Aus dem Lehrstuhl für Stoffwechselbiochemie

Biomedizinisches Centrum München (BMC)

Institut der Ludwig-Maximilians-Universität München

Vorstand: Prof. Dr. Dr. Christian Haass

The long non-coding RNA "upstream to Slitrk3" (RUS) affects chromatin organization by binding to Brd2 and Smarca5.

Dissertation

zum Erwerb des Doktorgrades der Naturwissenschaften

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Marius Frederik Schneider

aus

Aalen

Jahr

2021

Mit Genehmigung der Medizinischen Fakultät

der Universität München

Betreuer: Prof. Dr. Dr. h.c. Christian Haass

Zweitgutachter: Prof. Dr. rer.nat. Axel Imhof

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 27.01.2022

Contents

List of Figures

List of Tables

1 Abbreviations

Table 1: Abbreviations

Bmi1	polycomb complex protein BMI-1
Brd2	bromodomain-containing protein 2
Brd4	bromodomain-containing protein 4
CBP	CREB binding protein
ceRNA	competing RNA
CHART	capture hybridization analysis of RNA targets
ChIRP	chromatin isolation by RNA isolation
ciRNA	circular RNA
CN	cortical neurons
DNA	deoxynucleic acid
dNTPs	deoxynucleosidetriphosphate
DSB	double strand break
DSG	disuccinimidyl glutarate
EMSA	electromobiliy shift assay
eRNA	enhancer RNA
ESC	embryonic stem cells
FGF	fibroblast growth factor
Firre	functional intergenic repeating RNA element
FITC	5/(6)carboxyfluorescine
FXR1	Fragile X mental retardation syndrome-related protein 1
FXR2	Fragile X mental retardation syndrome-related protein 2
GFAP	Glial fibrillary acidic protein
Glast1	GLutamate Aspartate Transporter 1 / Excitatory amino acid transporter 1
H2A	histone protein H2A
H2B	histone protein H2B
H3	histone protein H3
H3K27Ac	acetylated H3 lysine 27
H3K27Me3	trimethylated H3 lysine 27
H3K4Me1	monomethylated H3 lysine 4
H3K4Me3	trimethylated H3 lysine 4
H3K9Me2	dimethylated H3 lysine 9
H3K9Me3	trimethylated H3 lysine 9
H4	histone protein H4
H4K12Ac	acetylated H4 lysine 12
H4K20	acetylated H4 lysine 20
HAT	histone acetyltransferase
HDAC	histone deacetylase
hnRNP	heterogeneous RNP

HP1	heterochromatin protein 1
IPS	induced pluripotent stem cells
KD	knock-down
KO	Knock-out
Lbr	lamin B receptor
LC-MS	liquid chromatography coupled to mass spectrometry
LINE	long interspersed nuclear element
lncRNA	long non-coding RNA
Lsd1	lysine-specific histone demethylase 1A
LTR	long terminal repeat
Map2	microtubule associated protein 2
Mapt	microtubule associated protein tau
MBP	maltose binding protein
MS2	MS2 aptamer
MS2BP	MS2 binding protein
NE	neuro-epithelial
NHS	n-hydroxysuccinimide
NPC	neural precursor cell
NSC	neural stem cell
OE	overexpression
p300	p300 acetyltransferase
PCR	polymerase chain reaction
PIC	preinitiation complex
PRC2	polycomb repressive complex 2
PSF	PTB-associated splicing factor
Puf60	poly(U) bBinding splicing factor 60
RACE	rapid amplification of cDNA ends
RAP	RNA antisense purification
RNA	ribonucleic acid
RNP	ribonucleoprotein particle
rNTPs	ribonucleosidetriphosphate
rRNA	ribosomal RNA
RT	reverse transcription
RT-PCR	real time PCR
SEM	standard deviation of the mean
Seq	sequencing
Sf3b	splicing factor 3b
SINE	short interspersed nuclear element
Sirt	sirtuin
Smarca5	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 (Snf2h)
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
SRFSF1	serine argnine rich splie factor 1
STD	standard deviation
SVZ	subventricular zone
TAD	topologically associating domain
TAMRA	5/(6)carboxytetramethylrhodamin
TBP	Tata-box binding-protein
TE	transposable element
TEG	triethyleneglycolyl-glyceryl
Terc	telomerase RNA component
TF	transcription factor
TFO	triplex forming oligonucleotide
TGFbeta	transforming growth factor beta
tRNA	transfer RNA
TSS	transcription start site

