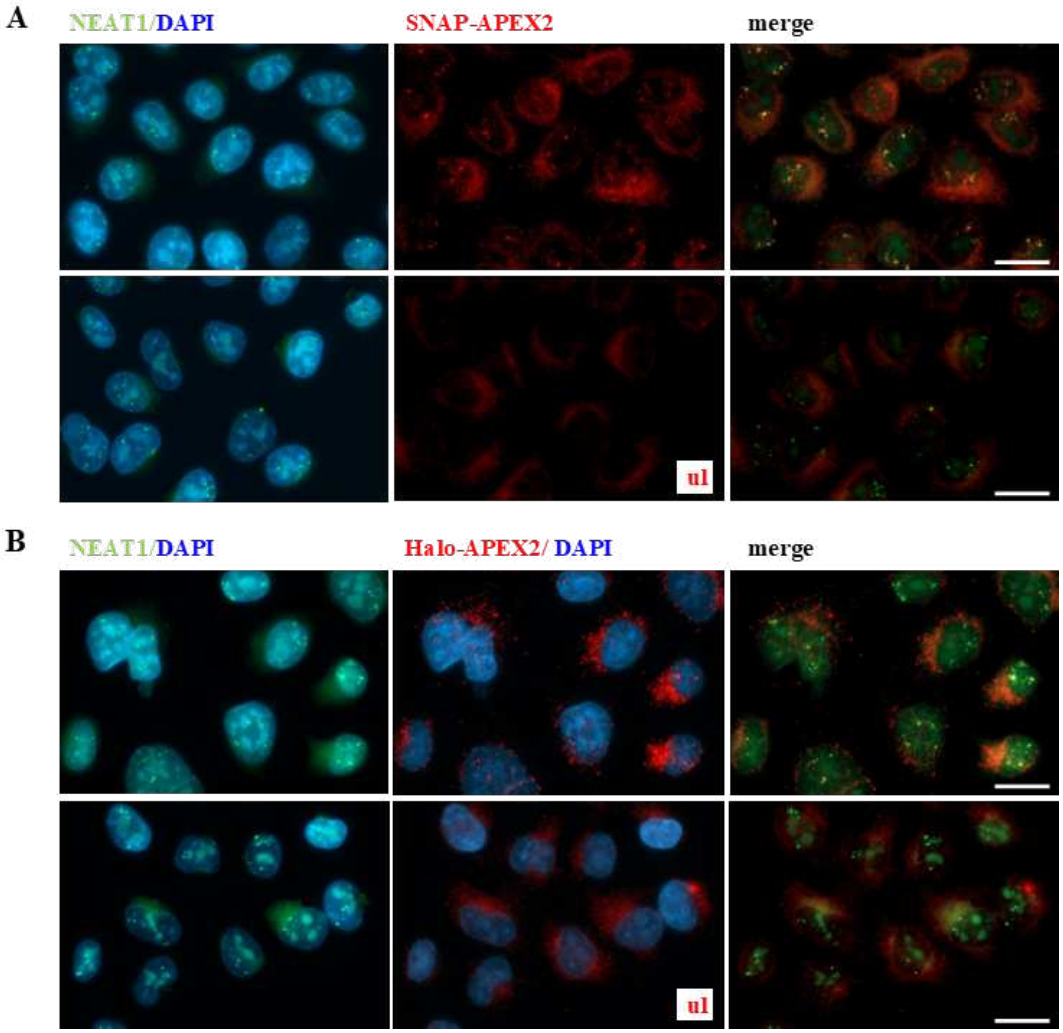


Figure 22.: Microscopy analysis of SNAP- and Halo-APEX2 enzyme localization. (A) The SNAP-APEX2 localization was visualized with an anti-His antibody and checked for co-localization with NEAT1-Atto488 (green signals). SNAP-APEX2 correctly localizes to paraspeckles and shows co-localization, while the negative control uses unlabeled NEAT1 oligonucleotides (ul) and shows no red spots inside the nucleus but some background in the cytoplasm. **(B)** The Halo-APEX2 enzyme localization was assessed in the same way as described for (A). Co-localization was not as strong as for SNAP-APEX2, and some background and spots in the red channel can be observed again for the negative control, but no co-localization with the unlabeled probes (ul). DNA in cell nuclei was stained with DAPI (blue signals). All images are maximum intensity z-stacks. Scale bar = 20 µm.

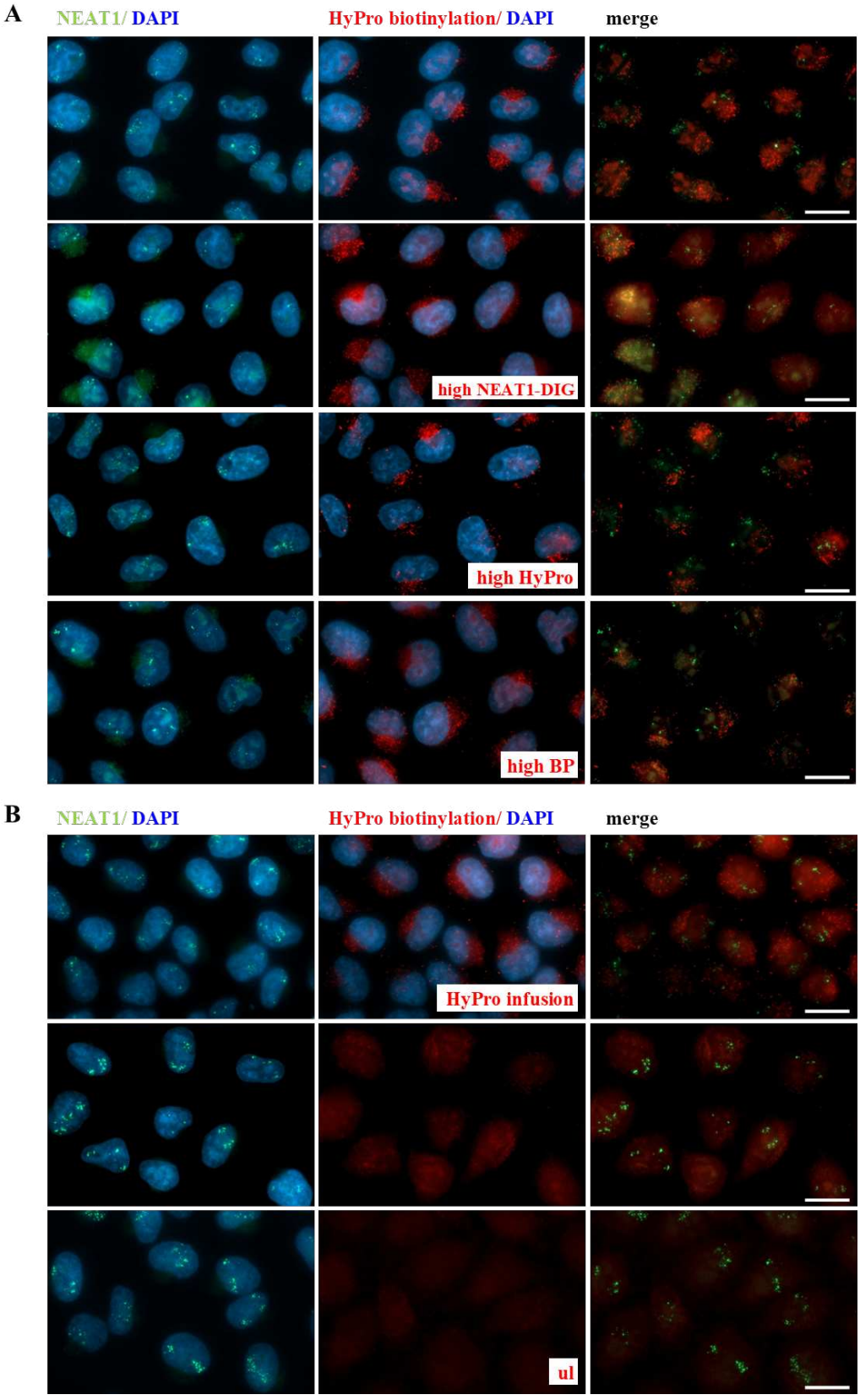


Figure 23.: Microscopy analysis of *in vitro* biotinylation by HyPro.

Figure 23.: (Continued from previous page.) **(A)** Different biotinylation conditions were tested in fixed and permeabilized HeLa cells. The staining with Streptavidin-Alexa594 looked similar for all samples, with biotinylation not co-localizing with NEAT1-Atto488 probes (green signals). **(B)** Biotinylation controls. HyPro infusion served as positive control which resulted in equal biotinylation over the cells. Below is shown the negative control without HyPro, showing background signal from the fluorescently labeled streptavidin. The negative control with unlabeled NEAT1 probes (ul, lowest panel) shows also some background signal and scattered speckles. DNA in cell nuclei was stained with DAPI (blue signals). All images are maximum intensity z-stacks. Scale bar = 20 µm.

with unlabeled probes in earlier experiments.
Taken together, these results indicate that, while HyPro exhibits biotinylation activity, the labeling does not specifically co-localize with NEAT1. Why the biotinylation does not localize to the target RNA as expected is unclear, as the probes were shown to localizes to paraspeckles under the right conditions. One suggestion is that the enzyme may not be efficiently targeted and/ or stably bound to the hybridized probes under the tested conditions. Further optimization will be required to improve specificity and ensure proper RNA-directed biotinylation. This could include an improved permeabilization protocol, more or different probes and an alternative binding chemistry.

To further evaluate and compare the *in vitro* biotinylation efficiency of HyPro, the control enzymes SNAP-APEX2 and Halo-APEX2 were tested. These enzymes had previously been established as controls, and their localization to nuclear paraspeckles was confirmed. The goal was to assess whether they could successfully mediate biotinylation at the expected sites.
The results of SNAP-APEX2-mediated biotinylation are shown in Figure 24(A). Fluorescently labeled streptavidin staining revealed a biotinylation signal that co-localized with NEAT1-Atto488 probes, indicating that biotinylation occurred at or near paraspeckles. The enzyme appeared highly active, efficiently depositing biotin at the target sites. A small number of biotinylated spots did not co-localize with NEAT1, possibly reflecting minor off-target labeling. The negative control, in which unlabeled NEAT1 probes were used, showed only a few scattered red spots, but these appeared weak and diffuse.
In contrast, Halo-APEX2-mediated biotinylation produced a markedly different pattern (Figure 24(B)). Here, the "biotinylation signal" was entirely excluded from the nucleus, showing no co-localization with NEAT1-Atto488 probes. This pattern was distinct from SNAP-APEX2 and the expected paraspeckle localization. The negative control, using unlabeled NEAT1 probes, produced a similar distribution of signal as with the labeled probes, suggesting that the observed biotinylation was not specifically directed to the NEAT1 target.
Taken together, these results demonstrate that while SNAP-APEX2 successfully deposits biotin at nuclear paraspeckles, Halo-APEX2 fails to do so, instead exhibiting a nuclear exclusion pattern. These findings highlight the differences in enzyme activity and localization, confirming that SNAP-APEX2 is a suitable control for RNA-directed biotinylation in this system.

The hybridization-proximity labeling approach demonstrated that DIG-labeled probes correctly localize to nuclear paraspeckles, as evidenced by their co-localization with NEAT1-Atto488 probes. The HyPro enzyme was shown to be catalytically active and also co-localized with

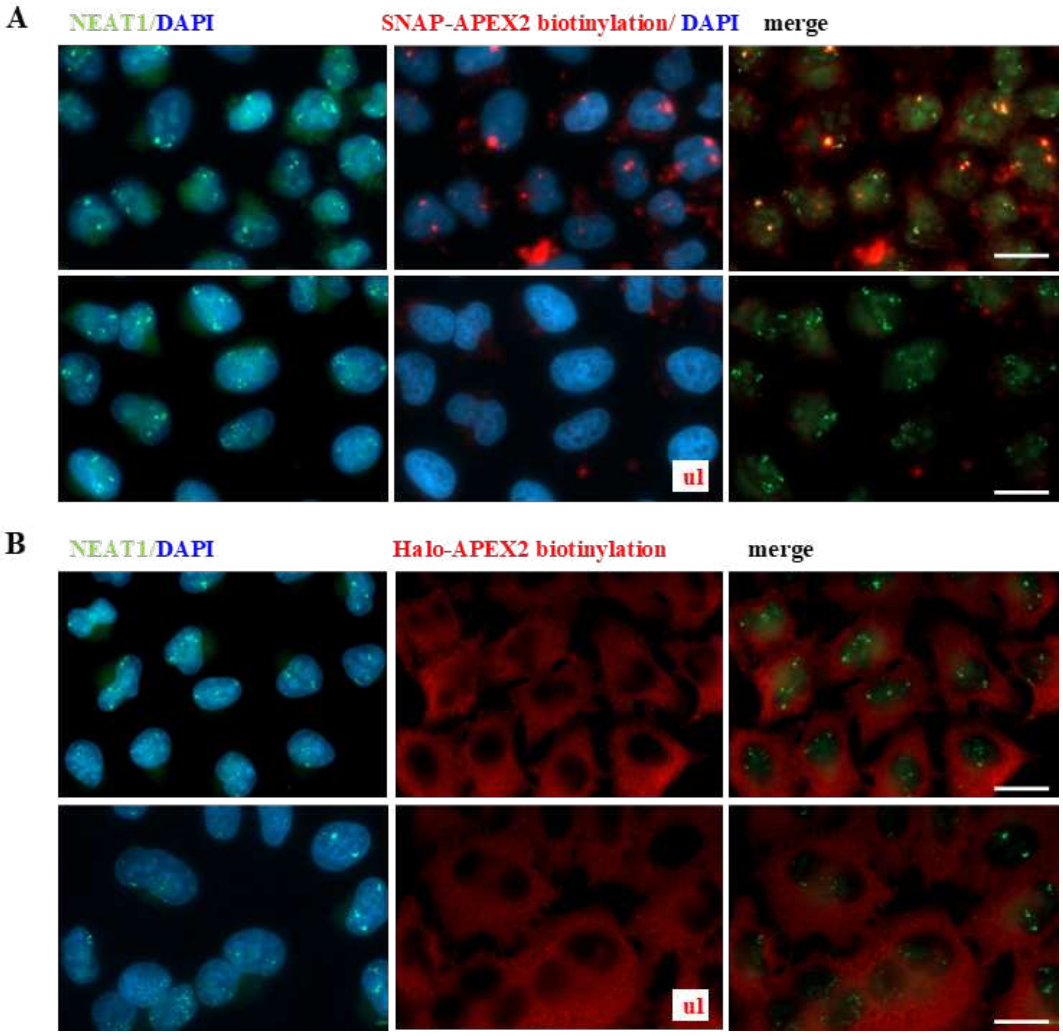


Figure 24.: ***In vitro*** **biotinylation using SNAP-APEX2 and Halo-APEX2. (A)** Biotinylation signal of SNAP-APEX2 co-localized with NEAT1-Atto488 (green signals), while the negative control with unlabeled NEAT1 probes (ul) shows only scattered and blurry red patterns. **(B)** With Halo-APEX2 the biotinylation signal showed no overlap with NEAT1-Atto488 signal but can be observed only in the cytoplasm of cells. The negative control appeared the same (ul). DNA in cell nuclei was stained with DAPI (blue signals). All images are maximum intensity z-stacks. Scale bar = 20 µm.

NEAT1 probes under certain conditions. However, despite its proper localization, HyPro-mediated biotinylation did not co-localize with NEAT1. In contrast, SNAP-APEX2 successfully directed biotinylation to paraspeckles, highlighting key differences in enzyme performance. These findings suggest that further optimization is required for HyPro to achieve RNA-directed biotinylation in this system.

10 Research project 3: β-actin interactome analysis

10.1. Introduction

10.1.1. (m)RNA localization within cells

RNA localization is a fundamental mechanism that fine-tunes cellular physiology in response to intracellular and environmental cues. A local enrichment or gradient of RNA transcripts within the cell can be achieved by local protection from degradation, diffusion-coupled local entrapment or by active and directed transport.
As introduced in section 2.2, mRNAs are the most researched RNA class, and consequently, their localization is best understood. By targeting mRNAs to specific subcellular compartments, cells create local translation hotspots. This process enables precise spatial and temporal control of gene expression and is conserved across both prokaryotic and eukaryotic cells [129].
The regulation of gene expression through mRNA localization offers several advantages. Firstly, as mentioned, translation can be regulated at a local level in response to stimuli, bypassing the need for transcriptional activation in the nucleus and subsequent transport of proteins. Thereby, the response to environmental clues is much faster. Second, localized mRNAs can be translated repeatedly in order to produce multiple protein copies at the site of need, which is more efficient than synthesizing proteins elsewhere and transporting them individually. Additionally, local translation can help to generate cell polarity by restricting protein synthesis to specific compartments [130].
The main influence on the final localization of mRNAs have the so-called "zipcodes," cis-acting sequences found within the 3' untranslated region (UTR) or, although less frequently, in the 5' UTR or coding sequence. These localization elements often form secondary structures, such as stem loops, that are recognized by RNA-binding proteins (see section 3.2 for an overview). These proteins bind to the zipcode sequence, forming ribonucleoprotein complexes (RNPs) that

direct the mRNA to its destination.
mRNA localization to subcellular sites involves several steps. The assembly of RNPs begins in the nucleus, where RNA-binding proteins (RBPs) associate with pre-mRNAs during processing. Once in the cytoplasm, RNAs and RBPs are often incorporated into RNA granules, which also contain ribosomal subunits, translation factors, decay enzymes, helicases, and scaffold proteins. Within these granules, mRNAs are translationally repressed until they reach their destination. Some RNPs are transported along microtubule or actin networks by motor proteins, such as kinesins, dyneins, and myosins. Upon reaching their final destination, mRNAs are typically anchored in an actin-dependent manner, ensuring localized translation [4], [131].

10.1.2. The β-actin mRNA

β-actin mRNA is one of the best-studied examples of localized mRNAs and serves as a model for understanding the mechanisms of mRNA localization. The β-actin mRNA is translated into the β-isoform of the cytoskeleton protein actin, which polymerizes into filaments. These actin filaments, or microfilaments, are key components of the cytoskeleton and play a central role in processes like cell adhesion, directional migration, and intracellular transport [132].
After export from the nucleus, β-actin mRNA is transported via an active, directed mechanism. Once localized, β-actin mRNA is translated, producing actin protein. This spatially restricted translation results in a high local concentration of actin, activating nucleation and leading to actin polymerization. Depending on the cell type, cell polarity, directional motility, and protrusion dynamics are influenced and regulated via this mechanism of spatially restricted gene expression [133], [134].
In fibroblasts, β-actin mRNA is localized near the leading edge of the cell, proximal to the lamellipodia (see Figure 25).

This localization is dependent on the β-actin zipcode, a conserved sequence in the 3' UTR of the mRNA [133]. The zipcode is a cis-acting localization element recognized by the RNA-binding protein ZBP1 (Zipcode Binding Protein 1, also known as IGF2BP1). ZBP1 is required for transport, localization, and translational control of β-actin mRNA [135]. Disruption of the zipcode or ZBP1 function leads to impaired mRNA localization and reduced cell motility.
In neurons, β-actin mRNA localizes to the growth cones of developing neurites, where its localized translation drives growth cone motility and directionality. Like in fibroblasts, the delocalization of the mRNA in neurons led to abnormalities such as growth cone retraction and impaired guidance [135].

10.1.3. The β-actin mRNA interactome

Studying the interactome of β-actin mRNA, including the associated transport mechanisms of the RNA-RBP complexes, is essential to fully understand how the localization contributes to

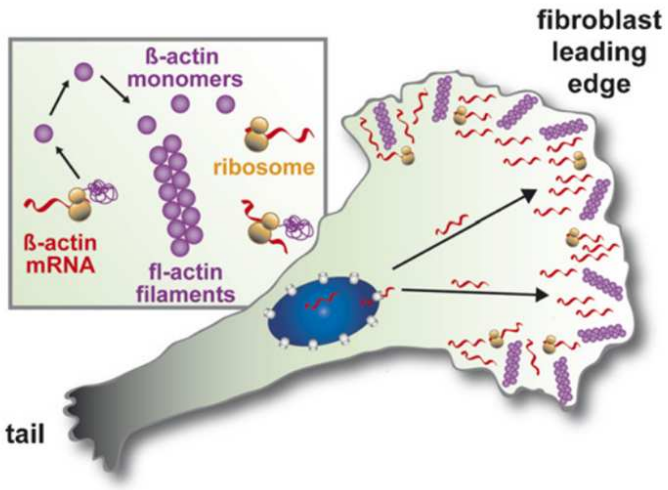


Figure 25.: Transport and translation of the β-actin mRNA in fibroblasts. The β-Actin mRNA is transported to the leading edge of fibroblasts, where it is translated and the actin proteins are assembled into actin filaments. From [130].

cell motility and spatially regulated gene expression in different cell types.
The application of the RNA-BioID assay in mouse embryonic fibroblasts (MEFs) yielded an unbiased β-actin interactome during serum-induced transport [1]. The data analysis identified both known and previously unknown RNA-associated proteins. Naturally, the publication from Mukherjee et al. raised many new follow-up questions, among them: When and during which process of its life cycle is β-actin mRNA bound by the interactor? In which region of the RNA does the RBP bind to β-actin mRNA? How does the binding of the RBP influence β-actin mRNA transport, or what is the effect of the binding of the RBP to β-actin mRNA? And lastly, how do β-actin interactors themselves interact, to which other proteins or in which complexes do they bind?
In the next sections, four candidates discovered by Mukherjee et al. (2019) are shortly introduced in order to further investigate their binding to the β-actin mRNA in the following.

10.1.3.1. The RBP heterogeneous nuclear ribonucleoprotein K (hnRNP K)

Heterogeneous nuclear ribonucleoprotein K (hnRNP K) is a 65 kDa RNA-binding protein that belongs to the hnRNP family (see section 3.1). It is widely expressed across human cell types and localized to the nucleus and cytoplasm, but also found in mitochondria [136]. hnRNP K plays diverse roles in transcription, translation, RNA splicing, and chromatin remodeling, and is essential for various cellular processes [20].
Structurally, hnRNP K contains three highly conserved KH domains, which bind RNA and DNA with strong specificity for poly(C)-rich sequences such as 5'-dTCCC. In addition to its KH domains, hnRNP K possesses a K-interactive (KI) region, which mediates interactions with other proteins, and a nuclear shuttling domain (KNS), allowing bidirectional transport across

the nuclear envelope [136].
Among its many functional roles, hnRNP K also interacts with the TATA-binding protein (TBP) to positively regulate gene expression [137]. It binds constitutively and inducibly to promoters and transcribed regions, as shown by ChIP analyses [138]. In neurons, hnRNP K is essential for axonal growth, neuronal development and synaptic maturation. Its knockdown impairs axonal outgrowth and reduces synaptic contacts. Additionally, hnRNP K regulates the expression of mRNAs encoding microtubule- and neurofilament-associated proteins, coordinating cytoskeletal organization critical for axon formation, polarity, and proper neuronal connectivity. hnRNP K also interacts with actin-regulatory proteins, such as N-WASP, to regulate actin polymerization and cytoskeletal organization. Those interactions lead to a repression of filopodia and protrusion formation of cells while promoting cytoskeletal remodeling required for synaptic development [139].
Beyond its role in cytoskeletal regulation, hnRNP K interacts with various mRNAs and lncRNAs, including the lncRNA NEAT1, where the binding influences paraspeckle formation and lincRNA-p21, where the cooperation inhibits p53-responsive genes [136].
Clinically, hnRNP K functions as both, tumor suppressor and oncogene, depending on the context. Complete loss of hnRNP K is embryonic lethal, while haploinsufficient mice exhibit severe developmental defects and reduced survival. Dysregulation of hnRNP K, including overexpression and cytoplasmic mislocalization, is associated with cancer progression and poor prognosis, underscoring its relevance in human disease [18], [137].

10.1.3.2. The RBP TAR DNA-binding protein 43 kDa (TDP-43)

Transactive response (TAR) DNA-binding protein 43 kDa (TDP-43) is an ubiquitously expressed member of the hnRNP family, known for its diverse roles in RNA metabolism. TDP-43 contains two RNA recognition motifs (RRMs) and a glycine-rich C-terminal region. It is predominantly localized in the nucleus, where it participates in RNA transcription, pre-mRNA splicing, mRNA transport, and microRNA processing. With both, nuclear localization (NLS) and nuclear export signal (NES), TDP-43 shuttles between the nucleus and cytoplasm, suggesting a role in mRNA transport [140].
TDP-43 associates with specific mRNAs, including β-actin and CaMKIIα [140]. Using iCLIP, it was confirmed that TDP-43 binds preferentially to clusters of UG-rich sequences *in vivo* [141]. Within motor neurons, TDP-43 is found in highly mobile RNA granules that are actively transported along axons. These granules co-localize with other RNA-binding proteins, such as FMR1, ZBP1, and HuD, which are known to regulate mRNA transport and local translation. Similar to FMR1 and ZBP1, TDP-43 may coordinate the transport and localized translation of mRNAs in response to neuronal stimuli, potentially playing a role in dendritic mRNA regulation. Pointing to this is the fact that mutant TDP-43 disrupts this process, impairing the transport of mRNP granules and leading to defects in mRNA localization [142], [143]. Additionally, loss of TDP-43 is toxic to cells, leading to embryonic lethality, motor dysfunction, and increased apoptosis. In motor neurons, TDP-43 depletion results in dysmorphic nuclear shapes, disruption of the cell cycle, and reduced viability [143].

Clinically, TDP-43 is strongly implicated in neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD). Mutations in the TARDBP gene, which encodes TDP-43, account for approximately 2-5% of ALS cases. TDP-43 is also the major component of ubiquitinated cytoplasmic, nuclear, and neuritic inclusions in ALS and FTLD brains, underscoring its critical role in motor neuron degeneration [144], [145].

10.1.3.3. The RBP Polypyrimidine tract-binding protein 1 (PTBP1)

Polypyrimidine tract-binding protein 1 (PTBP1, also known as hnRNP I) is a widely expressed RNA-binding protein and a member of the heterogeneous nuclear ribonucleoprotein (hnRNP) family. It regulates numerous post-transcriptional processes, including splicing, localization, translation and mRNA stability [146]. PTBP1 has a molecular weight of 56.5 kDa and possesses an N-terminal nuclear localization signal (NLS) and nuclear export signal (NES), which allow it to shuttle between the nucleus and cytoplasm. It also contains four RNA recognition motifs (RRMs), connected by flexible linker regions [21]. The RRMs of PTBP1 preferentially bind to pyrimidine-rich sequences, with optimal motifs such as UCUUC, UUCU, or CUCU [147].
In the nucleus, PTBP1 is a well-characterized regulator of alternative splicing. It binds polypyrimidine tracts near the 3' splice sites of pre-mRNAs, repressing exon inclusion and promoting exon skipping [148], [149]. In neurons, PTBP1 is expressed only during the embryonic stages, where it regulates alternative splicing, while its neuronal paralog nPTB performs similar functions in both embryonic and adult neurons [150].
Furthermore, PTBP1 contributes to cytoplasmic mRNA metabolism in various ways. It regulates mRNA stability, 3'-end processing, and cap-independent translation through internal ribosome entry sites (IRESs). Notably, PTBP1 binds several mammalian mRNAs, including GAP-43, Tau, CaMKIIα, and β-actin, highlighting its critical role in mRNA regulation [148].
Regarding the β-actin mRNA, PTBP1 was shown to interact with the zipcode in the 3' UTR [146]. This interaction facilitates β-actin mRNA transport to neurites, where localized translation supports neuronal polarity and axonal growth. Depletion of PTBP1 leads to reduced actin protein accumulation in neurite terminals, impairing actin dynamics and axon elongation [146].

10.1.3.4. The RBP heterogeneous nuclear ribonucleoprotein R (hnRNP R)

Heterogeneous nuclear ribonucleoprotein R (hnRNP R) belongs to the subfamily of hnRNPs (see section 3.1) and plays a role in pre-mRNA processing, RNA metabolism and transport. hnRNP R is a 65 kDa protein, containing three RNA recognition motifs, an acidic-binding domain (AcD), and an RGG (Arg-Gly-Gly) box, which enables interaction with other hnRNPs. The RNA-binding domains facilitate specific interactions with target RNAs, including β-actin mRNA. hnRNP R is primarily localized in the nucleus, as well as in the cytoplasm, axons, and axon terminals, where it plays a pivotal role in RNA transport and local translation [151].

hnRNP R is essential for the axonal localization of β-actin mRNA, which depends on its interaction with U-rich sequences within the 3' UTR of the transcript [152]. hnRNP R and β-actin mRNA co-localize in axons and growth cones, and this localization is critical for proper axon elongation and presynaptic differentiation. Although hnRNP R repression does not alter β-actin mRNA levels, knockdown of hnRNP R in motor neurons significantly reduces axon growth, while dendrite growth and neuronal survival remain unaffected. Similarly, morpholino-mediated knockdown of hnRNP R in zebrafish embryos leads to defects in axonal extension, as well as reduced spontaneous movement, further emphasizing its role in neuronal development [153].
The role of hnRNP R in β-actin mRNA localization is closely tied to its interaction with the survival motor neuron (SMN) protein. hnRNP R binds SMN in motor neuron cytosol, and their co-localization is observed in axons and axon terminals [152]. While SMN does not bind β-actin mRNA directly, the SMN–hnRNP R complex facilitates β-actin mRNA transport to axons and growth cones, where localized translation supports axon elongation. In SMN-deficient motor neurons, reduced β-actin mRNA and protein levels in distal axons correlate with impaired axon growth.
These observations have a clinical impact, as in a subtype of frontotemporal lobar degeneration (FTLD), hnRNP R and hnRNP Q (with which hnRNP R shares a high level of sequence homology) were identified in neuronal cytoplasmic and intranuclear inclusions in the frontal cortex and hippocampus of FTLD patients. Additionally, an increased expression of hnRNP R is detected in several FTLD disease groups [154].

10.2. Results

The primary goal of this part of the study was to prove and further analyze the interactions between β-actin mRNA and four (potential) RNA-binding candidates (introduced above in 10.1.3) previously identified by Mukherjee et al. in 2019 [1] in the RNA-BioID screen in MEF cells. As stated above, for some of these candidates interaction with β-actin mRNA was reported in neurons, but not for MEFs.
To achieve this goal, a protocol for RNA co-immunoprecipitation (Co-IP) with MEF cell lysate was established. The Co-IP experiment usually included two control conditions to validate the specificity of the interactions: a bead-only control (magnetic Protein G Dynabeads incubated with lysate), to control for unspecific binding to the magnetic Protein G Dynabeads and an antibody control (beads incubated with antibody), to control for efficient antibody-bead coupling. In the case of hnRNP R, no specific antibody was available, so hnRNP R was expressed as a fusion protein with GFP and GFP-coupled agarose beads were used for the pull-down. Here, untransfected lysate served as a negative control. The general workflow is described in the following. First, antibody-coated beads were incubated with MEF cell lysate to allow protein complexes to bind. After incubation, supernatant samples were taken for Western blot analysis, and the beads were washed to remove non-specific interactions. Samples were collected after the first and last wash step. Interacting proteins were then eluted by boiling the beads.

Elution samples, along with supernatant and wash samples, were separated by SDS-PAGE and transferred to a PVDF membrane. The membrane was probed with the respective primary antibody, and visualized via a secondary antibody and chemiluminescence. Input samples served as positive control to confirm the presence of the target protein in the lysate.
For RNA interaction analysis, after successful Co-IP, the protocol was performed again under RNase-free conditions to enrich for RNA-protein interactions. Pulled-down RNAs were isolated and converted to cDNA via reverse transcription. RT-qPCR, including a negative and positive control, was performed to confirm the association of β-actin mRNA with the target protein. This approach aimed to validate β-actin mRNA-protein interactions and provide insights into their biological relevance.

10.2.1. Co-Immunoprecipitation of hnRNP K

To investigate the interaction between hnRNP K and β-actin mRNA, Co-Immunoprecipitation (Co-IP) was performed using an anti-hnRNP K antibody and MEF cell lysate. Initial attempts to immunoprecipitate hnRNP K were unsuccessful, likely due to insufficient cell lysate and/or inefficient antibody coupling to the beads.
To optimize the protocol, additional antibody and cell lysate were used, and four different conditions were tested: (1) a bead control sample (2) an antibody control sample, with only beads and antibody (3) 3 µg of anti-hnRNP K with lysate sample; and (4) 5 µg of anti-hnRNP K with lysate sample.
After coupling the antibody to magnetic Protein G Dynabeads in sample (2)-(4), supernatant samples were taken. Following incubation of the antibody-coupled beads of sample (3) and (4) with 600 µg MEF cell lysate, additional supernatant samples were collected. The beads were then washed three times with NT2 buffer, with samples collected after the first wash. Finally, elution samples from the beads and all previously collected samples were analyzed via SDS-PAGE and Western blot (see Figure 26).
The input sample (first lane in both blots, Figure 26), serving as a positive control, confirms the presence of hnRNP K in the MEF lysate, with a strong band at ∼65 kDa. No detectable heavy or light antibody chain bands can be observed after the antibody to bead coupling step, indicating that antibody binding to the beads was successful (SN1 samples, left blot). However, analysis of the supernatant and first wash step samples revealed that only a small fraction of hnRNP K was captured, as evidenced by weak bands at 65 kDa in the elution samples. This suggests that although the antibody to bead coupling worked, large amounts of hnRNP K could not be bound by the antibody and some was washed away, so the enrichment of hnRNP K was limited. The antibody control and α-hnRNP K with lysate approaches show weak bands at ∼55 kDa in the elution, likely referring to the heavy chain of the IgG antibody, again confirming successful antibody-to-bead coupling.
To increase and improve the enrichment, a modified protocol was tested in which the lysate was incubated with the antibody before bead addition. Despite this adjustment, substantial amounts of hnRNP K remained unbound in the supernatant, and the elution samples showed no improvement. Further optimizations, including adjustments of incubation conditions and

reagent concentrations, also failed to enhance enrichment.

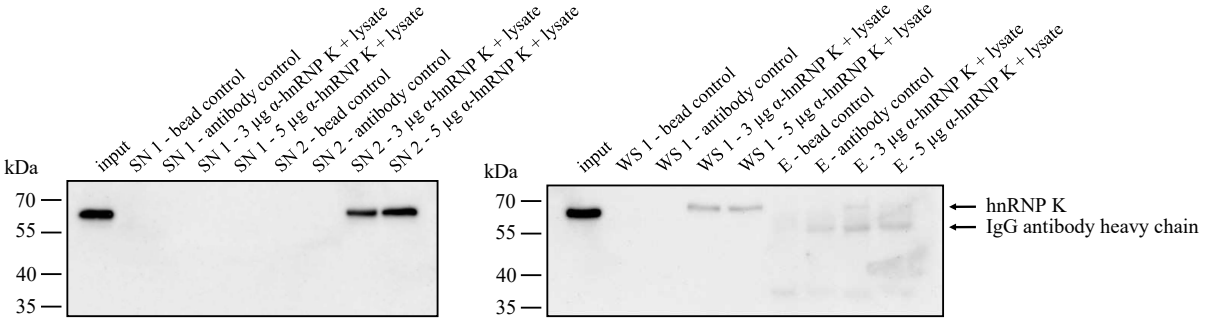


Figure 26.: Western blot analysis of Co-Immunoprecipitation of hnRNP K with α-hnRNP K. Co-Immunoprecipitation was performed including two controls (bead and antibody control). Two different antibody amounts were tested (3 µg and 5 µg), which were incubated with MEF cell lysate each. Both input samples show a strong band at ∼65 kDa, confirming the presence of hnRNP K in the lysate. The supernatant 2 and first wash step samples of the α-hnRNP K with lysate approaches show bands at 65 kDa. The elution of these approaches resulted in weak bands at 65 kDa. SN = supernatant, WS = washing step, E = elution.

In conclusion, additional optimization or the use of a more specific antibody is necessary to reliably immunoprecipitate hnRNP K for downstream analyses.

10.2.2. Co-Immunoprecipitation of TDP-43

To study the interaction between TDP-43 and β-actin mRNA, Co-IP was performed using an anti-TDP-43 antibody with MEF cell lysate. Initial attempts were unsuccessful, likely due to an inefficient and nonspecific antibody.
To achieve a successful pull-down, a different TDP-43 antibody was used, and four conditions were tested: (1) a bead control, (2) an antibody control, (3) 5 µL of α-TDP-43 with lysate, and (4) 10 µL of α-TDP-43 with lysate. After coupling the antibody to magnetic Protein G Dynabeads in conditions (2)-(4), conditions (3) and (4) were incubated with 500 µg of MEF cell lysate, and supernatant samples were taken. The beads were then washed, and samples were collected after the first and third wash steps. Elution samples and all previously collected fractions were analyzed via SDS-PAGE and Western blot (see Figure 27).

The input sample (positive control) confirms the presence of TDP-43 in the MEF lysate, with a band at 43 kDa. The supernatant and first wash step samples of the α-TDP-43 with lysate approaches also show bands at 43 kDa whereby the bands of the wash step samples are weaker than the bands of the supernatant samples. This reveals that only a small fraction of TDP-43 was captured by the antibody-bound beads. The strong bands in the supernatant indicate that most of the TDP-43 protein remained unbound, while weak bands at 43 kDa in the first wash step suggest partial loss during washing. The elution samples show only minimal enrichment

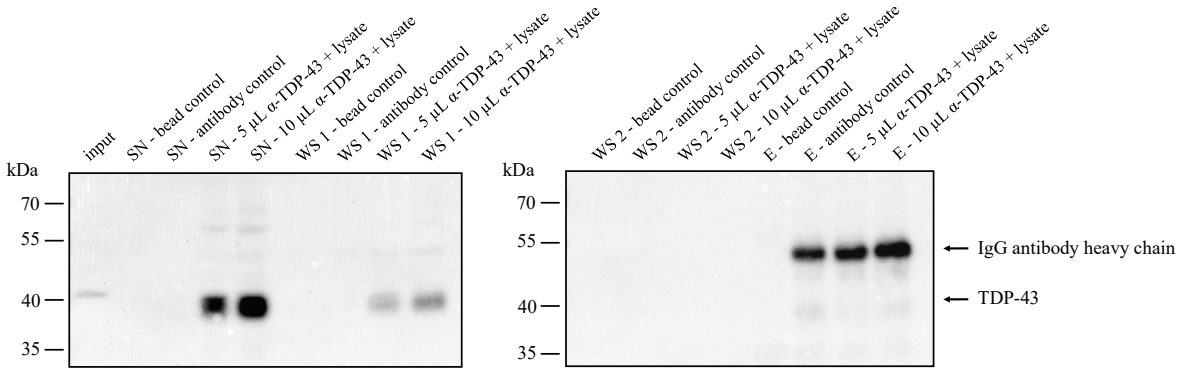
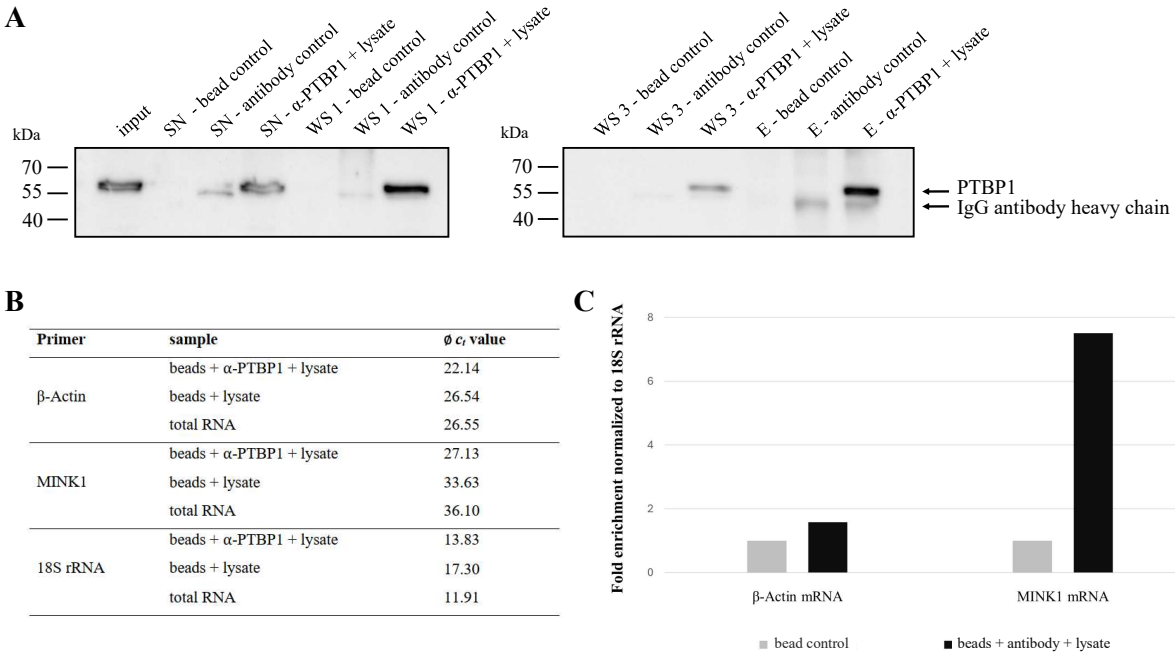


Figure 27.: Western blot analysis of Co-Immunoprecipitation of TDP-43 with α-TDP-43. Co-Immunoprecipitation was performed including a bead and antibody control and two different antibody amounts (5 µL and 10 µL). All samples, including an input sample, were analyzed by SDS-PAGE and Western blot. The input sample shows a weak band at the expected size of 43 kDa. The elution of the α-TDP-43 with lysate approaches resulted in very faint bands at ∼43 kDa. SN = supernatant, WS = washing step, E = elution.

of TDP-43, indicating inefficient pull-down.
To improve the yield, modified protocols were tested. Despite adjustments, including varying incubation conditions and reagent concentrations, a significant amount of TDP-43 remained unbound in the supernatant, and the elution sample showed no successful enrichment.
In summary, while the Co-IP demonstrated that TDP-43 could be pulled down, the enrichment was minimal, and further optimization is required to achieve reliable and robust results.

10.2.3. Co-Immunoprecipitation of PTBP1 and analysis of associated RNAs by RT-qPCR

To examine the interaction between PTBP1 and β-actin mRNA, Co-Immunoprecipitation was performed using MEF cell lysate and an anti-PTBP1 antibody. The initial trial successfully isolated PTBP1, as evidenced by a specific band at ∼56 kDa in the elution sample. Building on this result, a second Co-IP experiment was conducted, doubling the amount of magnetic Protein G Dynabeads to allow for both Western blot analysis and subsequent RNA isolation for RT-qPCR (see Figure 28).
For the second Co-IP, three conditions were tested: (1) a bead control, (2) an antibody control and (3) the experimental condition (Dynabeads with 5 µg α-PTBP1 and 3 mg MEF lysate). After antibody coupling and incubation with the lysate, supernatant samples were taken from each condition. The beads were then washed three times, after the first and last wash step, samples from each approach were taken. All samples from the supernatant, wash steps, and eluates were analyzed by SDS-PAGE and Western blot. The input sample served as a positive control, confirming the presence of PTBP1 in the MEF lysate with a strong band at ∼56 kDa (Figure 28(A), first lane). The supernatant and the samples from the first wash step of the antibody control show weak bands at 55 kDa which match the heavy chain of the IgG antibody.