TTS triplex targeting site	
TTS-DNA	triplex targeting site DNA
vsv-g	vesicular stomatitis virus G protein

Table 2: IUPAC amino acid codes

IUPAC amino acid code	Three letter code	Amino acid
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Table 3: IUPAC nucleotide code

IUPAC nucleotide code	Base
A	adenine
C	cytosine
G	guanine
T (or U)	thymine (or uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base
. or -	gap

2 Summary

2.1 Zusammenfassung

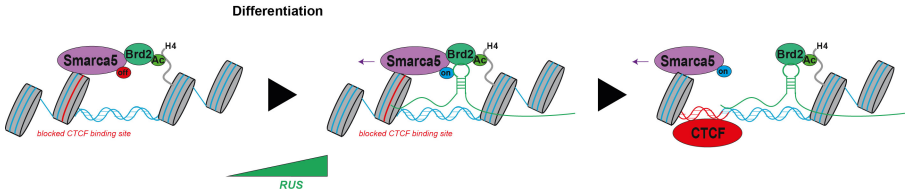


Figure 1: **Grafische Zusammenfassung:** Die lncRNA RUS bindet an Brd2 besetzte Regionen und aktiviert den Chromatinremodellierer Smarca5. Dadurch kommt es zum ortsspezifischen Chromatin-Remodellieren und zur Chromatinbindung von CTCF.

Im Laufe der Evolution hat das zentrale Nervensystem der Säugetiere in seiner Komplexität zugenommen. Ebenfalls nahm die Anzahl einer neuen Klasse von RNA- Pol II Transkripten zu, die als lange nichtkodierende RNAs (lncRNAs) definiert sind. Daher wird vermutet, dass lncRNAs für die massive Zunahme der Gehirngröße sowie die Diversifizierung der neuronalen Zelltypen verantwortlich sind [1]. LncRNAs sind oft Gewebe-spezifisch exprimiert und wirken als zusätzliche Ebene bei der Genregulation. Allerdings ist die biologische Funktion der überwiegenden Mehrheit der Neuronenspezifischen lncRNAs noch unbekannt. Das Ziel der vorliegenden Arbeit war, in verschiedenen Transkriptom-Datensätzen nach konservierten, funktional relevanten lncRNAs zu suchen und deren molekularen Mechanismus zu bestimmen. Um die mutmaßlichen biologischen Funktionen möglicher Kandidaten in der Neurogenese zu testen, wurde die jeweilige Expression in sich differenzierenden neuralen Stammzellen mittels lentiviral transduzierten shRNAs im Sinne eines Knockdowns (KD) heruntergefahren. Die erzielten Phänotypen wurden mit einem neuentwickelten Algorithmus ausgewertet.

Ein Kandidat, dessen Knockdown (KD) die Anzahl an Neuronen signifikant reduzierte, war eine lncRNA, die im Genom upstream zu Slitrk3 liegt – einem Gen, welches wichtig für die Entwicklung von Neuronen ist - und der wir aufgrund dessen den Namen RUS (RNA upstream to Slitrk3) gegeben haben. Eine detaillierte Bewertung des KD-Phänotyps durch histochemische Färbungen und RNA-Sequenzierung zeigte eine verminderte Zellproliferation, ein Stillstand der Zelldifferenzierung und damit einhergehend ein erhöhtes Zellsterben. RUS ist im Zellkern angereichert und in Maus und Mensch konserviert, insbesondere an seinem 5'-Ende. Die Überexpression von RUS nach KD im Sinne eines Rescue-Experimentes stellte die Bildung von Neuronen wieder her. Die Überexpression einer Deletionskontruktes, dem die konservierte 5'-Domäne fehlte, konnte den KD-Phänotyp jedoch nicht retten. Dies ließ vermuten, dass diese 5' konservierte Domäne wichtig für die Funktion der lncRNA ist und als Bindestelle für genregulatorische Proteine dient.

Um interagierende Proteine zu isolieren und mittels LC-MS zu bestimmen, wurde eine neue *in vivo* MS2 basierte RNA-Affinitätsaufreinigung entwickelt. Dafür wurden MS2-getaggte RNAs in Neuro2A-Zellen stabil integriert und überexprimiert. Zur Bestimmung der Proteine, die spezifisch an der konservierten Domäne am 5' Ende binden, wurden diejenigen Proteine aufgereinigt, die an der RUS, der 5' Deletionsmutante und der Domäne selbst binden. Ein Vergleich der Interaktome dieser drei Konstrukte untereinander ergab, dass RUS mittels seiner 5'Domäne spezifisch mit den Bet (Bromodomain and Extra-Terminal motif) Proteinen Brd2, Brd4, und dem ISWI (Imitation SWItch) Chromatin-Remodellierer Smarca5 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5) interagiert. Brd2 und Smarca5 sind bevorzugt an DNA-Stellen zu finden, die von dem Isolator-Protein CTCF gebunden werden [2–4].

ChIRP-Seq (Chromatin Isolation by RNA purification coupled with high troughput sequencing) Experimente gegen RUS in Wt Zellen und ChIP-Seq (Chromatin Immunuprecipitation) Experimente gegen Brd2 in ctrl und RUS KD Zellen zeigten, dass RUS und Brd2 gemeinsam an 13 hippocampal spezifischen CTCF Seiten im Genom gebunden haben, die reich an GAATG Wiederholungen sind. Dabei übte RUS weder einen rekrutierenden noch repulsiven Effekt auf Brd2 aus. Die Spezifität von RUS gegenüber dem repetitiven GAATG-Motiv blieb jedoch schwer fassbar. Als Hypothese dient, dass die RUS ihre Ziele während der Transkription erkennt. Dafür konnte ein konserviertes Motiv in der nicht gespleißten Form der RUS gefunden werden, das mit den repetitiven Regionen einen RNA-DNA-Triplex ausbildete. Dieselbe Stelle kann aber auch mit den Transkripten aus repetitiven Regionen RNA-RNA-Doppelstränge bilden.
Da RUS im Hippocampus am höchsten exprimiert ist, wurde angenommenen, dass RUS eine aktivierende Funktion auf Smarca5 ausübt. CTCF und Smarca5 ChIP Experimente kombiniert mit quantitativer RT-PCR in ctrl, RUS und Smarca5 KD Zellen bestätigten diese Annahme. Wohingegen der Verlust von Smarca5 zu einer reduzierten Smarca5-Bindung und somit auch zu einer verminderten CTCF-Bindung an allen getesteten CTCF Bindestellen führte, führte der RUS Verlust spezifisch nur an RUS Bindestellen zu einer verminderten CTCF-Bindung. Die Bindung von Smarca5 blieb nach dem RUS KD erhalten. RUS aktiviert möglicherweise die Chromtin-Remodellier-Aktivität von Smarca5 ohne dessen Chromatin Bindung zu regulieren. Dadurch stellt die ortsspezifische Aktivierung des Chromatin-Remodellierens durch lncRNAs einen neuen, vielseitigen und noch nicht beschriebenen Mechanismus für den Organismus dar, um Chromatin zeitlich und lokal spezifisch zu organisieren.