B

Primer	sample	ø c_t value
β-Actin	beads + α-PTBP1 + lysate	22.14
	beads + lysate	26.54
	total RNA	26.55
MINK1	beads + α-PTBP1 + lysate	27.13
	beads + lysate	33.63
	total RNA	36.10
18S rRNA	beads + α-PTBP1 + lysate	13.83
	beads + lysate	17.30
	total RNA	11.91

Figure 28.: RNA Co-Immunoprecipitation of PTBP1. **(A)** Western blot analysis of Co-Immunoprecipitation of PTBP1. Co-Immunoprecipitation was performed with two controls (bead only and bead + antibody). All samples, including an input samples, were analyzed by SDS-PAGE and Western blot. The input sample served as a positive control, confirming the presence of PTBP1 in the lysate with a band at ∼56 kDa. The elution sample also shows a clear band at 56 kDa, indicating a successful pull-down of PTBP1. SN = supernatant, WS = washing step, E = elution. **(B)** and **(C)** show the RT-qPCR analysis of co-immunoprecipitated RNA with PTBP1. Co-immunoprecipitated RNA with PTBP1 was isolated from the beads and reverse transcribed into cDNA. Total RNA was isolated as well and used as control. The abundance of each RNA in all three samples was assessed by RT-qPCR with RNA-specific primers, performed by Melike Basbuga. **(B)** The detected average c_t values of the RT-qPCR +RT samples with β-Actin, MINK1 (positive control), and 18S rRNA (reference) are shown. **(C)** Bar graph with calculated fold enrichment, ribosomal 18S rRNA was used as reference. In the approach with α-PTBP1, MINK1 mRNA was more enriched than β-actin mRNA. However, with α-PTBP1 more β-actin mRNA was enriched than in the approach without α-PTBP1.

This indicates that a minor amount of the antibody is washed off at every step. Similarly, in lanes of the supernatant and wash step samples of the α-PTBP1 with lysate sample, bands at 56 kDa can be recognized, indicating that not the complete amount of PTBP1 is bound by the antibody-coupled beads and some is washed off. However, in the last wash step, no band is visible in the antibody control sample and the PTBP1 band is drastically weaker in the antibody + lysate sample, confirming that three wash steps were sufficient. While the elution sample shows a distinct band at 56 kDa, indicating successful pull-down of PTBP1, weaker bands corresponding to the IgG heavy and light chains can also be observed in the eluate. Nevertheless, the Co-IP effectively enriched PTBP1.

To investigate whether PTBP1 interacts with β-actin mRNA, RT-qPCR was performed. Since the co-immunoprecipitation of PTBP1 was successful, as can be concluded from the Western

blot results, RNA associated with PTBP1 after the pull-down and total RNA from MEFs was isolated and reverse transcribed into cDNA.
GAPDH mRNA, a housekeeping gene known not to interact with PTBP1 [1], was used as a negative control. According to Xue et al. (2009) PTBP1 acts as an alternative splicing regulator of Misshapen-like kinase 1 (MINK1) mRNA [155]. Therefore, MINK1 mRNA was used as a positive control. Ribosomal 18S rRNA was used as a reference RNA to normalize the obtained c_t values of the target RNAs. For each primer pair, technical replicates of +RT and one sample -RT as control were prepared. The -RT control served to ensure that only RNA, which was reverse transcribed into cDNA is used as template. Additionally, a water control was run for each primer pair.
The detected average c_t values (see Figure 28(B)) were analyzed following the description in subsection 7.6.8. The fold enrichment is shown in Figure 28(C). β-actin mRNA is 1.6 fold enriched in the α-PTBP1 sample over the control sample without antibody, supporting a specific interaction between PTBP1 and β-actin mRNA. The positive control revealed strong enrichment of MINK1 mRNA in the pull-down, confirming the specificity of the Co-IP. However, GAPDH mRNA showed similar accumulation to β-actin mRNA (data not shown), rendering it unsuitable as a negative control.
In summary, the Co-IP and RT-qPCR experiments provided evidence of a specific interaction between PTBP1 and β-actin mRNA, although the choice of negative control was ineligible and requires optimization to validate the results more robustly.

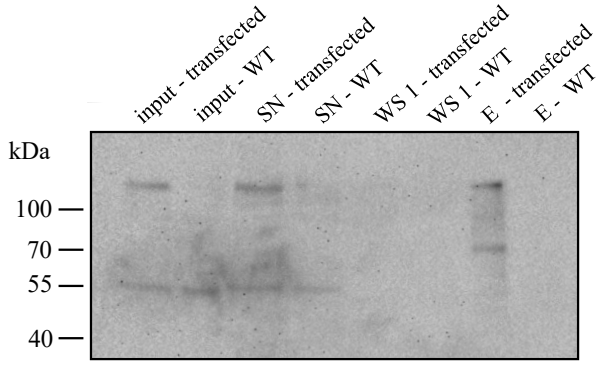
10.2.4. Co-Immunoprecipitation of hnRNP R and RT-qPCR analysis

To investigate the interaction between hnRNP R and β-actin mRNA, an affinity tag was used for immunoprecipitation due to the lack of a specific hnRNP R antibody. The hnRNP R coding sequence was amplified from the plasmid p2309 using primers 7261 and 7262. The PCR product was purified, digested, and ligated into the digested plasmid p2283, resulting in the successful construction of plasmid p2315.
A test transfection of MEF cells with plasmid p2315 was conducted to confirm the expression and correct localization of the hnRNP R fusion protein, which included a C-terminal HA and eGFP tag. Fluorescence microscopy verified the correct expression and proper folding - at least of the GFP part of the fusion protein - based on the observed GFP signal (data not shown).
For the pull-down experiment (see Figure 29 (A)), MEF cells were transfected on a larger scale with plasmid p2315, while untransfected MEF cells (WT) served as a negative control. In the input and supernatant samples from the transfected MEF cell lysates (lane 1 and 3), bands were detected above 100 kDa. Although the expected molecular weight of the fusion protein was 99.4 kDa, these bands can be attributed to the fusion protein, as no corresponding bands were observed in the input or supernatant sample from untransfected MEFs (lane 2 and 4). Additionally, a band at approximately 55 kDa was detected across lanes 1 to 4, likely due to non-specific binding of the primary anti-HA antibody to an endogenous protein. However, the absence of this band in the wash step samples indicates specificity of the pull-down. The elution sample from the transfected MEF cell lysate approach again shows a band above 100 kDa,

confirming successful pull-down of the fusion protein. The band size is consistent with that observed in the input sample (lane 1). A second band at 100 kDa in the elution sample may represent a degradation product of the fusion protein. In contrast, no bands were detected in the elution sample from the untransfected MEF lysate approach, indicating specificity of the anti-GFP nanobody.

Since pull-down of the hnRNP R fusion protein was successful, RNA associated with the hnRNP R fusion protein was isolated from the beads, along with total RNA from MEFs. To perform RT-qPCR, the RNA was reverse transcribed into cDNA. For each condition, technical duplicates (+RT) along with one -RT sample were prepared. Additionally, a water control was run for each primer pair.
To validate the specificity of the hnRNP R–RNA interaction, 7SK RNA, a non-coding RNA and known interactor of hnRNP R, served as a positive control [151]. GAPDH mRNA was used as a negative control and ribosomal 18S rRNA was used as a reference RNA to normalize the obtained c_t values. From the detected average c_t values of the RT-qPCR +RT samples (listed in Figure 29(B)), normalized Δc_t and fold enrichment (shown in Figure 29(C)) was calculated. GAPDH mRNA seems to bind unspecifically to beads and/ or the antibody and therefore cannot be used as a negative control.
The bar graph in Figure 29(C) reveals that the positive control 7SK ncRNA is highly enriched in the GFP-based hnRNP R pull-down, confirming the specificity of the assay. Comparison between transfected and untransfected samples show higher enrichment of RNA in the transfected samples, further validating the specificity of the anti-GFP nanobody coupled to magnetic agarose beads. Importantly, β-actin mRNA displayed a two fold enrichment, suggesting a specific interaction between hnRNP R and β-actin mRNA.
These results confirm that the established pull-down protocol enriched RNA interacting with hnRNP R and support the conclusion that β-actin mRNA associates with hnRNP R.

A



B

Primer	sample	∅ c_t value
β-Actin	transfected	21.03
	untransfected	25.46
	total RNA	27.32
7SK	transfected	19.97
	untransfected	27.47
	total RNA	25.54
18S rRNA	transfected	14.33
	untransfected	16.67
	total RNA	13.48

C

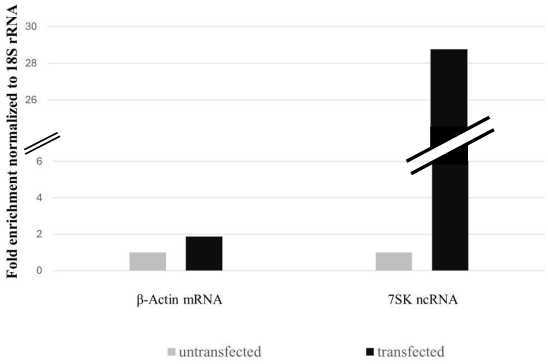


Figure 29.: RNA Co-Immunoprecipitation of hnRNP R. **(A)** Western blot analysis of Co-Immunoprecipitation of hnRNP R, done by Melike Basbuga. Co-Immunoprecipitation was performed under two different conditions: GFP-coupled magnetic agarose beads were incubated with untransfected (control) or transfected MEF cell lysate. The transfected cells were expressing hnRNP R-HA-eGFP from plasmid p2315 for pull-down with α-GFP. All samples, including an input sample, were analyzed by SDS-PAGE and Western blot. The input and supernatant sample of the transfected MEF cell lysate in lane 1 and 3 show a band >100 kDa, although a molecular weight of ∼ 100 kDa was expected for the fusion protein. The elution sample from transfected MEFs also shows a band over 100 kDa, indicating that the pull-down was successful. The elution sample of the approach with untransfected MEF cell lysate shows no bands and this indicates that the anti-GFP nanobody is specific. SN = supernatant, WS = washing step, E = elution. **(B)** and **(C)** show the RT-qPCR analysis of co-immunoprecipitated RNA with hnRNP R, RT-qPCR was performed by Melike Basbuga. RNA pulled-down with the hnRNP R fusion protein was isolated from the beads and reverse transcribed into cDNA. The abundance of each RNA in all three samples was assessed by RT-qPCR with RNA-specific primers. **(B)** The detected average c_t values of the RT-qPCR +RT samples with β-Actin mRNA, 7SK ncRNA (positive control) and 18S rRNA (reference) are shown. **(C)** Bar graph with fold enrichment calculated from average values of two independent RT-qPCR experiments, ribosomal 18S rRNA was used as reference. Comparison between the transfected and untransfected samples show that the 7SK non-coding RNA is the most enriched. β-actin mRNA shows a two fold enrichment in transfected over untransfected samples.

11

Discussion and Outlook

11.1. Discussion of Project 1

This project aimed to identify RNA-binding proteins (RBPs) that interact with lncRNAs, focusing on the lncRNAs NORAD and OIP5-AS1 as candidates. As this class of RNA is very heterogeneous and for most members the cellular role is elusive, identifying binding partners should help shedding some light on their largely unexplored functions.
To achieve this, the plan was to adapt the RNA-BioID assay, originally developed by Mukherjee et al. (2019). The original study used a stable MEF cell line in which β-actin mRNA was tagged with 24xMS2 loops in its 3' UTR, from Lionnet et al. (2011), which was then transfected with a 2xMCP-eGFP-BirA fusion protein to facilitate proximity labeling.
Building on this approach, the experimental strategy in this thesis consisted of two main objectives to establish the RNA-BioID assay for lncRNAs. First, stable cell lines expressing either 2xMCP-eGFP-TurboID or 2xMCP-eGFP-microID should be generated, to evaluate the most effective biotinylation system. Second, tagging the endogenous loci of OIP5-AS1 and NORAD with 24xMS2 loops via the CRISPR/Cas9 system is aimed. Following successful integration, optimization of biotin labeling and streptavidin pull-down would be performed, leading to mass spectrometry analysis of enriched proteins.
By establishing this lncRNA-BioID assay, this study aimed to uncover the interactomes of OIP5-AS1 and NORAD, providing new insights into their molecular functions.

11.1.1. Proximity labeling

As the first step in establishing the lncRNA-BioID assay, a biotin ligase needed to be stably integrated into cells. Traditional PL enzymes, such as BioID, exhibit slow labeling kinetics, re-

quiring prolonged incubation with exogenous biotin to achieve sufficient labeling. To overcome these limitations, newer ligases have been developed, including TurboID [41] and microID [40]. To evaluate the performance of both enzymes for the need of this thesis, constructs expressing 2xMCP-eGFP-TurboID and 2xMCP-eGFP-microID were generated. A transient test transfection proved that both fusion proteins function as expected, so stable cell lines were created. Therefore, the HeLa-EM2-11ht cell line was chosen due to its already integrated Flp-in system which offers ease and robustness for stable transfection. The integration of both constructs was successful and monoclonal cell lines were created. Correct localization of the fusions constructs was assessed by live cell microscopy, showing that GFP signal was primarily observed in the nucleus as expected (Figure 8(C) and Figure 7(C)). Both biotin ligases were active and gave satisfying biotinylation signal when incubated with biotin as proven by Western blot analysis (Figure 8(B) and Figure 7(B)). Importantly, the doxycycline-inducible promoter showed tight regulation, preventing unwanted background expression but $\sim$ 90% expression upon addition of doxycycline as shown by flow cytometry (Figure 8(D) and Figure 7(D)).
One challenge with TurboID is its tendency for strong background labeling, even before exogenous biotin is added, likely due to its ability to utilize trace biotin present in standard media [41], [45]. This issue was already evident in transient experiments, which is why a doxycycline-inducible promoter was used for the stable cell lines. The time-course induction experiment (see Figure 9) revealed that an 8-hour induction resulted in sufficient expression for suitable biotinylation signals with less background in Western blot analysis. This seems a promising method to control the otherwise aberrant background labeling.
In contrast, microID did not exhibit background labeling but showed slower labeling kinetics compared to TurboID (2 hours vs. 10 minutes). While TurboID appears to be the more suitable enzyme for the lncRNA-BioID system, microID may be advantageous in cases where minimizing the fusion protein size and reducing steric hindrance are priorities.

Overall, both TurboID and microID provide significant advantages over traditional PL enzymes. TurboID offers significantly faster labeling kinetics, while microID provides a compact and efficient alternative.
The choice of PL enzymes obviously is much bigger (as described in section 4.1), and they all come with certain advantages as well as disadvantages. A recent publication by Schreiber et al. (2024) compared APEX2-TDP-43 versus miniTurbo-TDP-43 with each other [156]. While APEX2-TDP-43 identified about four times more high-confidence proximal interactors relative to miniTurbo-TDP-43, a greater proportion of these proteins were previously identified as TDP-43 interactors [156]. This once more underscores the importance of carefully choosing the PL enzyme considering its activity, the amino acid targeted for biotinylation and specific experimental conditions.

11.1.2. Split-enzymes as alternative labeling enzymes?

As described before, there is a variety of biotin ligases developed by now (see page 15). Additionally, split versions like split-BioID, split-TurboID and splitAirID were engineered [157],

[158], [159], [160]. These enzymes are originally biotin ligases, which were divided into two inactive fragments. When each half is fused to two interacting proteins of interest (POI), the split fragments reassemble to an active enzyme after the POIs come in proximity, leading to biotinylation only in the vicinity of the interacting POIs (depicted in Figure 30(A)). This leads to specific and locally constraint biotinylation, only in certain contexts when the two POIs interact. These split versions allow investigation of specific protein complexes, even in case one of the POI is present in several different complexes [158].
Establishing a split-BioID assay for RNAs would greatly reduce unspecific background labeling that occurs while the biotin ligase is not yet bound to the aptamer tagging the RNA of interest. Each half of the split enzyme could be fused to a MCP protein, alternatively one half could be fused to MCP and the second one attached to a known protein interactor in order to probe a subset of the RNA interactome (see Figure 30(D)).
As an initial test how well the split enzymes perform and whether this new development is promising, a new plasmid was cloned containing both split-TurboID halves, fused to the proteins FKBP (12-kDa FRB-binding protein) or FRB (FKBP-rapamycin-binding domain) that interact in the presence of rapamycin, under a doxycycline inducible promoter (p2363). The same plasmid, but instead of split-TurboID containing split-microID (unpublished split-version of microID [40]) was also tested for comparison (p2231). Both plasmid were transiently transfected and the biotinylation activity was assessed by Western blot (see Figure 30(B) and (C)) As positive control, the full length version of both enzymes was also transfected. Both split-enzymes were expressed upon the addition of doxycycline. While without exogenous biotin no biotinylation signal could be observed, which is an advantage compared to the full length TurboID, a certain degree of reconstitution of functional enzyme happened without rapamycin, as can be seen in the biotinylation activity when biotin was added. However the biotinylation signal got stronger when rapamycin was supplemented. Split-TurboID seems to be almost as active as the full length version, while split-microID is drastically less active. In comparison, the split-TurboID seems to perform better and is an interesting candidate for the realization of an RNA-splitBioID assay as depicted in Figure 30(D).

11.1.3. CRISPR

For the precise tagging of lncRNAs with 24xMS2 loops, the second step in order to establish a proximity labeling assay for lncRNAs, two main strategies can be considered: inserting the MS2 loops into the endogenous gene using CRISPR/Cas9 or introducing an exogenous vector that expresses the lncRNA with the desired tag. Both approaches have advantages and limitations. The use of a vector to express a tagged lncRNA construct can, in some cases, offer a faster alternative. However, this leads to overexpression of the lncRNA, because both the endogenous and vector-derived transcript are expressed simultaneously, leading to non-physiological levels of the RNA. Additionally, cloning large lncRNA constructs, such as OIP5-AS1 (8.8 kb) tagged with 24xMS2 loops (1.3 kb), is technically demanding due to the sheer size and repetitive nature of the sequence. Another challenge when cloning the construct is the placement of the MS2 loop array. Due to the existence of several splice variants and a rudimental annotation the

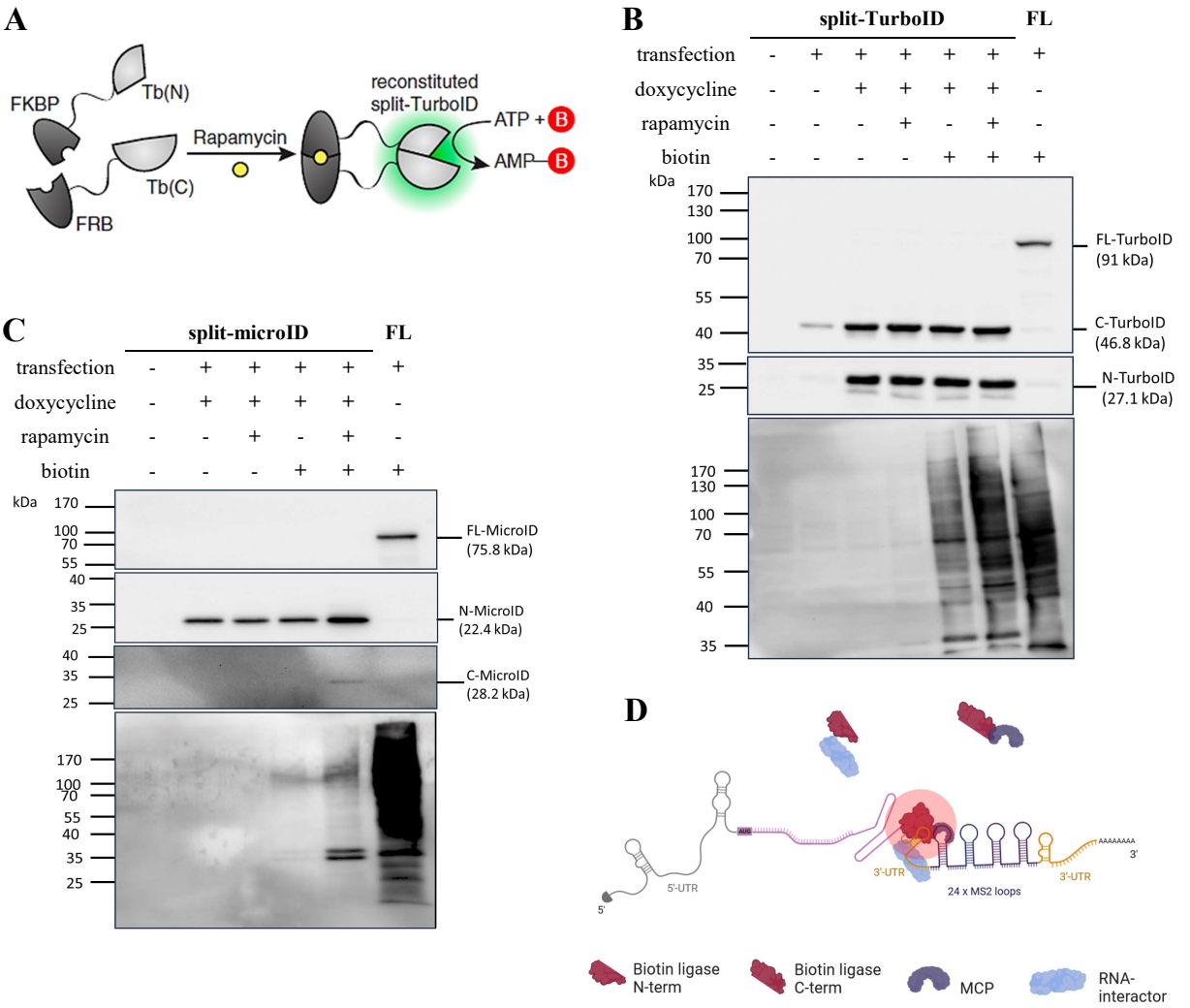


Figure 30.: SplitBioID (A) Schematic display of the split-TurboID assay. Each half is fused to either FKBP (FRB-binding protein) or FRB (FKBP-rapamycin-binding domain) that interact in the presence of rapamycin. Upon addition of rapamycin, the functional TurboID enzyme is reconstituted and biotinylates its surrounding. Adapted from [159]. **(B)** Split-TurboID (p2363) was transiently transfected. Upon addition of doxycycline, both split halves are expressed. The enzyme partly is reconstituted without rapamycin, as can be seen in the biotinylation activity when biotin is added. However, without biotin, but with rapamycin, no biotinylation can be detected. If both chemicals are substituted, the enzyme is nearly as active as the FL TurboID. Biotinylation time was 2 h 50 min. FL = full length **(C)** Split-microID (p2231) was transiently transfected. Upon addition of doxycycline, the N-terminal half was stably expressed, the C-terminal half was either less expressed or unstable. The enzyme is slightly active without rapamycin, as there is biotinylation activity when biotin is added. Without biotin, but with rapamycin, no biotinylation can be seen. If both chemicals are present, the enzyme does biotinylate, but is less active than the FL microID. Biotinylation time was 4 h. FL = full length **(D)** Simplified schematic display of a possible RNA-BioID assay with split-TurboID.

position of the 24xMS2 array risks disrupting critical secondary structures, functional motifs, or regulatory elements of the lncRNA.
In contrast, CRISPR/Cas9-mediated tagging directly at the endogenous locus provides a more elegant and physiologically relevant approach. By ensuring that the tagged lncRNA is expressed under its native promoter, natural expression levels and regulatory controls are preserved. Nevertheless, careful consideration is required when determining the insertion site for MS2 loops to avoid disrupting the lncRNA's secondary structure, functional motifs, or overall expression. Furthermore, the polyploid nature of HeLa cells can complicate the tagging of all alleles. Despite these challenges, CRISPR-based tagging circumvents the effects of additional lncRNA copies and should provide a more accurate representation of endogenous lncRNA biology.