Durch die Wechselwirkung mit Brd2 und Brd4 und durch das Zufügen neuer CTCF Stellen, können lncRNAs möglicherweise Enhancer-Promotoren-Schleifen, neue Insulator-Barrieren aufbauen als auch die zellspezifische Chromatin-Organization regulieren, um so ihre regulatorische Funktion auszuüben. Weitere biochemische Experimente sowie die Anwendung neuer Färbetechnik wären nötig, um die Aktivierung des Chromtin-

Remodellierens und das Zusammenspiel der RUS mit ihren Interaktionspartnern zu beschreiben.

2.2 Summary

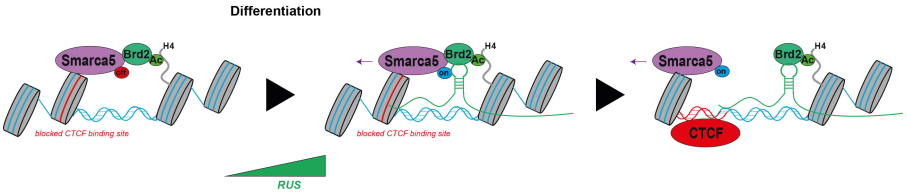


Figure 2: **Graphical abstract:** The lncRNA RUS binds to Brd2 occupied regions and activates the remodeling activity of Smarca5 to induces CTCF binding and chromatin reorganization.

During evolution, the mammalian central nervous system has increased in complexity. Likewise, the number of a new class of RNA Pol II transcripts, defined as long non-coding RNAs (lncRNAs), increased. Therefore, lncRNAs are thought to be responsible for the massive increase in brain size and the diversification of neuronal cell types [1]. LncRNAs are often expressed in a tissue-specific manner and act as an additional layer to increase gene regulation complexity. However, the biological function of the vast majority of neuron-specific lncRNAs is still unknown. This work aimed to search from different transcriptome datasets for conserved functionally relevant lncRNAs and describe their molecular mechanism. To test potential candidates' putative neurogenic functions, we knocked down (KD) their expression in differentiating neural stem cells (NSC). The obtained phenotypes were analyzed using a newly developed algorithm.

One candidate whose knockdown (KD) significantly reduced the number of neurons was a lncRNA located upstream to the neurodevelopmental gene Slitrk3, which we named RUS (RNA upstream to Slitrk3) based on this. A detailed assessment of the KD phenotype by histochemical staining and RNA sequencing showed that the KD of RUS resulted in a decreased cell proliferation and an arrest of cell differentiation, concomitant with an increase in cell death. RUS is enriched in the nucleus and conserved in mice and humans, particularly at its 5' end. In a rescue experiment, the overexpression of RUS in KD cells restored neuron formation. However, overexpression of a deletion construct of RUS lacking the conserved 5' domain failed to rescue the KD phenotype. These results suggested that the 5' conserved domain is important for lncRNA function and serves as a binding site for gene regulatory proteins.

To isolate interacting proteins and determine them by LC-MS, a new *in vivo* MS2-based RNA affinity purification was developed. For this, MS2-tagged RNAs were stably integrated and overexpressed in Neuro2A cells. To determine the proteins specifically binding to the conserved domain at the 5' end, I purified proteins binding to the RUS, the 5' deletion mutant, and the domain itself. The three constructs' interacting-proteins were compared to reveal that RUS interacts via its 5' end with the Bet proteins (Bromod-

omain and Extra-Terminal motif) Brd2, Brd4, the ISWI (Imitation SWItch) chromatin remodeler Smarca5 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5). Brd2 and Smarca5 preferentially act at DNA sites bound by the insulator protein CTCF [2–4]. ChIRP-Seq (Chromatin Isolation by RNA purification coupled with high throughput sequencing) experiments against RUS in Wt cells and ChIP-Seq (Chromatin Immunoprecipitation) experiments against Brd2 in ctrl and RUS KD cells showed that RUS and Brd2 bound together at 13 hippocampal-specific CTCF sites in the genome, rich in GAATG repeats. Here, RUS exerted neither a recruiting nor a repulsive effect on Brd2. However, the specificity of RUS toward the repetitive GAATG motif remains elusive. We hypothesized that RUS recognizes its targets during transcription. In line with this, a conserved motif in the non-spliced form of RUS was found, which should form an RNA-DNA triplex with the repetitive regions. However, the same site can also form RNA-RNA double strands with the transcripts from repetitive regions.

Since RUS is most highly expressed in the hippocampus, it was assumed that RUS exerts an activating function on Smarca5. CTCF and Smarca5 ChIP experiments combined with quantitative RT-PCR in ctrl, RUS, and Smacra5 KD cells confirmed this assumption. Whereas a loss of Smarca5 resulted in a reduced Smarca5 binding and a reduced CTCF binding at all CTCF binding sites tested, RUS loss resulted in reduced CTCF binding selectively at RUS binding sites. The binding of Smarca5 after the RUS KD was preserved. RUS possibly activates the chromatin remodeling activity of Smarca5 without regulating its chromatin binding. Thus, site-specific activation of chromatin remodeling by lncRNAs represents a novel and yet undescribed mechanism for the organism to specifically organize chromatin temporally and locally.
By interacting with Brd2 and Brd4 and adding new CTCF sites, lncRNAs may be able to establish enhancer-promoter loops, new insulator barriers as well as to regulate cell-specific chromatin organization to exert their regulatory function. Further biochemical experiments and the application of new staining techniques would be needed to confirm the activation of chromatin remodeling and to determine the interplay of RUS with its interacting partners.

3 Introduction

3.1 Neurons

Neurons are our sensory cells in our central and peripheral nervous systems, transmitting and processing information. Thereby, neurons can be categorized into sensory-, motoric-, interneurons neurons and neurons of the brain. A neuron itself can be divided into its different compartments like the cell body called soma, branched structures called dendrites, and axons. Axons are long fibers that can reach up to 1 m in length in the human body. Dendrites are shorter fibers typically located close by the soma (Figure 3) [5, 6].
Neurons form complex circuits to process all sensory, motoric, and cognitive tasks. Thereby, neurons are interconnected between the synapsis protruding from the axon and dendrites and communicate by chemical transmission involving the release of small chemical compounds defined as neurotransmitters. Once an axon is stimulated, an action potential moves towards the axonal end. Arrived at the synapsis, vesicle storing the neurotransmitters fuse with the synaptic plasma membrane resulting in a neurotransmitter. After release, receptors at the postsynaptic membrane of the downstream dendrites are stimulated by the neurotransmitter. Most neurotransmitter receptors are ligand-gated ion channels that either open or close after stimulation. Whether an ion channel opens or closes depends primarily on the neurotransmitter. The release of ions by channel opening results in depolarization of the postsynaptic membrane. The postneuron gets excited. Vice versa, ion channel closing increases intracellular ion concentration. That results in a hyperpolarization of the postsynaptic membrane and postneuron inhibition. Thus the releasing neurotransmitter determines whether a synapse acts excitatory or inhibitory primarily depends on the releasing neurotransmitter. Several neurotransmitters, such as glutamate, dopamine, serotonin, and acetylcholine, are distinct for neuronal subtypes [5]. Furthermore, neuronal subtypes differ not only by their releasing neurotransmitter but also by their morphology and anatomical location. Thereby, each neuronal subtype is integrated into its defined network and circuit.
During evolution, the central nervous system increased in volume and became more complex, and more anatomical substructures evolved. Thus, not only the total number of neurons but also the number of neuronal subtypes increased massively. The formation of the plethora of different types of neurons is orchestrated by complex genetic circuits, gene -regulatory, and signaling pathways that evolved simultaneously during evolution [7].