In order to achieve the tagging of OIP5-AS1 and NORAD lncRNA with 24xMS2 loops, a CRISPR strategy was planned based on the publication of Cheng and Qiu, 2017. First, gRNAs for both lncRNAs were designed and tested. The most efficient one was combined with the slightly modified donor template (originally received from Cheng and Qiu, 2017) and transfected into the TurboID-expressing HeLa-EM2-11ht cell line that had already been established. In summary, both NHEJ-based CRISPR approaches, following the published protocol of Cheng and Qiu [117], to integrate MS2 loops into the OIP5-AS1 or NORAD gene failed, although for different reasons. In the case of OIP5-AS1, the first approach was too inefficient, and no clones survived the puromycin selection. The second attempt achieved partial success but resulted in the integration of only a fragment of the repair construct. As demonstrated by genotyping PCR, only the puromycin resistance cassette was integrated into the genome. Southern blot and smiFISH analysis confirmed that no MS2 loops were present in the puromycin resistant cell lines. This was likely due to degradation of the linearized repair plasmid by exonucleases inside the cell. Only cells that contained the part with the puromycin gene and promoter in their genome could survive the antibiotic selection (see Figure 12). For the NORAD lncRNA, the integration was successful at the first trial, but closer analysis of the cells revealed that the MS2 tag was not expressed as part of the NORAD lncRNA. This issue could not be solved by Cre-mediated recombination (CmR), which removes the backbone of the repair plasmid and leaves only the 24xMS2 loops integrated. After transient transfection of the Cre recombinase, the NORAD expression was disturbed and down-regulated (see Figure 14(C)), but still no MS2 expression could be observed by smiFISH.
The fact that the plasmid-based CRISPR-mediated integration of the MS2 tag succeeded for NORAD immediately, but failed for OIP5-AS1 is unexpected, especially given that both approaches used the same repair strategy and cell line. One possible explanation is differences in chromatin accessibility at the target loci. If one locus resides in a more open chromatin region, Cas9 and the repair template may have had better access compared to a more tightly packed heterochromatin genomic environment. Additionally, sequence-specific factors such as secondary structures within the genomic DNA or RNA transcripts might have influenced repair efficiency. Another consideration is differential expression or transcriptional activity of the lncRNAs. NORAD has a higher expression level compared to OIP5-AS1 and an actively transcribed locus may be more prone to disruption and NHEJ repair. This was also reflected by other authors, as Fueller et al. (2020) stated critically that their choice of highly expressed genes as convenient reporters to validate their CRISPR tagging method might bias conclu-

sions about the efficiency [115]. Further investigation into these factors, including chromatin accessibility assays and RNA stability measurements, could shed light on why CRISPR editing succeeded in one case but not the other.
Even with the achieved targeted integration of the MS2 loops into the NORAD genome, no NORAD-24xMS2 expression could be observed, neither before CmR nor afterwards. Therefore, a closer examination of the original publication by Cheng and Qiu (2017) was necessary. The title claims "Long non-coding RNA tagging and expression manipulation via CRISPR/Cas9-mediated targeted insertion" [117], yet nowhere in the paper do the authors provide evidence that their candidate lncRNAs were tagged and actually also expressed as tagged versions. Their main approach involved designing sgRNAs within the transcriptional termination sites of six different lncRNAs, co-transfecting CRISPR plasmids, and selecting for resistant cells following integration events.
To analyze the resulting cell lines, the authors measured lncRNA expression levels by RT-qPCR before and after Cre-mediated recombination (CmR). They reported that all six lncRNAs showed a significant upregulation, with at least two-fold increase in expression. Upon Cre transfection, which excised the plasmid backbone, expression levels partially decreased but did not fully normalize. From this, they concluded: "We successfully tagged lncRNAs." Notably, their so-called "tagging" was only confirmed at the genomic level via junction PCR and at the RNA level via RT-qPCR. However, they used exclusively gene-specific primers to perform RT-qPCR and measure expression level changes. Nowhere in their study do they provide any evidence that the MS2 loops were actually transcribed, except for one lncRNA by an indirect experiment. Without direct detection of the inserted tag at the RNA level, their conclusions remain entirely unsubstantiated.
From a purely technical standpoint, this study did manage to replicate their work — smiFISH confirmed disrupted NORAD expression following CmR. The assertion that this method enables "lncRNA tagging and expression manipulation" is highly misleading; at best, it achieves only expression manipulation [117].

Both NHEJ-based CRISPR approaches to integrate MS2 loops into the OIP5-AS1 and NORAD loci failed, albeit for different reasons. To overcome these issues, the microhomology-mediated end joining (MMEJ) repair pathway was explored as an alternative. MMEJ utilizes short microhomology arms (MHAs) flanking the Cas9-induced double-strand break (DSB) to facilitate precise DNA repair.
This new approach was based on a study by Chen et al. (2020), in which the authors knocked in an MS2 cassette at the distal end of the NEAT1 locus using CRISPR/Cas9-induced MMEJ. To select for successfully tagged cells and monitor lncRNA expression levels, an IRES-GFP sequence was inserted downstream of the MS2 loops, driven by the endogenous promoter. GFP-positive cells were then enriched via sorting before the authors went for clonal isolation and genotyping. In contrast to the above-discussed publication [117], this study demonstrated that their strategy effectively labeled the endogenous lncRNA NEAT1 without disrupting its physiological properties [118]. Given its successful application of CRISPR-based MMEJ to a lncRNA, this method was adapted for use in the present study.
A key feature of the strategy presented by Chen et al. is the stabilization of the linear DNA repair template, a problem that had been encountered in the previous CRISPR attempt. To

prevent rapid degradation, the 5' ends of primers were thiophosphorylated at their last five nucleotides. The repair template design consisted of 20-nt MHAs flanking the Cas9 cleavage site and a 25–30 nt overlap with a universal template plasmid. This ensures the creation of a protected and efficient linear repair template by PCR from the universal template plasmid [118].
Although Chen et al. successfully applied this method to lncRNA tagging, they were not the first to use MMEJ for precise insertions. Nakade et al. (2014) demonstrated that targeted integration could be achieved via very short microhomologies in both cultured cells and animals [114]. Interestingly, their study reported an observation that was frequently encountered in this work as well: while 5' junctions tended to be precisely repaired via microhomology sequences, 3' junctions were often joined imprecisely. This discrepancy is likely influenced by the selection strategy used. Since antibiotic selection excludes out-of-frame clones, inaccuracies at the 3' junction may not impact overall selection efficiency. This could explain why junction PCRs between the MS2 loops and the genome frequently failed in this study.

Despite initial hopes, the MMEJ-based approach adopted from Chen et al. (2020) ultimately failed to achieve the desired outcome. For both lncRNA candidates, PCR and Southern blotting confirmed the presence of MS2 loops in the genome. However, the integration did not occur at the intended OIP5-AS1 or NORAD loci.
While the MMEJ strategy successfully enabled MS2 loop insertion, it is likely that the CRISPR event was unspecific. Southern blot analysis revealed significant off-target integration and potential genomic rearrangements. Additionally, smiFISH detected no MS2 loop expression, rendering the generated clones unsuitable for further experiments.
To address these issues, a CRISPR approach with improved specificity is needed. Increasing the length of homology arms could enhance repair precision following DSB induction, thereby improving the likelihood of successful and site-specific integration.

As a final attempt to achieve precise MS2 loop integration, a strategy utilizing the homology-directed repair (HDR) pathway was employed. HDR, a less error-prone mechanism compared to NHEJ or MMEJ, was expected to facilitate accurate insertion of the 24xMS2 cassette into the OIP5-AS1/NORAD locus. Based on the approach suggested by Spille et al. (2019), a repair plasmid was designed with nearly 900 bp homology arms flanking a selectable blasticidin resistance gene and the 24xMS2 loops [60].
Despite employing HDR and donor templates with extended homology arms, integration of the 24xMS2 cassette into the OIP5-AS1 locus was unsuccessful. Although the MS2 array was incorporated into the genome of all analyzed clones, Southern blotting and genotyping PCRs revealed that integration did not occur at the intended locus. Unexpected band patterns in the Southern blot suggested off-target insertions, and smFISH analysis showed no detectable MS2 loop expression as part of the OIP5-AS1 transcript. The occurrence of off-target integrations even with HDR was unexpected, emphasizing the necessity of thoroughly validating CRISPR knock-in cell lines through multiple assays and experiments. Notably, off-target integration events may occur at sites not included in the predicted candidate list, highlighting the importance of whole-genome sequencing to ensure the complete absence of unintended modifications [112].

Following this failure, further CRISPR attempts at the NORAD locus were not pursued, and the project was ultimately discontinued. These results highlight the challenges of achieving precise lncRNA tagging and suggest that further optimization of the CRISPR approach is needed, potentially involving alternative gRNA designs or repair templates. Additionally, Spille et al. reported that insertion frequency or "CRISPR efficiency" is generally low ($< 1\%$) for inserts of 1 kb and above. The commonly used repair template here consisted of a CMV promoter ($\sim$ 500 nt) driving the expression of a resistance gene ($\sim$ 500 nt) and the 24xMS2 cassette ($\sim$ 1.3 kb) so overall more than 2 kb [60]. Further, the authors mentioned to obtain 50–100 distinct colonies per T25 flask after a CRISPR attempt. In this work, surviving cells typically grew into 10 - 20 small colonies per ⌀10 cm dish. This huge discrepancy likely already hints to poor CRISPR efficiency. [60]

To increase CRISPR efficiency and minimize indels in both on- and off-target sites due to error-prone repair after DNA DSB, recently it got more popular to use the partially inactivated Cas9 D10A nickase enzyme. This mutant variant of the original Cas9 introduces only single-strand nicks and according to Bollen et al. (2022) enables fast, efficient, and indel-free knock-ins [161]. This method was also applied by Yarnall et al. (2022), who used a CRISPR–Cas9 nickase fused to a reverse transcriptase and serine integrase. This fusion enzyme is capable of integrating inserts of several kb in a single delivery reaction with high efficiencies in cell lines [162]. However, as these promising methods were mostly targeted against highly expressed mRNAs (like the β-actin gene) so far, it is debatable whether it would work for lncRNAs as well.

11.2. Conclusion and Outlook of Project 1

Since the CRISPR-based tagging attempts failed in this project, the broader applicability of the RNA-BioID assay, at least in its current form, must be questioned. As a result, alternative strategies are needed to investigate the lncRNA interactome.
Although this study focused on protein rather than RNA interactors, Moreira and colleagues compared proximity labeling and affinity capture [163]. They demonstrated that both techniques can yield valuable insights. As previously stated in chapter 4, proximity labeling records a temporal history of protein interactions that occur during the labeling window. However, BioID tends to result in a higher number of false positives, particularly for proteins with diffuse localization (e.g., cytoplasmic), when compared to immunoprecipitation (IP). Furthermore, BioID is less effective at identifying low molecular weight proteins, as it depends on the presence of surface-exposed lysine residues, making detection efficiency partly dependent on protein size and structure. Nevertheless, the ability to capture interactions *in vivo* is a major advantage of proximity labeling over IP, which requires cell lysis and exposure of the extract to a solid phase for binding. IP-based affinity capture provides only a static snapshot of stable interactions.
Ultimately, the usefulness of each single method is context dependent, and they should rather be viewed as complementary than mutually exclusive approaches. Interestingly, Moreira et

al. found limited overlap between proteins identified by BioID and those retrieved by affinity capture [163].

Given the limitations encountered with CRISPR-based approaches, an alternative and more feasible method for identifying lncRNA interactomes is RNA immunoprecipitation using antisense oligonucleotides (ASOs). Until CRISPR strategies become more robust for lncRNA tagging, this approach remains a valid option for analysis of lncRNA interactors.
Here, the identification of lncRNA associated proteins is typically performed using multiple, often chemically modified ASOs. These are commonly used in combination with a cross-linking step to preserve native RNA–protein interactions. Cross-linking to covalently stabilizes protein–RNA contacts can be achieved either by UV irradiation [164], or by formaldehyde fixation [76]. Following cross-linking, proteins bound to the RNA of interest are captured through hybridization with ASOs that are e.g. covalently bound to magnetic resins or labeled with biotin. After washing steps to reduce nonspecific binding, associated proteins are identified by mass spectrometry [76], [164]. This strategy, especially when combined with tiling oligonucleotides spanning the full length of the RNA, has been successfully applied to map the interactomes of several lncRNAs, including the well-studied Xist RNA [165].

Compared to proximity labeling methods like BioID, ASO-based approaches offer direct, sequence-specific capture of the RNA of interest and can provide higher specificity, particularly valuable when precise localization or isoform-specific binding is critical. Given the disadvantages of IP mentioned before, another method based on *in vitro* proximity labeling is discussed in the following chapter.

11.3. Discussion of Project 2

The second project of this thesis aimed to identify RNA-binding proteins (RBPs) that interact with lncRNAs using proximity labeling, but in contrast to the first project, an *in vitro* approach with fixed cells rather than live ones was used.
To achieve this aim, the HyPro (hybridization-proximity labeling) assay, originally developed by Yap et al. (2021), was applied. In the original study, fixed cells were incubated with the purified HyPro enzyme to identify the interactomes of three non-coding RNAs, including the lncRNA NEAT1. The HyPro protein consists of a DIG-binding domain, which anchors the enzyme to DIG-labeled antisense oligonucleotides (ASOs), fused to the peroxidase APEX2 to catalyze biotin labeling of nearby proteins [64] [66].
Building on this strategy, the assay was first established using the nuclear lncRNA NEAT1 as a model target. In a second step, the goal was to extend the method to the cytoplasmic lncRNAs OIP5-AS1 and NORAD. Following successful establishment, optimization of the biotin labeling conditions and streptavidin pull-down would be performed, ultimately enabling the identification of enriched proteins by mass spectrometry.
By establishing this *in vitro* proximity labeling assay, the project aimed to uncover the interac-

tomes of OIP5-AS1 and NORAD, thereby providing insights into their molecular functions.

As an initial step, the HyPro enzyme was successfully expressed, purified, and validated for both peroxidase and digoxigenin-binding activity. Meanwhile, the DIG-labeled NEAT1 probes were confirmed to specifically recognize NEAT1, as demonstrated by their precise co-localization with NEAT1-Atto488 probes (see Figure 20).
However, the binding of the HyPro enzyme was less specific and robust than expected. Using an anti-His antibody to detect the HyPro enzyme, clear co-localization with paraspeckles was only seen at the lowest DIG-probe concentration (see Figure 21(A)). For comparison, SNAP-APEX2 and Halo-APEX2 enzymes demonstrated clear and specific localization to nuclear paraspeckles (see Figure 22), and thus will be used as positive controls in subsequent experiments.
To analyze the *in vitro* biotinylation activity, several experimental conditions were tested. However, HyPro-mediated proximity labeling did not specifically co-localize with the NEAT1 signal (see Figure 23(A)). The reason for this failure remains unclear, as the DIG-labeled probes correctly localize to paraspeckles. In contrast, SNAP-APEX2 effectively deposited biotin at the expected site (see Figure 24(A)).

Searching explanations for the observed differences, one possible explanation is that the HyPro enzyme may not efficiently target or stably bind the hybridized DIG-probes under the tested conditions. Although the parameters from the published protocol were used [66] the SNAP- and Halo-tag seem superior in stably binding the labeled oligonucleotides. Further optimization is needed to improve specificity and achieve effective RNA-directed biotinylation by HyPro. This could involve adjusting probe design, concentrations, or adopting alternative binding chemistries, as successfully applied in SNAP-APEX2 and Halo-APEX2 systems.
Beyond probe binding, other factors may contribute to the failure of correct HyPro-mediated biotinylation. Besides protein quality, the size of the labeling proteins and consequently their accessibility to the nucleus might play a role. As all three constructs share APEX2, they only differ in the size of their targeting domains. The HaloTag (33 kDa) is the largest, which could explain the failure to biotinylate the NEAT1 surrounding effectively. However, the DIG10.3 domain in HyPro is smaller (15 kDa) than the SNAP-tag (19 kDa). So size alone cannot explain the lack in biotinylation specificity for HyPro, nonetheless it might be a contributing factor (as confirmed in personal discussion with Daniel Hofacker (Stafforst group, IFIB)).
A further influence could be the charge or isoelectric point (pI) of the labeling proteins. As the nucleus harbors many histones that are highly basic proteins, a lower pI could facilitate nuclear access. However, the pI of DIG10.3 is between that of SNAP-tag and HaloTag, indicating that charge alone is likely a minor factor for the observed differences.

11.4. Conclusion and Outlook of Project 2

The *in vitro* proximity labeling assay represents a promising strategy to investigate the interactome of lncRNAs, but a significant difference in the performance of labeling enzymes was

observed depending on the targeting domain.
All three tested enzymes here (HyPro, SNAP-APEX2 and Halo-APEX2) are peroxidase based, but instead of using APEX2 as a labeling enzyme, biotin ligases could offer a valuable alternative. As already tested by Daniel Hofacker (Stafforst group, IFIB), they performed robustly, offering more flexibility as a wider variety of enzymes is available. As an interesting future option, the recently developed ultraID enzyme [40], with a size of only 20 kDa, could offer advantages when coupled to a binding domain. Given that size may have contributed to the challenges faced with Halo-APEX2 and HyPro, ultraID presents a promising candidate.
However, it remains to be determined whether the *in vitro* proximity labeling approach will work for cytoplasmic lncRNAs such as OIP5-AS1 and NORAD. Unlike NEAT1, these RNAs are not highly concentrated in discrete nuclear structures like paraspeckles. This results in lower local concentrations of the labeling enzyme at a discrete spot in the cell. It needs to be investigated whether the biotinylation signal will be sufficiently strong to allow detection by microscopy and analysis by mass spectrometry.
Nevertheless, *in vitro* proximity labeling offers advantages compared to live-cell PL approaches. As labeling occurs in fixed and permeabilized cells, labor-intensive genetic modifications are not required. This broadens the applicability of PL to a wider range of biological systems, including complex genomic loci that are challenging to manipulate by CRISPR. Importantly, many lncRNA genes overlap with protein-coding genes or originate from bidirectional promoters [68]. Therefore, CRISPR-mediated editing carries a significant risk of disrupting neighboring or overlapping genes. In such cases, *in vitro* PL offers a valuable and less invasive alternative for studying lncRNA interactomes.

11.5. Discussion of Project 3

The localization of β-actin mRNA plays a crucial role in establishing cell polarity and directional motility [133]. To better understand the mechanisms governing β-actin mRNA localization, it is essential to identify its interacting proteins. In 2019, Mukherjee et al. developed the RNA-BioID assay and applied it to β-actin mRNA in MEFs, uncovering a set of putative interactors. From this unbiased screen, four members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family — hnRNP K, hnRNP R, PTBP1, and TDP-43 — were selected for further investigation. Their interaction with β-actin mRNA was analyzed by Co-IP and subsequent RT-qPCR of co-immunoprecipitated RNAs.

11.5.1. Co-IP of hnRNP K and TDP-43

Heterogeneous nuclear ribonucleoprotein K (hnRNP K) is a ubiquitously expressed RNA-binding protein involved in pre-mRNA processing, mRNA metabolism, and transport. It has also been implicated in the localization of β-actin mRNA [136].