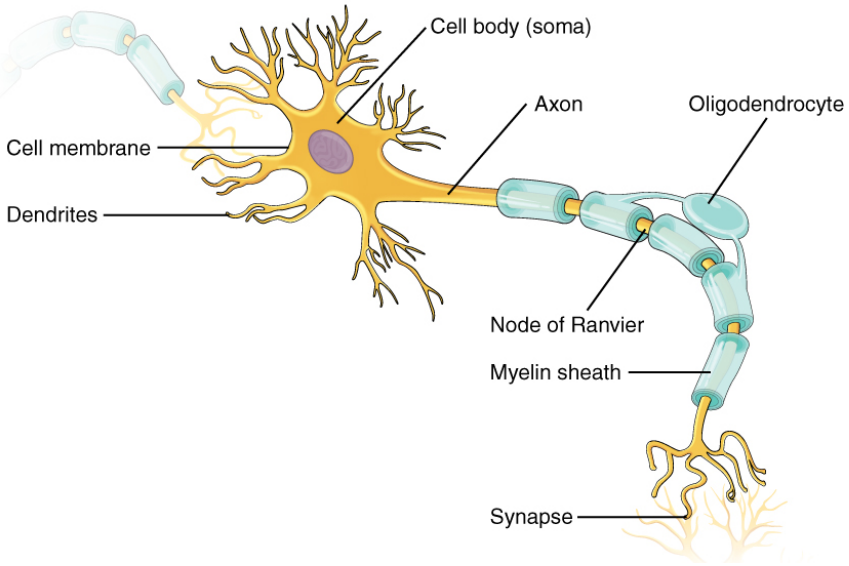


Figure 3: Schematic structur of a neuron insulated by an oligodendrocyte. Picture depicted from [5]

3.2 Central nervous development

Embryonic nervous system development starts when the neural tube is formed from the ectoderm [8]. The neural tube subdivides into distinct regions: the forebrain, midbrain, hindbrain, and spinal cord. The neural tube is harboring the neuroepithelia cells – the first neural progenitors. These cells are highly polarized along their apical-basal axis and interconnected by tight – and adherens junction [9]. The maintenance of the pool of neural progenitors and neurons' formation is regulated by symmetric – and asymmetric cell division. Symmetric division of a progenitor cell either produce two daughter progenitor cells or two daughter neurons. The progenitor cell's asymmetric division produces one progenitor daughter and one neuron daughter or two progenitor daughter cells. Although both progenitor cells produced by asymmetric division are self-renewal, one daughter cell differs by its molecular and morphological identity from its mother cell. Thus, asymmetric cell division of precursor cells results in different subtypes of neural precursors and neurons. Those precursor cells differ not only by their morphology but also by their neuronal lineage commitment [9, 10]. The asymmetric division of neuroepithelia cells (NE) produces the first pioneer neurons forming the subplate and Caja-Retzius neurons forming the marginal zone (Figure 4). Besides neuron formation, the asymmetric division of NE cells produces apical radial glial cells. Formation of radial glia cells is happening exhausting in rodents at E12.5 [11, 12]. Apical radial glia cells maintain attachment to the apical surface; however, they lose tight junctions and adherens junction and express typical marks of astrocytes as GFAP and Glast1.
During embryonic-neurogenesis, newly formed neurons migrate to the basal lamina using radial glial cells as a scaffold. Migration is inhibited at the marginal zone by Caja-Retzius neurons producing the migrating inhibitory factor relin [13]. Apical radial glia cells further divide into radial glia cells and outer radial glia cells, which lose the connec-