To analyze the interacting RNAs of hnRNP K, Co-IP was performed. The first attempts were unsatisfactory, and the hypothesis that the antibody coupling to the beads was the limiting step was refuted in a new pull-down experiment, where samples were taken after the coupling step to control for remaining antibody in the supernatant (see Western blot analysis in Figure 26). To improve the protocol, an alternative approach was tested in which the lysate was first incubated with the anti-hnRNP K antibody before adding the beads. This adjustment aimed to prevent potential steric hindrance that could impair antibody binding to the protein when pre-coupled to the beads. However, the elution still only showed minor enrichment.

TAR DNA-binding protein 43 kDa (TDP-43) is a multifunctional RNA-binding protein involved in RNA processing, mRNA transport, and local translation. It plays a crucial role in neuronal development and axon dynamics and has been suggested to contribute to β-actin mRNA transport [144].
To investigate potential RNA interactions, Co-IP was performed. Initial attempts were unsuccessful, as shown by Western blot analysis (see Figure 27) which confirmed the presence of TDP-43 in MEF lysate, but the elution fractions showed only minor enrichment. Despite multiple protocol optimizations, the results remained unchanged.

To improve Co-IP efficiency for both hnRNP K and TDP-43, pre-clearing the lysate with magnetic Protein G Dynabeads prior to immunoprecipitation could help reduce non-specific protein binding. Another possible explanation for the unsuccessful pull-down is that the antibodies may preferentially recognize denatured hnRNP K/TDP-43, as both antibodies performed well in Western blot but failed to recognize native proteins in Co-IP. Potentially, the antibody affinity is too weak, leading to too much protein loss during washes (see e.g. bands after the wash step, right blot in Figure 26). A solution to this could be cross-linking the antibody to Protein G Dynabeads using DSS (disuccinimidyl suberate) or BS3 (bis-sulfosuccinimidyl suberate) to prevent antibody dissociation and contamination from heavy and light chains. Additionally, the quality of the used antibodies may not be well-suited for immunoprecipitation.
In summary, further optimization or the use of a more suitable antibody may be necessary to achieve reliable and specific immunoprecipitation.

11.5.2. Co-IP of PTBP1 and analysis of associated RNAs by RT-qPCR

Polypyrimidine tract-binding protein 1 (PTBP1) is a well-characterized RNA-binding protein involved in nearly all aspects of mRNA regulation. It has been suggested to play a role in the localization of β-actin mRNA to the leading edge of fibroblasts [21].
To investigate whether PTBP1 interacts with β-actin mRNA, a Co-IP was performed using a PTBP1-specific antibody. After the antibody incubation with the MEF lysate, a portion of PTBP1 remained unbound in the supernatant, suggesting that the antibody may have been saturated. Reducing the amount of lysate while maintaining the same antibody concentration should solve this. Nevertheless, the elution sample contained a strong 56 kDa band, confirming successful PTBP1 immunoprecipitation.

To assess whether PTBP1 interacts with β-actin mRNA, RNA was isolated from the beads after pull-down, reverse transcribed into cDNA, and analyzed by RT-qPCR. A positive and negative control were included to validate the specificity of enrichment. MINK1 mRNA was selected as a positive control [155]. GAPDH mRNA was initially chosen as a negative control, based on previous findings that it does not interact with PTBP1 [1].
RT-qPCR results revealed a 1.6 fold enriched of β-actin mRNA in the beads + α-PTBP1 sample over the control sample without antibody, confirming successful enrichment of mRNA. As expected, MINK1 mRNA showed the highest enrichment, verifying its suitability as a positive control. However, GAPDH mRNA was also enriched, suggesting that it may interact non-specifically with the antibody or with another protein binding to PTBP1. Due to this inconsistency, GAPDH was deemed an unsuitable negative control. In order to obtain reliable and robust results, a new negative control must be included in further RT-qPCR experiments.

11.5.3. Co-IP of hnRNP R and analysis of associated RNAs by RT-qPCR

Heterogeneous nuclear ribonucleoprotein R (hnRNP R) is a widely expressed RNA-binding protein involved in RNA metabolism and has been suggested to play a role in β-actin mRNA localization [166]. To investigate this interaction, a Co-IP assay was performed.
Since no specific antibody for hnRNP R was available, a plasmid (p2315) was constructed by cloning the hnRNP R coding sequence with a C-terminal HA tag into another plasmid containing an eGFP tag. The GFP tag enables the isolation of hnRNP R via an anti-GFP nanobody coupled to magnetic agarose beads, while also allowing fluorescence microscopy-based verification of fusion protein expression and transfection efficiency. The HA tag facilitated Western blot detection. Sequencing confirmed the successful construction of p2315, and test transfection of MEFs demonstrated proper expression of the fusion protein.
For pull-down experiments, MEF cells were transfected with p2315, while untransfected MEFs served as a negative control. Western blot analysis confirmed fusion protein expression, with the input sample from transfected cells (lane 1, Figure 29(A)) showing a band above 100 kDa, absent in untransfected cells (lane 2). Although the expected molecular weight of the fusion protein was 99.4 kDa, post-translational modifications or incomplete denaturation could account for the observed shift. Since the sequence downstream of the GFP gene was not sequenced, it could be possible that the stop codon right after the GFP gene is mutated, leading to a larger fusion protein. The elution sample from transfected cells (lane 7) again displayed a band above 100 kDa, corresponding to the fusion protein, while no bands were detected in the elution sample of untransfected cells, confirming the specificity of the anti-GFP nanobody.
Following successful hnRNP R isolation, RNA was extracted from the beads, reverse transcribed into cDNA, and analyzed by RT-qPCR. The 7SK non-coding RNA was used as a positive control due to its known interaction with hnRNP R [151], while GAPDH mRNA served as a negative control.
RT-qPCR analysis showed the strongest enrichment for 7SK ncRNA, validating it as a positive control. Agarose gel electrophoresis of RT-qPCR probes confirmed specific amplification of 7SK, with no bands detected in the -RT control, ensuring the absence of genomic DNA

contamination. β-Actin mRNA was more enriched than the negative control, supporting an interaction with hnRNP R. However, GAPDH mRNA also showed some enrichment, suggesting potential nonspecific interactions with the GFP tag or beads. To further optimize the assay, additional wash steps could be introduced to reduce background binding, and DNase treatment could be refined to ensure complete removal of contaminating DNA. To reliably confirm the interaction between β-actin mRNA and hnRNP R, another negative control has to be found for further RT-qPCR experiments.

11.6. Conclusion and Outlook of Project 3

To sum up, the isolation of the RBPs hnRNP R and PTBP1 was successful, and their interaction with β-actin mRNA was confirmed. However, further optimization of the Co-IP protocol is required to improve the isolation efficiency of hnRNP K and TDP-43, as well as to conclusively prove the interaction between these RBPs with β-actin mRNA.
One possible approach to achieve the Co-IP of TDP-43 and hnRNP K is to use a tagged version of the protein expressed from a plasmid instead of the endogenous one. The transiently expressed, tagged protein (e.g. FLAG- or HA-tagged) could then be immunoprecipitated using a tag-specific antibody, likely improving enrichment and analysis.

To gain a deeper understanding of β-actin mRNA transport, localization, and translation, which is crucial for maintaining cell polarity and motility, further investigations are necessary. Since hnRNP R and PTBP1 were shown to interact with β-actin mRNA in Co-IP experiments, the next step is to validate and characterize these interactions using complementary methods. Several assays could provide more detailed insights into RBP-RNA interactions.
First, RNA immunoprecipitation followed by RNA sequencing (RNA-Seq) instead of qPCR could be performed. This would allow for a broader analysis of RNA targets associated with the RBP, providing a comprehensive view of its interactome and again confirm the interaction with β-actin mRNA.
Additionally, mutational analysis of the RNA-binding site could be performed to determine which RNA regions are critical for RBP binding. This could involve Co-IP, EMSA, or luciferase reporter assays to compare interaction strength between wild-type RNA and RNA constructs with deletions or mutations in predicted RBP binding sites, for example in the 3'UTR.
To validate the interaction *in vivo*, live-cell imaging using the MS2 system could be employed. In this approach, the already existing MEF cell line expressing β-actin mRNA tagged with MS2 stem-loops and a 2xMCP-BirA*-eGFP fusion protein [1] could be transfected with an RBP fused to a fluorescent protein, allowing visualization of potential RNA-RBP co-localization and transport in living cells. This technique would provide spatiotemporal insights into RNA dynamics. Alternatively, RNA FISH combined with immunofluorescence could be performed to assess RBP-RNA co-localization in fixed cells.
In conclusion, these experiments would confirm direct RBP-RNA interactions (IP combined with RNA-Seq), validate specificity (mutational analysis), and provide insights into function

and localization (live-cell imaging). Together, these approaches will contribute to a more comprehensive understanding of how these RBPs regulate β-actin mRNA localization and function.

11

Bibliography

[1] J. Mukherjee et al., "β-Actin mRNA interactome mapping by proximity biotinylation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 26, pp. 12 863–12 872, 2019.

[2] *The Nobel Prize in Physiology or Medicine*, https://www.nobelprize.org/prizes/medicine/, 2025.

[3] G. Meister, *RNA Biology: An Introduction*. Weinheim: Wiley-VCH, 2011.

[4] K. C. Martin and A. Ephrussi, "mRNA localization: Gene expression in the spatial dimension," *Cell*, vol. 136, no. 4, pp. 719–30, 2009.

[5] M. Corley, M. C. Burns, and G. W. Yeo, "How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms," *Molecular Cell*, vol. 78, no. 1, pp. 9–29, 2020.

[6] R. Weissinger, L. Heinold, S. Akram, R.-P. Jansen, and O. Hermesh, "RNA Proximity Labeling: A New Detection Tool for RNA–Protein Interactions," *Molecules*, vol. 26, no. 8, p. 2270, 2021.

[7] W. Gilbert, "Origin of life: The RNA world," *Nature*, vol. 319, no. 6055, pp. 618–618, 1986.

[8] D. K. Hendrix, S. E. Brenner, and S. R. Holbrook, "RNA structural motifs: Building blocks of a modular biomolecule," *Quarterly reviews of biophysics*, vol. 38, no. 3, pp. 221–43, 2005.

[9] International Human Genome Sequencing Consortium, "Finishing the euchromatic sequence of the human genome," *Nature*, vol. 431, no. 7011, pp. 931–945, 2004.

[10] J. S. Mattick et al., "Long non-coding RNAs: Definitions, functions, challenges and recommendations," *Nature reviews. Molecular cell biology*, 2023.

[11] International Human Genome Sequencing Consortium, "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[12] P. Patraquim, E. G. Magny, J. I. Pueyo, A. I. Platero, and J. P. Couso, "Translation and natural selection of micropeptides from long non-canonical RNAs," *Nature Communications*, vol. 13, no. 1, p. 6515, 2022.

[13] J. L. Charles Richard and P. J. A. Eichhorn, "Platforms for Investigating LncRNA Functions," *SLAS technology*, vol. 23, no. 6, pp. 493–506, 2018.

[14] B. M. Lewin, *Genes VIII*, 8th ed. Upper Saddle River, N.J: Pearson education international / Prentice Hall, 2004.

[15] L. A. Passmore and J. Coller, "Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression," *Nature Reviews Molecular Cell Biology*, vol. 23, no. 2, pp. 93–106, 2022.

[16] G. Ruffenach, L. Medzikovic, W. Sun, J. Hong, and M. Eghbali, "Functions of RNA-Binding Proteins in Cardiovascular Disease," *Cells*, vol. 12, no. 24, p. 2794, 2023.

[17] F. M. Fazal and H. Y. Chang, "Subcellular Spatial Transcriptomes: Emerging Frontier for Understanding Gene Regulation," *Cold Spring Harbor symposia on quantitative biology*, 2020.

[18] K. Schuschel, M. Helwig, S. Hüttelmaier, D. Heckl, J.-H. Klusmann, and J. I. Hoell, "RNA-Binding Proteins in Acute Leukemias," *International journal of molecular sciences*, vol. 21, no. 10, 2020.

[19] A. Castello et al., "Insights into RNA biology from an atlas of mammalian mRNA-binding proteins," *Cell*, vol. 149, no. 6, pp. 1393–406, 2012.

[20] K. Bomsztyk, O. Denisenko, and J. Ostrowski, "hnRNP K: One protein multiple processes," *BioEssays : news and reviews in molecular, cellular and developmental biology*, vol. 26, no. 6, pp. 629–38, 2004.

[21] P. Kafasla et al., "Defining the roles and interactions of PTB," *Biochemical Society transactions*, vol. 40, no. 4, pp. 815–20, 2012.

[22] D. Kovács et al., "Lipid exchange at ER–trans-Golgi contact sites governs polarized cargo sorting," *Journal of Cell Biology*, vol. 223, no. 1, e202307051, 2024.

[23] J. Quan et al., "Leveraging biotin-based proximity labeling to identify cellular factors governing early alphaherpesvirus infection," *mBio*, vol. 15, no. 8, C. B. Coyne, Ed., e01445–24, 2024.

[24] S. Markmiller et al., "Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules," *Cell*, vol. 172, no. 3, pp. 590–60 413, 2018.

[25] K. Bersuker et al., "A Proximity Labeling Strategy Provides Insights into the Composition and Dynamics of Lipid Droplet Proteomes," *Developmental Cell*, vol. 44, no. 1, pp. 97–1127, 2018.

[26] Q. Liu et al., "A proximity-tagging system to identify membrane protein-protein interactions," *Nature methods*, vol. 15, no. 9, pp. 715–722, 2018.

[27] H.-W. Rhee et al., "Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging," *Science (New York, N.Y.)*, vol. 339, no. 6125, pp. 1328–1331, 2013.

[28] J. D. Martell et al., "Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy," *Nature Biotechnology*, vol. 30, no. 11, pp. 1143–8, 2012.

[29] D. Kim, K. C. Birendra, W. Zhu, K. Motamedchaboki, V. Doye, and K. J. Roux, "Probing nuclear pore complex architecture with proximity-dependent biotinylation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 24, pp. 2453–61, 2014.

[30] K. Kwon and D. Beckett, "Function of a conserved sequence motif in biotin holoenzyme synthetases," *Protein Science*, vol. 9, no. 8, pp. 1530–1539, 2000.

[31] D. Beckett, E. Kovaleva, and P. J. Schatz, "A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation," *Protein Science : A Publication of the Protein Society*, vol. 8, no. 4, pp. 921–9, 1999.

[32] E. Choi-Rhee, H. Schulman, and J. E. Cronan, "Promiscuous protein biotinylation by Escherichia coli biotin protein ligase," *Protein science : a publication of the Protein Society*, vol. 13, no. 11, pp. 3043–50, 2004.

[33] E. N. Firat-Karalar and T. Stearns, "Probing mammalian centrosome structure using BioID proximity-dependent biotinylation," *Methods in Cell Biology*, vol. 129, pp. 153–170, 2015.

[34] A. Chapman-Smith and J. E. Cronan, "Molecular biology of biotin attachment to proteins," *The Journal of Nutrition*, vol. 129, no. 2S Suppl, pp. 477–484, 1999.

[35] K. J. Roux, D. Kim, M. Raida, and B. Burke, "A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells," *The Journal of cell biology*, vol. 196, no. 6, pp. 801–10, 2012.

[36] J. E. Cronan, "Biotin protein ligase as you like it: Either extraordinarily specific or promiscuous protein biotinylation," *Proteins: Structure, Function, and Bioinformatics*, vol. 92, no. 4, pp. 435–448, 2024.

[37] D. Kim et al., "An improved smaller biotin ligase for BioID proximity labeling," *Molecular biology of the cell*, vol. 27, no. 8, pp. 1188–96, 2016.

[38] K. Kido et al., "AirID, a novel proximity biotinylation enzyme, for analysis of protein-protein interactions," *eLife*, vol. 9, 2020.

[39] M. Ramanathan et al., "RNA-protein interaction detection in living cells," *Nature methods*, vol. 15, no. 3, pp. 207–212, 2018.

[40] L. Kubitz et al., "Engineering of ultraID, a compact and hyperactive enzyme for proximity-dependent biotinylation in living cells," *Communications biology*, vol. 5, no. 1, p. 657, 2022.

[41] T. C. Branon et al., "Efficient proximity labeling in living cells and organisms with TurboID," *Nature biotechnology*, vol. 36, no. 9, pp. 880–887, 2018.

[42] M. Larochelle, D. Bergeron, B. Arcand, and F. Bachand, "Proximity-dependent biotinylation mediated by TurboID to identify protein-protein interaction networks in yeast," *Journal of Cell Science*, vol. 132, no. 11, 2019.

[43] A. Mair, S.-L. Xu, T. C. Branon, A. Y. Ting, and D. C. Bergmann, "Proximity labeling of protein complexes and cell-type-specific organellar proteomes in Arabidopsis enabled by TurboID," *eLife*, vol. 8, 2019.

[44] Y. Zhang et al., "TurboID-based proximity labeling reveals that UBR7 is a regulator of N NLR immune receptor-mediated immunity," *Nature Communications*, vol. 10, no. 1, p. 3252, 2019.

[45] D. G. May, K. L. Scott, A. R. Campos, and K. J. Roux, "Comparative Application of BioID and TurboID for Protein-Proximity Biotinylation," *Cells*, vol. 9, no. 5, 2020.

[46] J. S. Rees, X.-W. Li, S. Perrett, K. S. Lilley, and A. P. Jackson, "Selective Proteomic Proximity Labeling Assay Using Tyramide (SPPLAT): A Quantitative Method for the Proteomic Analysis of Localized Membrane-Bound Protein Clusters," *Current Protocols in Protein Science*, vol. 80, pp. 19 271–192 718, 2015.

[47] C. Hopkins, A. Gibson, J. Stinchcombe, and C. Futter, "Chimeric molecules employing horseradish peroxidase as reporter enzyme for protein localization in the electron microscope," *Methods in Enzymology*, vol. 327, pp. 35–45, 2000.

[48] N. Kotani, J. Gu, T. Isaji, K. Udaka, N. Taniguchi, and K. Honke, "Biochemical visualization of cell surface molecular clustering in living cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 21, pp. 7405–9, 2008.

[49] S. S. Lam et al., "Directed evolution of APEX2 for electron microscopy and proximity labeling," *Nature Methods*, vol. 12, no. 1, pp. 51–4, 2015.

[50] B. T. Lobingier et al., "An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells," *Cell*, vol. 169, no. 2, pp. 350–36 012, 2017.

[51] A. Padrón, S. Iwasaki, and N. T. Ingolia, "Proximity RNA Labeling by APEX-Seq Reveals the Organization of Translation Initiation Complexes and Repressive RNA Granules," *Molecular Cell*, vol. 75, no. 4, pp. 875–8875, 2019.

[52] F. M. Fazal et al., "Atlas of Subcellular RNA Localization Revealed by APEX-Seq," *Cell*, vol. 178, no. 2, pp. 473–49 026, 2019.

[53] E. Prikas, A. Poljak, and A. Ittner, "Mapping p38α mitogen-activated protein kinase signaling by proximity-dependent labeling," *Protein Science*, vol. 29, no. 5, pp. 1196–1210, 2020.

[54] E. Schmidtmann, T. Anton, P. Rombaut, F. Herzog, and H. Leonhardt, "Determination of local chromatin composition by CasID," *Nucleus (Austin, Tex.)*, vol. 7, no. 5, pp. 476–484, 2016.

[55] X. D. Gao et al., "C-BERST: Defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2," *Nature Methods*, vol. 15, no. 6, pp. 433–436, 2018.

[56] S. A. Myers, J. Wright, R. Peckner, B. T. Kalish, F. Zhang, and S. A. Carr, "Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling," *Nature Methods*, vol. 15, no. 6, pp. 437–439, 2018.

[57] M. Lu and W. Wei, "Proximity labeling to detect RNA-protein interactions in live cells," *FEBS open bio*, vol. 9, no. 11, pp. 1860–1868, 2019.

[58] T. Lionnet et al., "A transgenic mouse for in vivo detection of endogenous labeled mRNA," *Nature Methods*, vol. 8, no. 2, pp. 165–70, 2011.

[59] J. F. Garcia and R. Parker, "MS2 coat proteins bound to yeast mRNAs block 5' to 3' degradation and trap mRNA decay products: Implications for the localization of mRNAs by MS2-MCP system," *RNA (New York, N.Y.)*, vol. 21, no. 8, pp. 1393–5, 2015.

[60] J.-H. Spille, M. Hecht, V. Grube, W.-K. Cho, C. Lee, and I. I. Cissé, "A CRISPR/Cas9 platform for MS2-labelling of single mRNA in live stem cells," *Methods (San Diego, Calif.)*, vol. 153, pp. 35–45, 2019.

[61] E. Tutucci, M. Vera, J. Biswas, J. Garcia, R. Parker, and R. H. Singer, "An improved MS2 system for accurate reporting of the mRNA life cycle," *Nature Methods*, vol. 15, no. 1, pp. 81–89, 2018.

[62] H. Laprade et al., "Single-Molecule Imaging of Telomerase RNA Reveals a Recruitment-Retention Model for Telomere Elongation," *Molecular Cell*, vol. 79, no. 1, pp. 115–1266, 2020.

[63] A. Smirnov, C. Schneider, J. Hör, and J. Vogel, "Discovery of new RNA classes and global RNA-binding proteins," *Current Opinion in Microbiology*, vol. 39, pp. 152–160, 2017.

[64] K. Yap, T. H. Chung, and E. V. Makeyev, "Hybridization-proximity labeling reveals spatially ordered interactions of nuclear RNA compartments," *Molecular cell*, 2021.

[65] N. Tsanov et al., "smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability," *Nucleic acids research*, vol. 44, no. 22, p. 165, 2016.

[66] K. Yap, T. H. Chung, and E. V. Makeyev, "Analysis of RNA-containing compartments by hybridization and proximity labeling in cultured human cells," *STAR Protocols*, vol. 3, no. 1, p. 101 139, 2022.

[67] M. Sauvageau, "Diverging RNPs: Toward Understanding lncRNA-Protein Interactions and Functions," *Advances in experimental medicine and biology*, vol. 1203, pp. 285–312, 2019.

[68] A. Goyal, K. Myacheva, M. Groß, M. Klingenberg, B. Duran Arqué, and S. Diederichs, "Challenges of CRISPR/Cas9 applications for long non-coding RNA genes," *Nucleic acids research*, vol. 45, no. 3, p. 12, 2017.

[69] H. Bunch, "Gene regulation of mammalian long non-coding RNA," *Molecular genetics and genomics : MGG*, vol. 293, no. 1, pp. 1–15, 2018.

[70] J. Barra and E. Leucci, "Probing Long Non-coding RNA-Protein Interactions," *Frontiers in molecular biosciences*, vol. 4, p. 45, 2017.

[71] J. H. Noh, K. M. Kim, W. G. McClusky, K. Abdelmohsen, and M. Gorospe, "Cytoplasmic functions of long noncoding RNAs," *Wiley interdisciplinary reviews. RNA*, vol. 9, no. 3, p. 1471, 2018.

[72] C. J. Ross, A. Rom, A. Spinrad, D. Gelbard-Solodkin, N. Degani, and I. Ulitsky, "Uncovering deeply conserved motif combinations in rapidly evolving noncoding sequences," *Genome biology*, vol. 22, no. 1, p. 29, 2021.

[73] M. C. Bridges, A. C. Daulagala, and A. Kourtidis, "LNCcation: lncRNA localization and function," *The Journal of cell biology*, vol. 220, no. 2, 2021.

[74] G. Pintacuda, A. N. Young, and A. Cerase, "Function by Structure: Spotlights on Xist Long Non-coding RNA," *Frontiers in molecular biosciences*, vol. 4, p. 90, 2017.

[75] A. Minajigi et al., "A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation," *Science (New York, N.Y.)*, vol. 349, no. 6245, 2015.

[76] C. Chu et al., "Systematic discovery of Xist RNA binding proteins," *Cell*, vol. 161, no. 2, pp. 404–16, 2015.

[77] J. M. Engreitz et al., "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome," *Science (New York, N.Y.)*, vol. 341, no. 6147, p. 1 237 973, 2013.

[78] S. J. Liu et al., "CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells," *Science (New York, N.Y.)*, vol. 355, no. 6320, 2017.

[79] C. Lin and L. Yang, "Long Noncoding RNA in Cancer: Wiring Signaling Circuitry," *Trends in cell biology*, vol. 28, no. 4, pp. 287–301, 2018.

[80] S. Lee et al., "Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins," *Cell*, vol. 164, no. 1-2, pp. 69–80, 2016.

[81] A. Tichon et al., "A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells," *Nature communications*, vol. 7, p. 12 209, 2016.

[82] M. M. Elguindy et al., "PUMILIO, but not RBMX, binding is required for regulation of genomic stability by noncoding RNA NORAD," *eLife*, vol. 8, 2019.

[83] U. Chorostecki, E. Saus, and T. Gabaldón, "Structural characterization of NORAD reveals a stabilizing role of spacers and two new repeat units," *Computational and structural biotechnology journal*, vol. 19, pp. 3245–3254, 2021.

[84] M. M. Elguindy and J. T. Mendell, "NORAD-induced Pumilio phase separation is required for genome stability," *Nature*, vol. 595, no. 7866, pp. 303–308, 2021.

[85] A. Tichon, R. B.-T. Perry, L. Stojic, and I. Ulitsky, "SAM68 is required for regulation of Pumilio by the NORAD long noncoding RNA," *Genes & development*, vol. 32, no. 1, pp. 70–78, 2018.

[86] M. Spiniello et al., "HyPR-MS for Multiplexed Discovery of MALAT1, NEAT1, and NORAD lncRNA Protein Interactomes," *Journal of Proteome Research*, vol. 17, no. 9, pp. 3022–3038, 2018.

[87] A. Ventura, "NORAD: Defender of the Genome," *Trends in genetics : TIG*, vol. 32, no. 7, pp. 390–392, 2016.

[88] F. Kopp et al., "PUMILIO hyperactivity drives premature aging of Norad-deficient mice," *eLife*, vol. 8, 2019.

[89] *NCBI*, https://www.ncbi.nlm.nih.gov/gene?Db=gene\&Cmd=DetailsSearch\&Term=729082, 2022.

[90] I. Ulitsky, A. Shkumatava, C. H. Jan, H. Sive, and D. P. Bartel, "Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution," *Cell*, vol. 147, no. 7, pp. 1537–50, 2011.

[91] E. Wanowska, M. Kubiak, I. Makałowska, and M. W. Szcześniak, "A chromatin-associated splicing isoform of OIP5-AS1 acts in cis to regulate the OIP5 oncogene," *RNA biology*, pp. 1–12, 2021.

[92] M. Goudarzi, K. Berg, L. M. Pieper, and A. F. Schier, "Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility," *eLife*, vol. 8, 2019.

[93] K. N. Smith, J. Starmer, S. C. Miller, P. Sethupathy, and T. Magnuson, "Long Noncoding RNA Moderates MicroRNA Activity to Maintain Self-Renewal in Embryonic Stem Cells," *Stem cell reports*, vol. 9, no. 1, pp. 108–121, 2017.

[94] H. J. Hunkler et al., "The Long Non-coding RNA Cyrano Is Dispensable for Pluripotency of Murine and Human Pluripotent Stem Cells," *Stem cell reports*, 2020.

[95] B. Kleaveland et al., "A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain," *Cell*, vol. 174, no. 2, pp. 350–36 217, 2018.

[96] J. Kim et al., "LncRNA OIP5-AS1/cyrano sponges RNA-binding protein HuR," *Nucleic acids research*, vol. 44, no. 5, pp. 2378–92, 2016.

[97] J. Kim et al., "LncRNA OIP5-AS1/cyrano suppresses GAK expression to control mitosis," *Oncotarget*, vol. 8, no. 30, pp. 49 409–49 420, 2017.

[98] M. Naemura et al., "The Long Noncoding RNA OIP5-AS1 Is Involved in the Regulation of Cell Proliferation," *Anticancer research*, vol. 38, no. 1, pp. 77–81, 2018.

[99] S. Ghafouri-Fard, S. Dashti, M. Farsi, B. M. Hussen, and M. Taheri, "A review on the role of oncogenic lncRNA OIP5-AS1 in human malignancies," *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, vol. 137, p. 111 366, 2021.

[100] C. Keryer-Bibens, C. Barreau, and H. B. Osborne, "Tethering of proteins to RNAs by bacteriophage proteins," *Biology of the Cell*, vol. 100, no. 2, pp. 125–38, 2008.

[101] F. Lim and D. S. Peabody, "RNA recognition site of PP7 coat protein," *Nucleic Acids Research*, vol. 30, no. 19, pp. 4138–44, 2002.

[102] D. Peabody, "The RNA binding site of bacteriophage MS2 coat protein.," *The EMBO Journal*, vol. 12, no. 2, pp. 595–600, 1993.

[103] H. E. Johansson, L. Liljas, and O. C. Uhlenbeck, "RNA Recognition by the MS2 Phage Coat Protein," *Seminars in Virology*, vol. 8, no. 3, pp. 176–185, 1997.

[104] E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long, "Localization of ASH1 mRNA Particles in Living Yeast," *Molecular Cell*, vol. 2, no. 4, pp. 437–445, 1998.

[105] E. Tutucci, N. M. Livingston, R. H. Singer, and B. Wu, "Imaging mRNA In Vivo, from Birth to Death," *Annual Review of Biophysics*, vol. 47, pp. 85–106, 2018.

[106] D. Fusco et al., "Single mRNA molecules demonstrate probabilistic movement in living mammalian cells," *Current biology: CB*, vol. 13, no. 2, pp. 161–167, 2003.

[107] G. Haimovich et al., "Use of the MS2 aptamer and coat protein for RNA localization in yeast: A response to "MS2 coat proteins bound to yeast mRNAs block 5' to 3' degradation and trap mRNA decay products: Implications for the localization of mRNAs by MS2-MCP system"," *RNA (New York, N.Y.)*, vol. 22, no. 5, pp. 660–6, 2016.

[108] J.-H. Yoon, S. Srikantan, and M. Gorospe, "MS2-TRAP (MS2-tagged RNA affinity purification): Tagging RNA to identify associated miRNAs," *Methods (San Diego, Calif.)*, vol. 58, no. 2, pp. 81–7, 2012.

[109] H. E. Johansson, D. Dertinger, K. A. LeCuyer, L. S. Behlen, C. H. Greef, and O. C. Uhlenbeck, "A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 16, pp. 9244–9, 1998.

[110] A. Graindorge et al., "In-cell identification and measurement of RNA-protein interactions," *Nature Communications*, vol. 10, no. 1, p. 5317, 2019.

[111] M. A. Mengstie and B. Z. Wondimu, "Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing," *Biologics: Targets and Therapy*, vol. Volume 15, pp. 353–361, 2021.

[112] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang, "Genome engineering using the CRISPR-Cas9 system," *Nature protocols*, vol. 8, no. 11, pp. 2281–2308, 2013.

[113] I. Tasan et al., "CRISPR/Cas9-mediated knock-in of an optimized TetO repeat for live cell imaging of endogenous loci," *Nucleic acids research*, vol. 46, no. 17, p. 100, 2018.

[114] S. Nakade et al., "Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9," *Nature communications*, vol. 5, p. 5560, 2014.

[115] J. Fueller et al., "CRISPR-Cas12a-assisted PCR tagging of mammalian genes," *The Journal of cell biology*, vol. 219, no. 6, 2020.

[116] I. Weidenfeld et al., "Inducible expression of coding and inhibitory RNAs from retargetable genomic loci," *Nucleic acids research*, vol. 37, no. 7, p. 50, 2009.

[117] T.-L. Cheng and Z. Qiu, "Long non-coding RNA tagging and expression manipulation via CRISPR/Cas9-mediated targeted insertion," *Protein & cell*, vol. 9, no. 9, pp. 820–825, 2017.

[118] B. Chen et al., "Live cell imaging and proteomic profiling of endogenous NEAT1 lncRNA by CRISPR/Cas9-mediated knock-in," *Protein & cell*, vol. 11, no. 9, pp. 641–660, 2020.

[119] J. A. West et al., "The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites," *Molecular Cell*, vol. 55, no. 5, pp. 791–802, 2014.

[120] Y. Zhou et al., "Expanding APEX2 Substrates for Proximity-Dependent Labeling of Nucleic Acids and Proteins in Living Cells," *Angewandte Chemie (International ed. in English)*, vol. 58, no. 34, pp. 11 763–11 767, 2019.

[121] B. Singer-Krüger, T. Fröhlich, M. Franz-Wachtel, N. Nalpas, B. Macek, and R.-P. Jansen, "APEX2-mediated proximity labeling resolves protein networks in Saccharomyces cerevisiae cells," *The FEBS Journal*, vol. 287, no. 2, pp. 325–344, 2020.

[122] D. Z. Bar, K. Atkatsh, U. Tavarez, M. R. Erdos, Y. Gruenbaum, and F. S. Collins, "Biotinylation by antibody recognition-a method for proximity labeling," *Nature Methods*, vol. 15, no. 2, pp. 127–133, 2018.

[123] I. Santos-Barriopedro, G. van Mierlo, and M. Vermeulen, "Off-the-shelf proximity biotinylation for interaction proteomics," *Nature communications*, vol. 12, no. 1, p. 5015, 2021.

[124] A. F. Tsue et al., "Oligonucleotide-directed proximity-interactome mapping (O-MAP): A unified method for discovering RNA-interacting proteins, transcripts and genomic loci in situ," *bioRxiv : the preprint server for biology*, 2023.

[125] C. E. Tinberg et al., "Computational design of ligand-binding proteins with high affinity and selectivity," *Nature*, vol. 501, no. 7466, pp. 212–216, 2013.

[126] G. V. Los et al., "HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis," *ACS Chemical Biology*, vol. 3, no. 6, pp. 373–382, 2008.

[127] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, and K. Johnsson, "A general method for the covalent labeling of fusion proteins with small molecules in vivo," *Nature Biotechnology*, vol. 21, no. 1, pp. 86–89, 2003.

[128] N. Porzberg, K. Gries, and K. Johnsson, "Exploiting Covalent Chemical Labeling with Self-Labeling Proteins," *Annual Review of Biochemistry*, 2025.

[129] S. Das, M. Vera, V. Gandin, R. H. Singer, and E. Tutucci, "Intracellular mRNA transport and localized translation," *Nature reviews. Molecular cell biology*, 2021.

[130] T.-G. Du, M. Schmid, and R.-P. Jansen, "Why cells move messages: The biological functions of mRNA localization," *Seminars in Cell & Developmental Biology*, vol. 18, no. 2, pp. 171–177, 2007.

[131] A. Khong and R. Parker, "The Landscape of Eukaryotic mRNPs," *RNA (New York, N.Y.)*, 2019.

[132] R. Dominguez and K. C. Holmes, "Actin Structure and Function," *Annual Review of Biophysics*, vol. 40, no. 1, pp. 169–186, 2011.

[133] E. A. Shestakova, R. H. Singer, and J. Condeelis, "The physiological significance of beta -actin mRNA localization in determining cell polarity and directional motility," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 13, pp. 7045–50, 2001.

[134] Z. B. Katz, A. L. Wells, H. Y. Park, B. Wu, S. M. Shenoy, and R. H. Singer, "β-Actin mRNA compartmentalization enhances focal adhesion stability and directs cell migration," *Genes & Development*, vol. 26, no. 17, pp. 1885–1890, 2012.

[135] J. Condeelis and R. H. Singer, "How and why does beta-actin mRNA target?" *Biology of the cell*, vol. 97, no. 1, pp. 97–110, 2005.

[136] Z. Wang et al., "The emerging roles of hnRNPK," *Journal of cellular physiology*, vol. 235, no. 3, pp. 1995–2008, 2020.

[137] M. Gallardo, M. J. Hornbaker, X. Zhang, P. Hu, C. Bueso-Ramos, and S. M. Post, "Aberrant hnRNP K expression: All roads lead to cancer," *Cell cycle (Georgetown, Tex.)*, vol. 15, no. 12, pp. 1552–7, 2016.

[138] J. Ostrowski, Y. Kawata, D. S. Schullery, O. N. Denisenko, and K. Bomsztyk, "Transient recruitment of the hnRNP K protein to inducibly transcribed gene loci," *Nucleic acids research*, vol. 31, no. 14, pp. 3954–62, 2003.

[139] C. Proepper et al., "Heterogeneous nuclear ribonucleoprotein k interacts with Abi-1 at postsynaptic sites and modulates dendritic spine morphology," *PloS one*, vol. 6, no. 11, p. 27 045, 2011.

[140] S. T. Warraich, S. Yang, G. A. Nicholson, and I. P. Blair, "TDP-43: A DNA and RNA binding protein with roles in neurodegenerative diseases," *The international journal of biochemistry & cell biology*, vol. 42, no. 10, pp. 1606–9, 2010.

[141] J. R. Tollervey et al., "Characterizing the RNA targets and position-dependent splicing regulation by TDP-43," *Nature neuroscience*, vol. 14, no. 4, pp. 452–8, 2011.

[142] M. Polymenidou et al., "Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43," *Nature neuroscience*, vol. 14, no. 4, pp. 459–68, 2011.

[143] C. Fallini, G. J. Bassell, and W. Rossoll, "The ALS disease protein TDP-43 is actively transported in motor neuron axons and regulates axon outgrowth," *Human molecular genetics*, vol. 21, no. 16, pp. 3703–18, 2012.

[144] A. Prasad, V. Bharathi, V. Sivalingam, A. Girdhar, and B. K. Patel, "Molecular Mechanisms of TDP-43 Misfolding and Pathology in Amyotrophic Lateral Sclerosis," *Frontiers in molecular neuroscience*, vol. 12, p. 25, 2019.

[145] K. Weskamp and S. J. Barmada, "TDP43 and RNA instability in amyotrophic lateral sclerosis," *Brain research*, vol. 1693, no. Pt A, pp. 67–74, 2018.

[146] S. Ma, G. Liu, Y. Sun, and J. Xie, "Relocalization of the polypyrimidine tract-binding protein during PKA-induced neurite growth," *Biochimica et biophysica acta*, vol. 1773, no. 6, pp. 912–23, 2007.

[147] M. B. Coelho et al., "Functional interactions between polypyrimidine tract binding protein and PRI peptide ligand containing proteins," *Biochemical Society transactions*, vol. 44, no. 4, pp. 1058–65, 2016.

[148] K. Sawicka, M. Bushell, K. A. Spriggs, and A. E. Willis, "Polypyrimidine-tract-binding protein: A multifunctional RNA-binding protein," *Biochemical Society transactions*, vol. 36, no. Pt 4, pp. 641–7, 2008.

[149] R. Spellman et al., "Regulation of alternative splicing by PTB and associated factors," *Biochemical Society transactions*, vol. 33, no. Pt 3, pp. 457–60, 2005.

[150] P. L. Boutz et al., "A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons," *Genes & development*, vol. 21, no. 13, pp. 1636–52, 2007.

[151] M. Briese et al., "hnRNP R and its main interactor, the noncoding RNA 7SK, coregulate the axonal transcriptome of motoneurons," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 12, pp. 2859–2868, 2018.

[152] B. Dombert, R. Sivadasan, C. M. Simon, S. Jablonka, and M. Sendtner, "Presynaptic localization of Smn and hnRNP R in axon terminals of embryonic and postnatal mouse motoneurons," *PloS one*, vol. 9, no. 10, p. 110 846, 2014.

[153] M. Glinka et al., "The heterogeneous nuclear ribonucleoprotein-R is necessary for axonal beta-actin mRNA translocation in spinal motor neurons," *Human molecular genetics*, vol. 19, no. 10, pp. 1951–66, 2010.

[154] L. M. Gittings, S. C. Foti, B. C. Benson, P. Gami-Patel, A. M. Isaacs, and T. Lashley, "Heterogeneous nuclear ribonucleoproteins R and Q accumulate in pathological inclusions in FTLD-FUS," *Acta neuropathologica communications*, vol. 7, no. 1, p. 18, 2019.

[155] Y. Xue et al., "Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping," *Molecular Cell*, vol. 36, no. 6, pp. 996–1006, 2009.

[156] K. J. Schreiber, E. Kadijk, and J.-Y. Youn, "Exploring Options for Proximity-Dependent Biotinylation Experiments: Comparative Analysis of Labeling Enzymes and Affinity Purification Resins," *Journal of Proteome Research*, vol. 23, no. 4, pp. 1531–1543, 2024.

[157] S. Munter et al., "Split-BioID: A proximity biotinylation assay for dimerization-dependent protein interactions," *FEBS Letters*, vol. 591, no. 2, pp. 415–424, 2017.

[158] I. M. Schopp et al., "Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes," *Nature Communications*, vol. 8, p. 15 690, 2017.

[159] K. F. Cho et al., "Split-TurboID enables contact-dependent proximity labeling in cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 22, pp. 12 143–12 154, 2020.

[160] G. A. Schaack, O. M. Sullivan, and A. Mehle, "Identifying Protein-Protein Interactions by Proximity Biotinylation with AirID and splitAirID," *Current protocols*, vol. 3, no. 3, p. 702, 2023.

[161] Y. Bollen et al., "Efficient and error-free fluorescent gene tagging in human organoids without double-strand DNA cleavage," *PLoS biology*, vol. 20, no. 1, p. 3 001 527, 2022.

[162] M. T. N. Yarnall et al., "Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases," *Nature biotechnology*, 2022.

[163] C. M. d. N. Moreira et al., "Impact of inherent biases built into proteomic techniques: Proximity labeling and affinity capture compared," *The Journal of biological chemistry*, vol. 299, no. 1, p. 102 726, 2022.

[164] B. Rogell, B. Fischer, M. Rettel, J. Krijgsveld, A. Castello, and M. W. Hentze, "Specific RNP capture with antisense LNA/DNA mixmers," *RNA (New York, N.Y.)*, vol. 23, no. 8, pp. 1290–1302, 2017.

[165] C. A. McHugh et al., "The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3," *Nature*, vol. 521, no. 7551, pp. 232–6, 2015.

[166] S. Cappelli, M. Romano, and E. Buratti, "Systematic Analysis of Gene Expression Profiles Controlled by hnRNP Q and hnRNP R, Two Closely Related Human RNA Binding Proteins Implicated in mRNA Processing Mechanisms," *Frontiers in molecular biosciences*, vol. 5, p. 79, 2018.

Part IV.

Appendix

12 Supplements to Project 1

12.1. Plasmid maps of generated constructs in this work

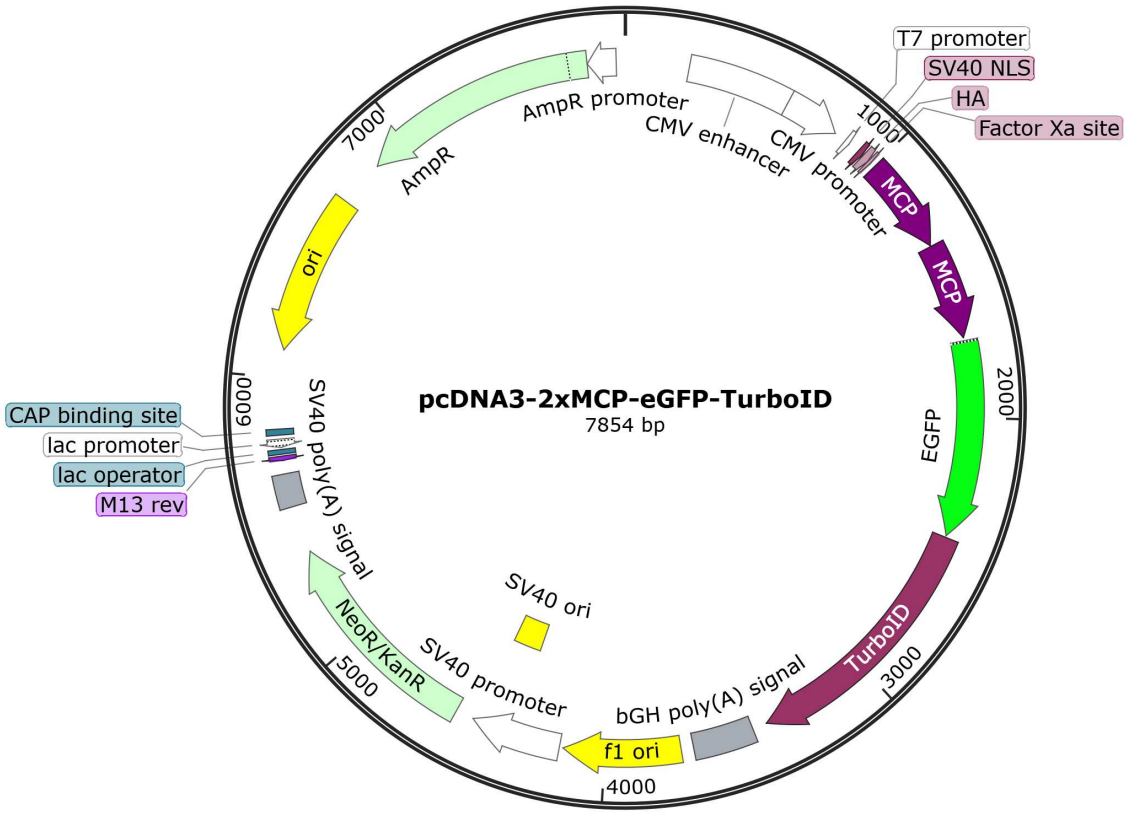


Figure 31.: Plasmid map of p2263, called pcDNA3-2xMCP-eGFP-TurboID. Used for test transfection (see section 8.2.1).

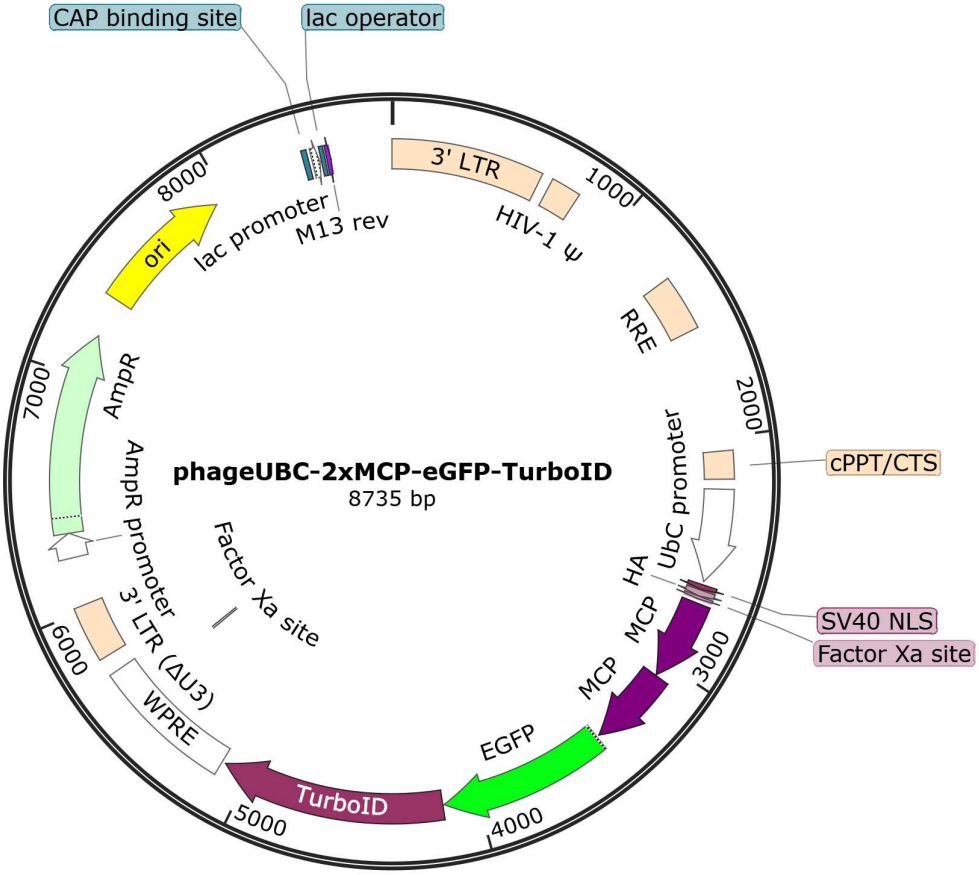


Figure 32.: Plasmid map of p2264, called phageUBC-2xMCP-eGFP-TurboID. Can be used for viral transfection of 2xMCP-eGFP-TurboID construct.

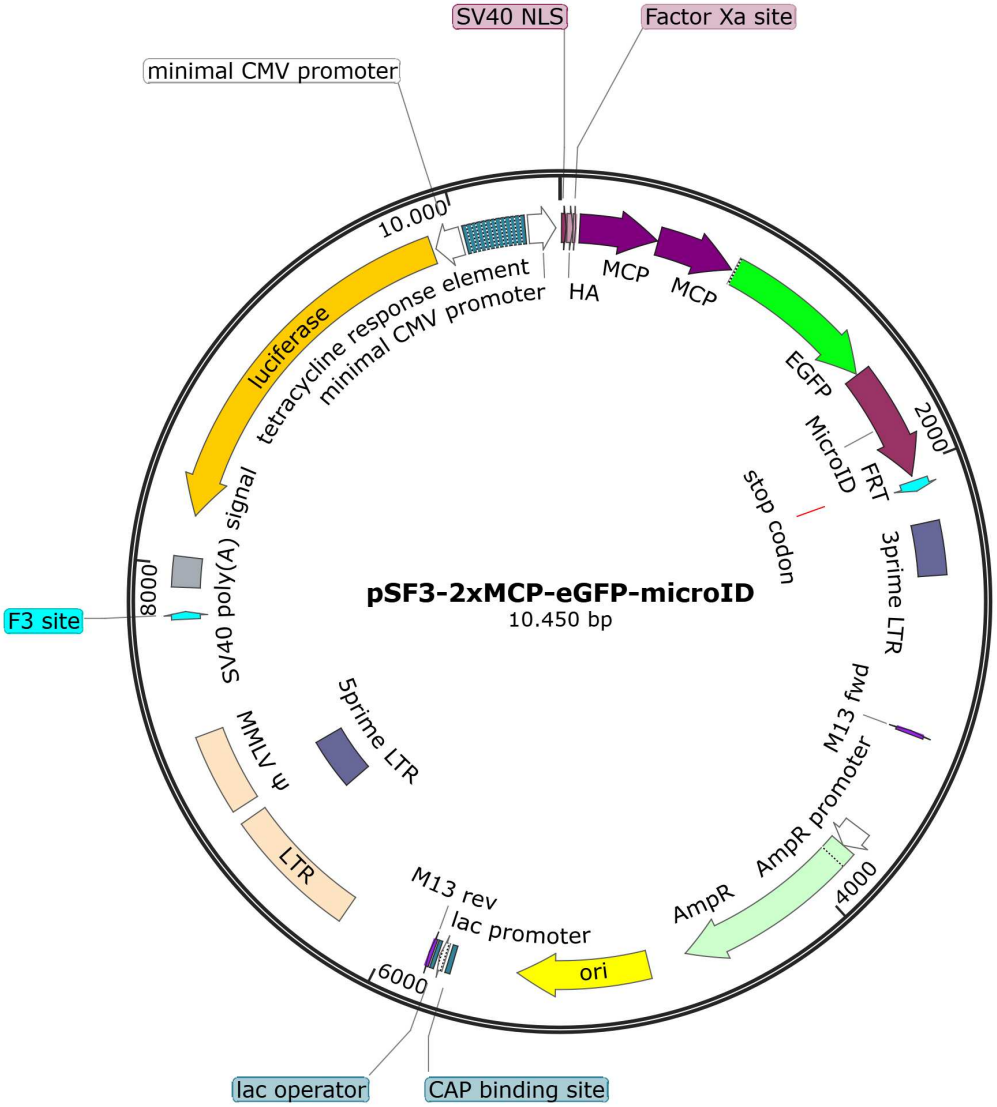


Figure 33.: Plasmid map of p2279, called pSF3-2xMCP-eGFP-microID. Used to create a stable HeLa cell line expressing 2xMCP-eGFP-microID from a doxycycline inducible promoter (see section 8.2.1).

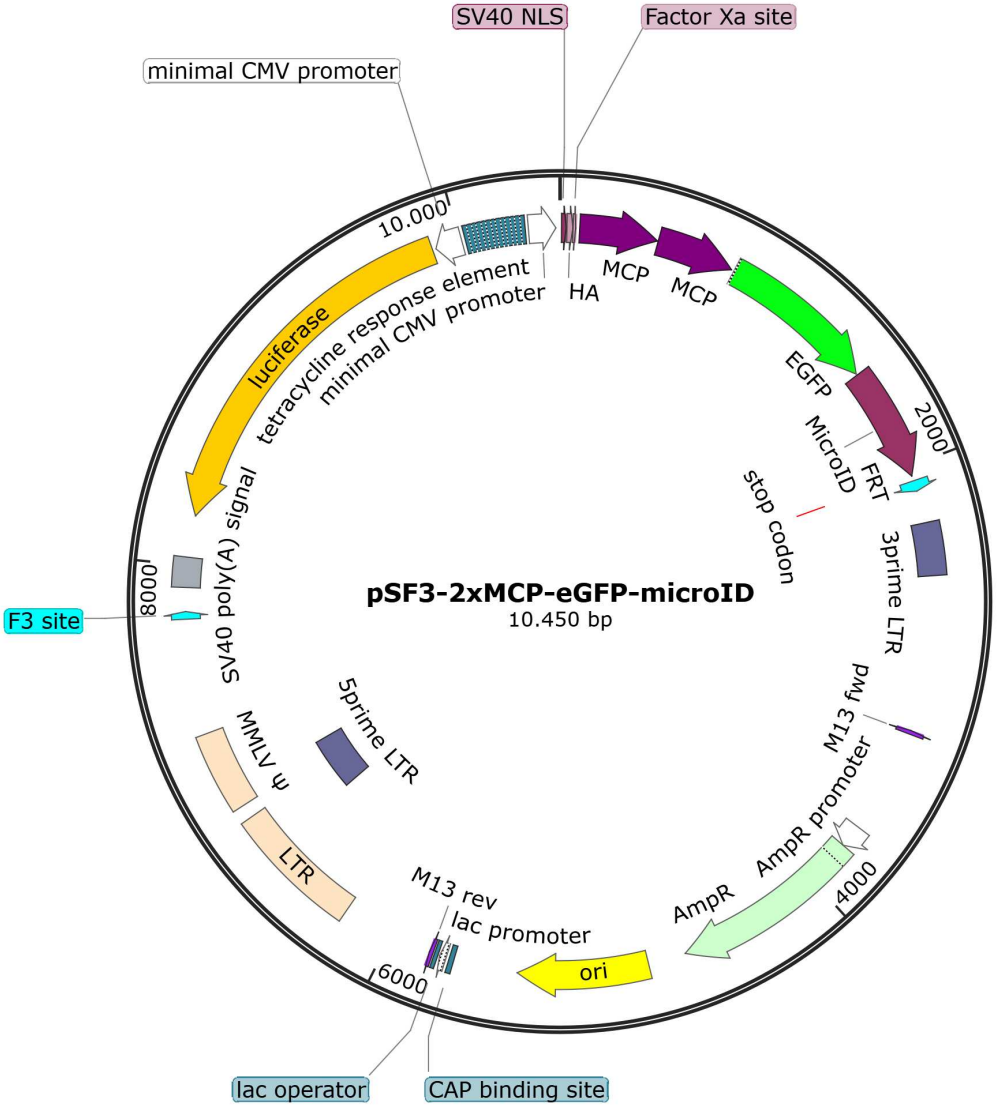


Figure 34.: Plasmid map of p2280, called pSF3-2xMCP-eGFP-TurboID. Used to create a stable HeLa cell line expressing 2xMCP-eGFP-TurboID from a doxycycline inducible promoter (see section 8.2.1).

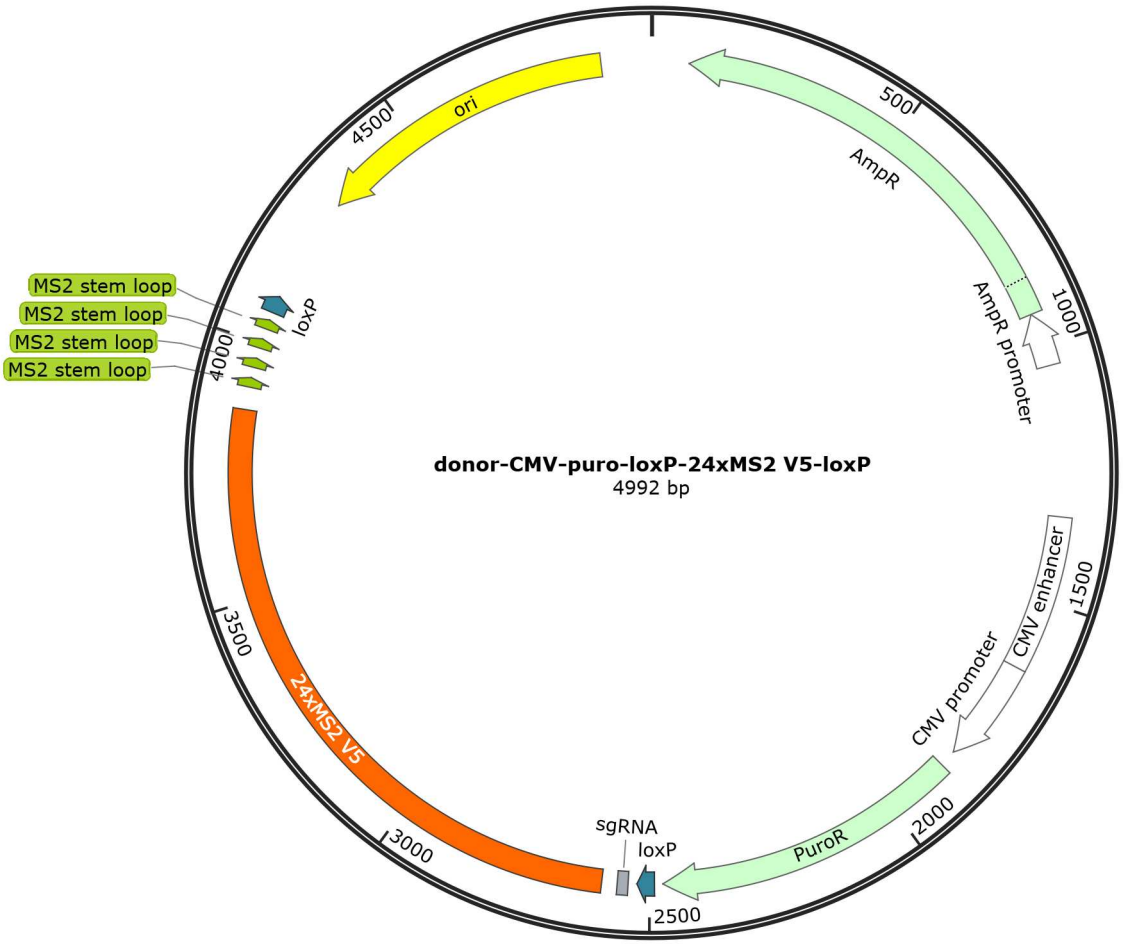


Figure 35.: Plasmid map of p2313, called donor-CMV-puro-loxP-24xMS2 V5-loxP. Used for NHEJ CRISPR attemps.

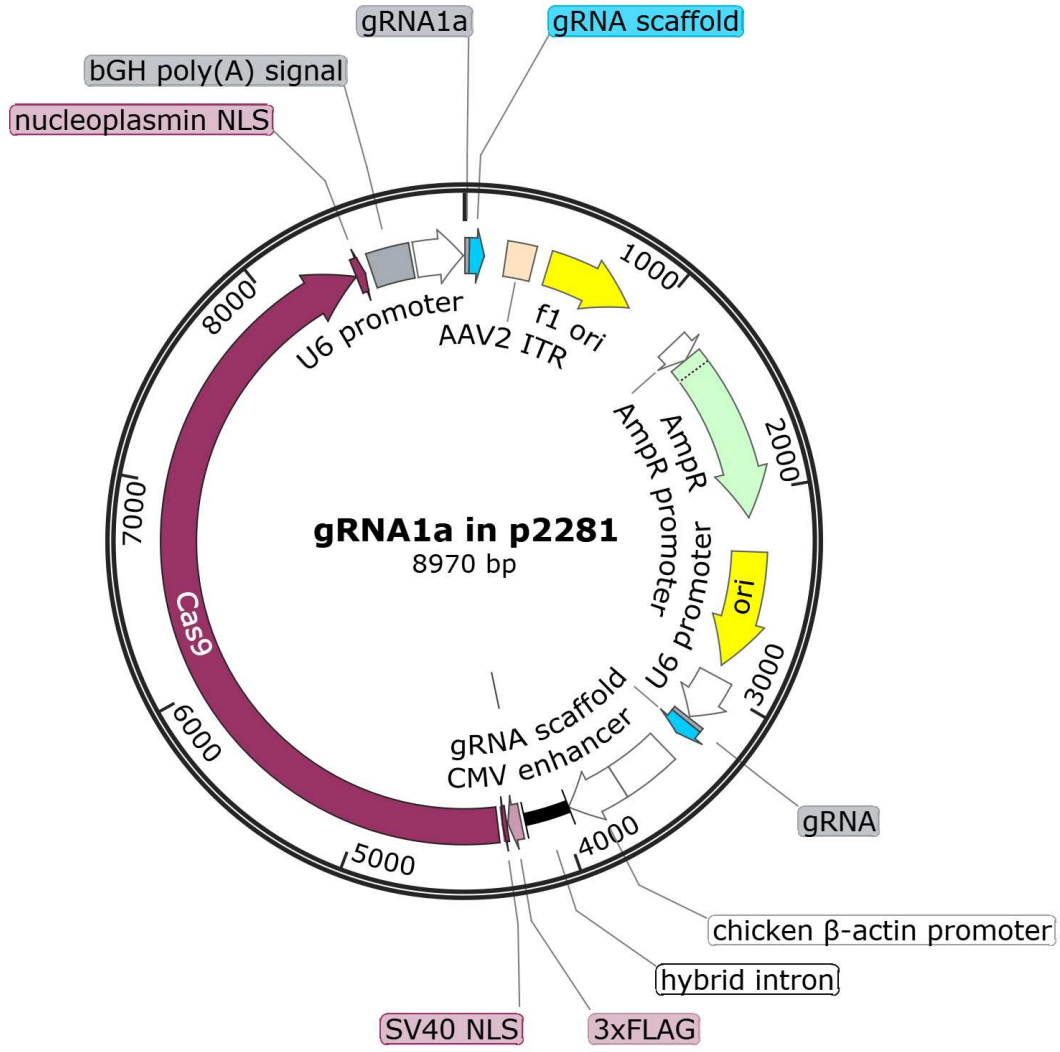


Figure 36.: Plasmid map of p2404, called gRNA1a in p2281. Expresses two gRNAs and the Cas9 enzyme, used for CRISPR. Four similar plasmids (p2405 - p2408) were generated besides this one, each containing a different gRNA targeting the lncRNA OIP5-AS1. The sequences of the gRNAs are listed in Table 13.

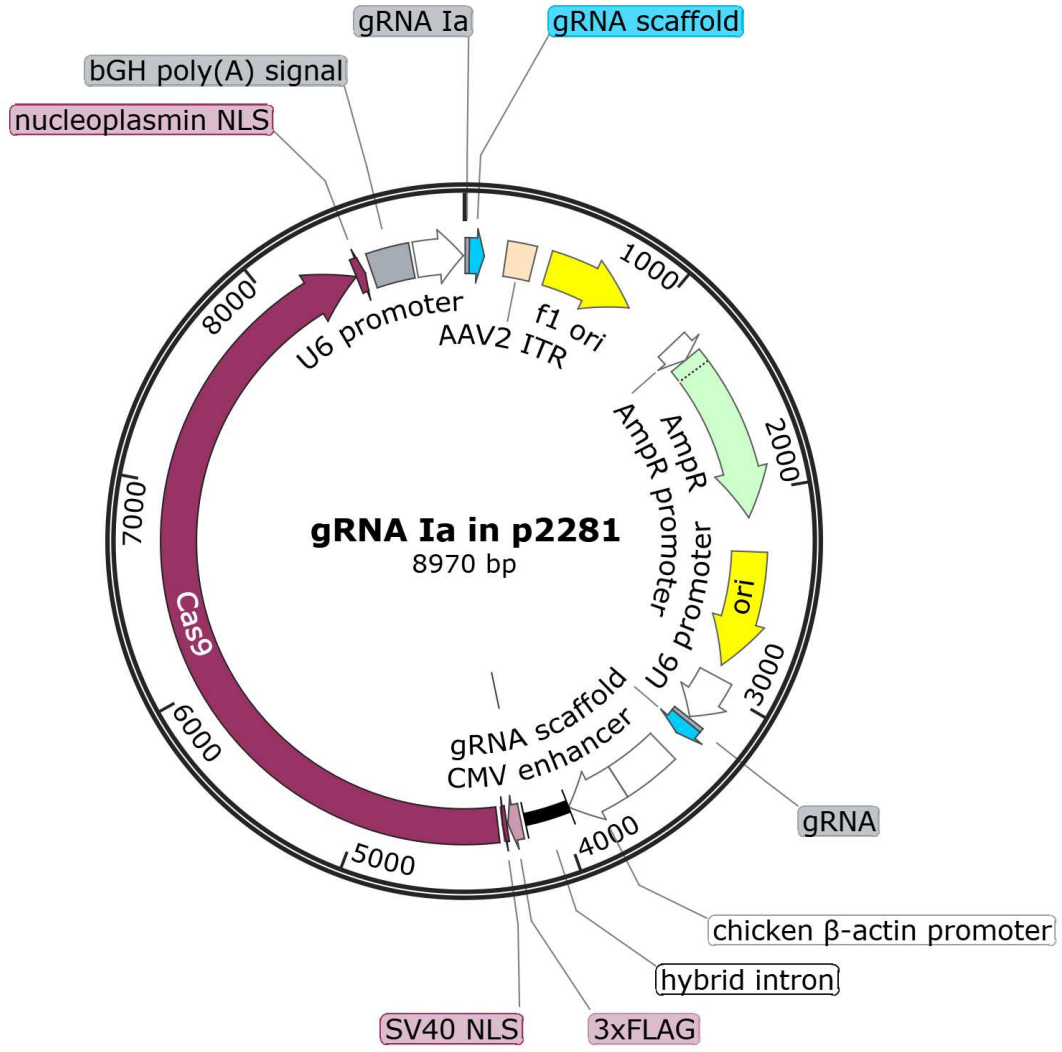


Figure 37.: Plasmid map of p2409, called gRNA Ia in p2281. Expresses two gRNAs and the Cas9 enzyme, used for CRISPR. Six similar plasmids (p2410 - p2415) were generated besides this one, each containing a different gRNA targeting the lncRNA Norad. The sequences of the gRNAs are listed in Table 13.

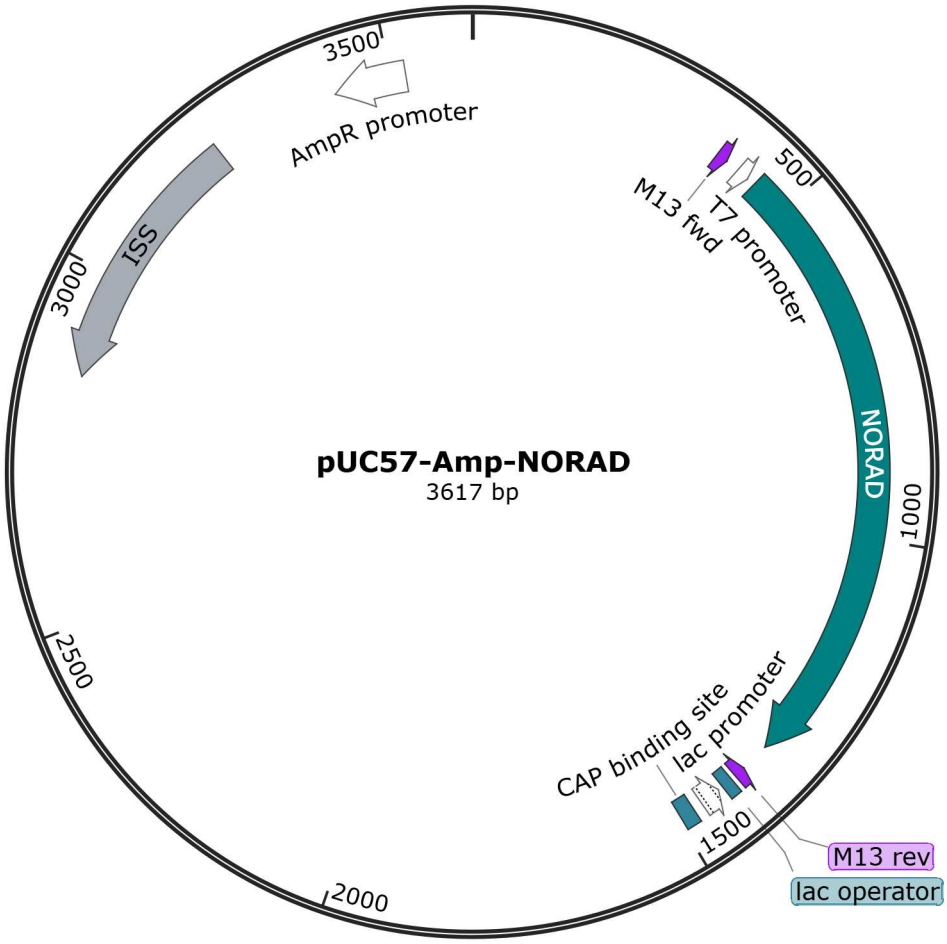


Figure 38.: Plasmid map of p2324, called pUC57-Amp-NORAD. Used to create a DIG-labelled Norad probe for Southern blot analysis.

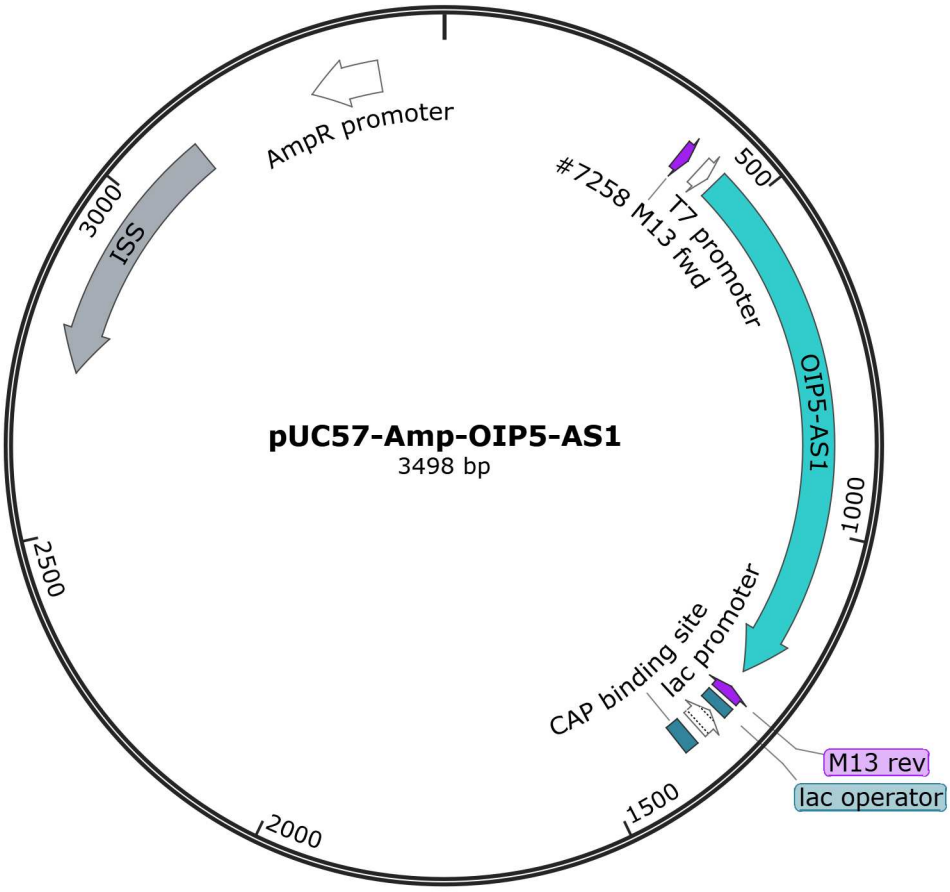


Figure 39.: Plasmid map of p2325, called pUC57-Amp-OIP5-AS1. Used to create a DIG-labelled OIP5-AS1 probe for Southern blot analysis.

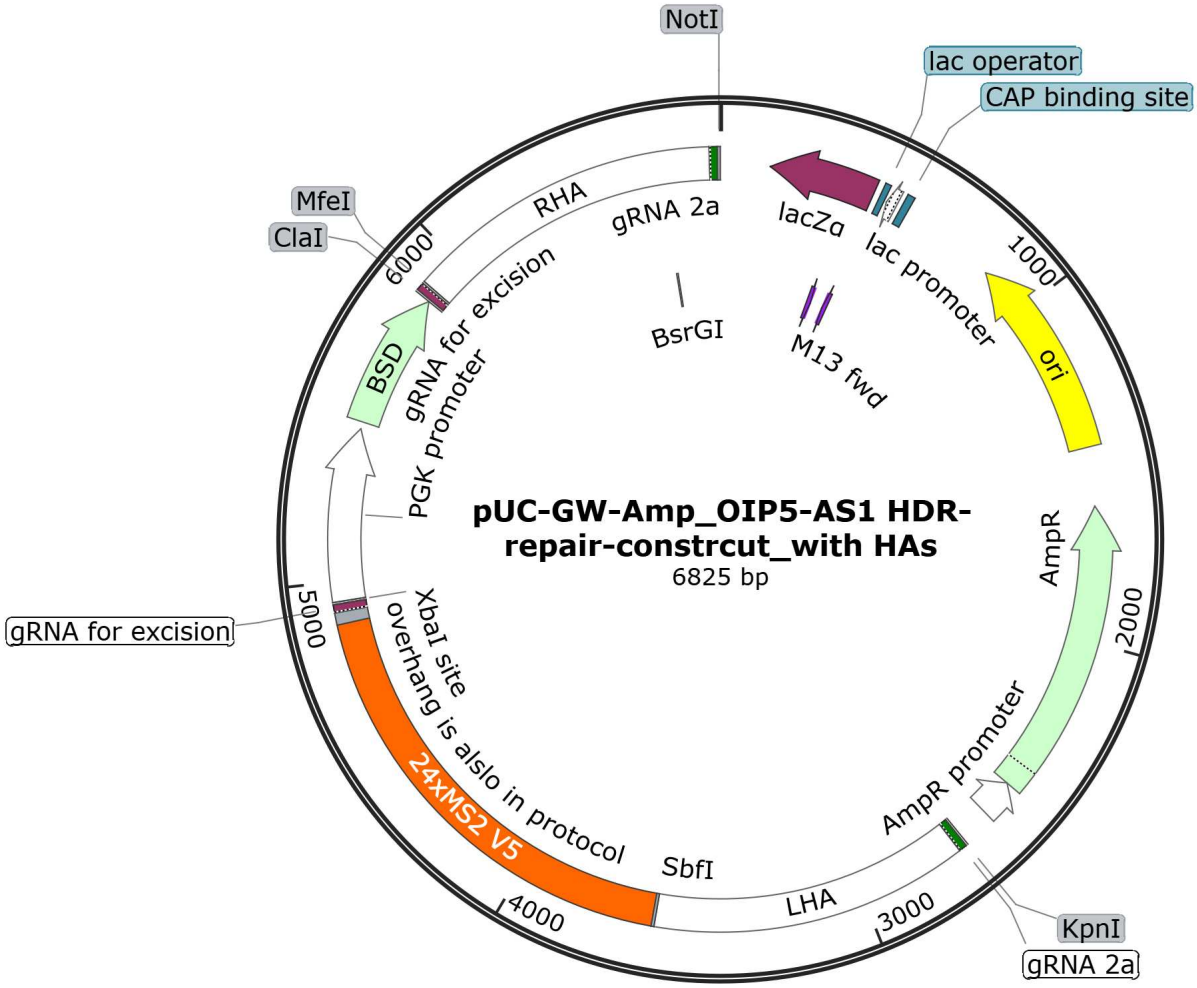


Figure 40.: Plasmid map of p2340, called pUC-GW-Amp_OIP5-AS1 HDR-repair-constrcut_with HAs. Used for HDR CRISPR attemps of OIP5-AS1. LHA = left homology arm, RHA = right homology arm.

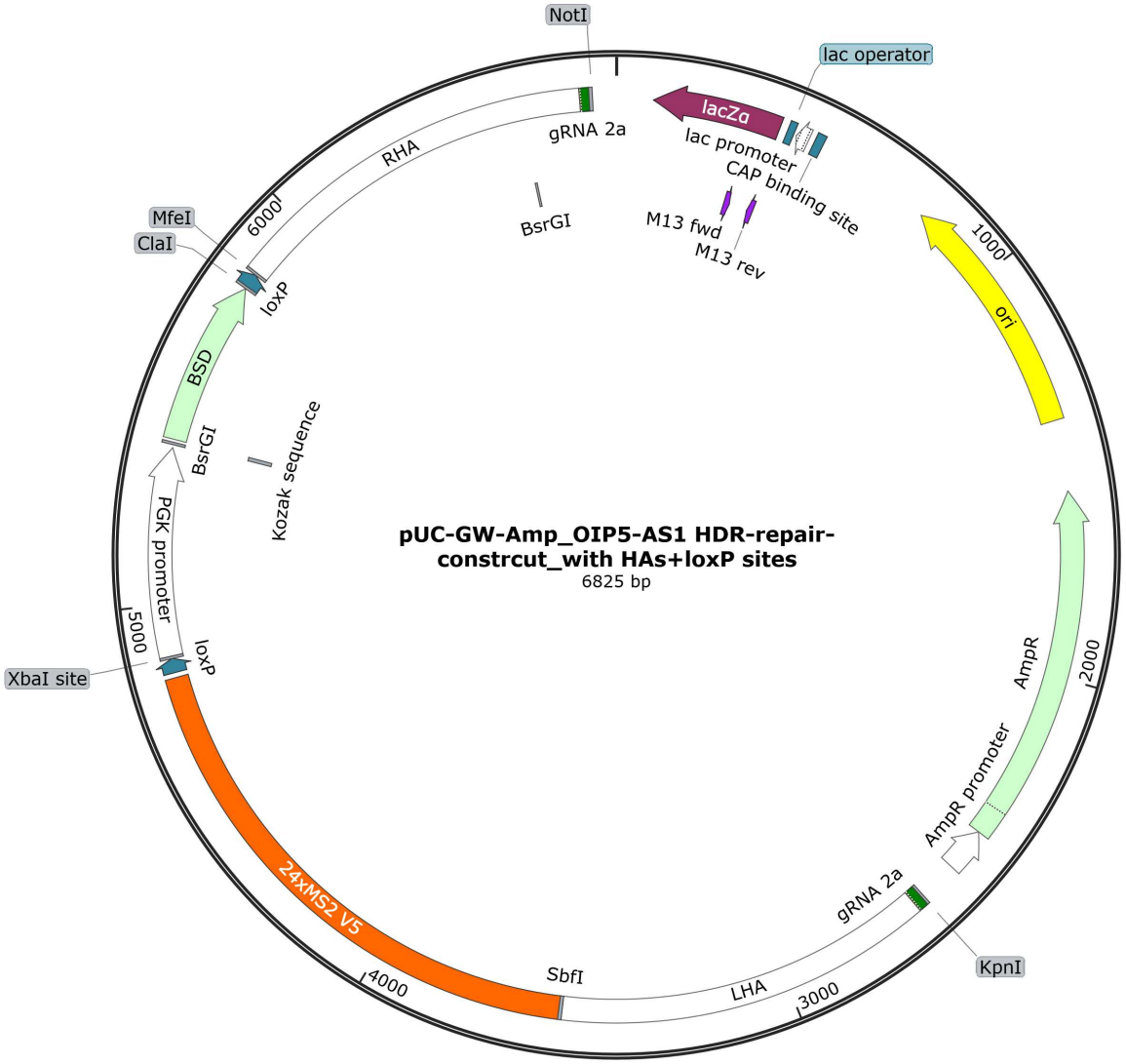


Figure 41.: Plasmid map of p2341, called pUC-GW-Amp_OIP5-AS1 HDR repair template_with HAs+loxP sites. Used for HDR CRISPR attemps of OIP5-AS1, contains losP sites instead of two gRNAs like p2340 for a potential removing of the antibiotic cassette. LHA = left homology arm, RHA = right homology arm.

13 Supplements to Project 3

13.1. Plasmid map of generated construct in this work

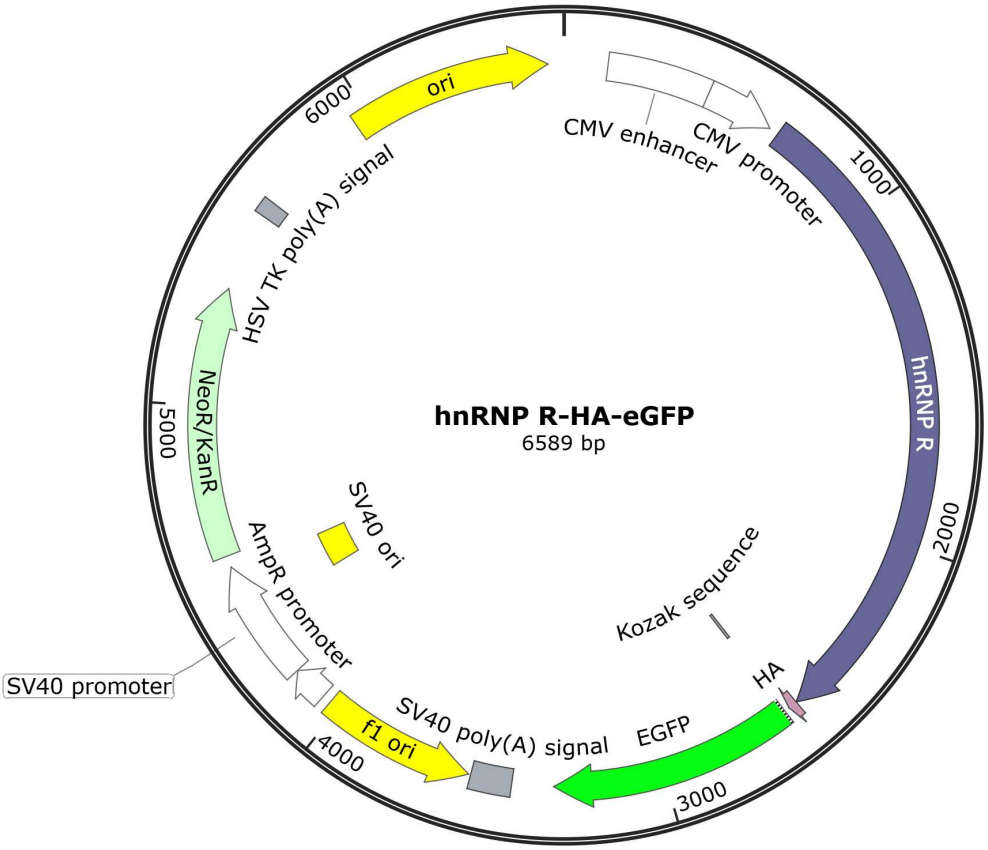


Figure 42.: Plasmid map of p2315, called hnRNP R-HA-eGFP. Used for GFP based pull-down of the RBP hnRNP R.

13 Acknowledgments

Lastly, I want to say "thank you" to all the people who accompanied me during this journey and made this work possible.

I want to express my gratitude to Prof. Dr. Ralf-Peter Jansen, for everything I've learned during my time at the institute. Thank you for your advice, support, and the opportunities to pursue my interests, including the visit to Israel and the Forschergruppen meetings.
I am grateful to Prof. Dr. Thorsten Stafforst for being my second supervisor. I wish to acknowledge your valuable feedback, input and advice throughout the journey, and for inspiring the cooperation with Daniel Hofacker.
I appreciate the time and effort of the examination committee to review my thesis and presentation.
Many thanks to the whole lab and my colleagues, including Lisa, Jonathan, Orit, Iliana, Ulrike, Ingrid, Ruth and Aaron. I'm deeply thankful for your company throughout the last years. Special thanks to Lisa and Jonathan, for your support and for going through this time together. I'm very grateful for all the joyful moments I've spent with Orit, Aaron and Iliana in the lab. Our countless discussions in the office, over lunch and coffee made this time memorable. Thank you for your patience with me. Again, thanks to Daniel Hofacker for the enzymes, reagents, helpful discussions and insights on the HyPro project. I want to extend these thanks to everyone on the 4th floor and the whole institute who contributed to a pleasant work environment. I am fortunate to have met all of you and especially for the new friendships. Thank you Vitasta and Luisa for your accompaniment.
I enjoyed supervising all of my students, namely Melike, Lina, Nina and Nadine. I benefited greatly from your contributions to this work, your feedback and your time. I hope you learned as much from me as I did from you.

This work would not exist without the Mensa and Cafeteria staff, who provided food, as well as coffee and chocolate. Thank you for sustaining me with energy to go through all of this. In addition, thanks to all the artists who created the soundtrack of my thesis, especially during the last months of writing.

During my time in Tübingen, I was lucky to be surrounded by wonderful people. I want to express my deep gratitude to my WG, as they had to live with me and all my emotions that come with a PhD and have stood steadfastly by me, encouraging me until the end. Special thanks to Leonie, Coco and Aaron, you made home always a good place to be. Thanks to Janek, Leo, Chris, Ronja, Thomas, Helen, Nils and Ronja for chatting, cheering me up, cooking, eating, drinking, chilling and all the big and small activities and trips. Great thanks to all of my friends for your support and love.
Thank you to all the people, but also to Yoga, for keeping me sane and to the Hochschulsport for all the stress relief. Additionally, I want to thank my physiotherapists and ergotherapist for fixing me again and again.

I am deeply thankful to my family, particularly my parents, my sister and my parents-in-law, for placing their trust in me, and consistently providing unconditional love and support throughout my journey.
Finally, I can never find the words to thank you enough, Jakob. I could not have done this without you. From all the people mentioned here, you know me best and I'm eternally grateful for your company throughout the last years. Thank you for being the best partner I could ask for.