tion to the apical surface. Neurons produced from radial glial cells migrate towards the marginal surface to create layers IV and V. During the late-phase of embryonic neurogenesis, radial glial cells and residual NE cells divide into basal progenitors (BP) [9]. BP cells are forming the subventricular zone (SVZ) and maintain self-renewal properties. Intriguingly, the symmetric division of one BPs to two daughter neurons massively increases the number of neurons that migrate to layer IV and produce the layers II-III [14]. After completing the migration, neurons further mature by forming one long axon and branched dendrites connected to other neurons' axons by synapsis formation regulated by repulsive and attractive cues. However, at the late stage of embryonic neurogenesis, many neurons are displaced by programmed cell death, and massive synapse pruning occurs [15]. Programmed cell death is considered to happen by neurons competing for neurotrophic factors. Only neurons established proper connections, obtain more neurotrophins, and are protected. Thus, programmed cell death plays an essential role in the establishment of functional neural circuits. This theory was raised by Oppenheim in 1989 and is called the neurotrophic theory. After embryonic neurogenesis is completed, neural precursor cells also differentiate into astrocytes and oligodendrocytes [10].

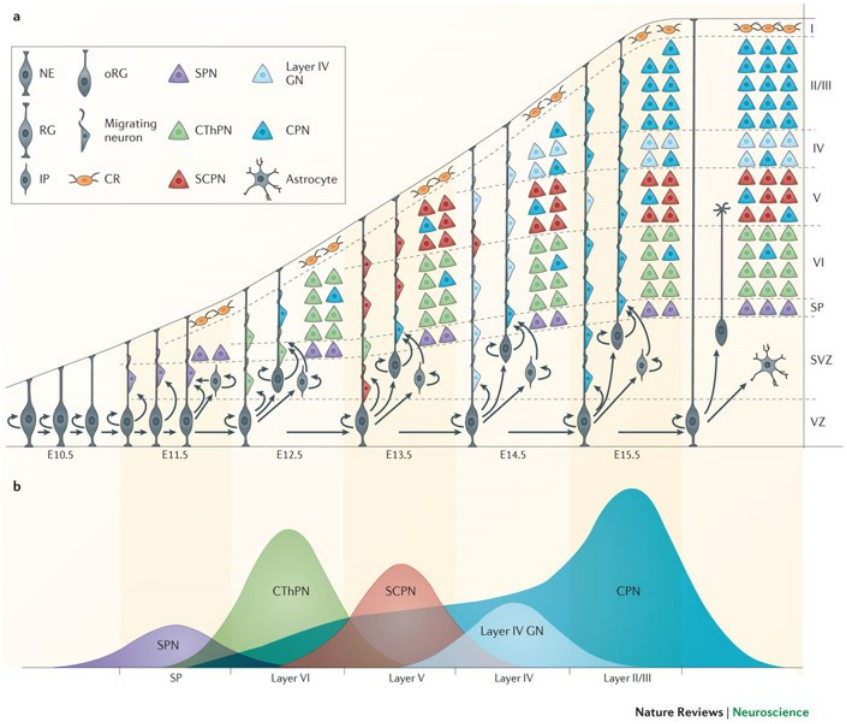


Figure 4: 6-layer neocortex formation: the layered structure of the neocortex is a consequence of multiple rounds of symmetric and asymmetric cell division producing different subtypes of neural progenitor cells and neurons. Neurons migrate on the scaffold of apical radial glia cells towards layer I (marginal zone) where migration is inhibited by relin producing Caja-Retzius neurons. The figure is depicted from [10].

3.3 Recapitulation neuron formation ex vivo

Studying the mechanism of neuron formation and maturation on the molecular level requires the recapitulation of the formation of neurons *ex vivo*. Therefore, three cell models are routinely used: stem-cell-derived neurons, neural-stem-cell-derived neurons, and cultured neurons.
Embryonic stem cells (ESC) and induced pluripotent stem cells (iPS) are self-renewal and pluripotent – meaning that they can differentiate into any cell-type. In contrast to ESC, pluripotency is artificially induced in iPS. Therefore, somatic cells from mice or human donors are reprogrammed by exogenous expression of the key pluripotent factors OCT3/4, SOX2, KLF4, and c-MYC [16]. IPS and ESC can be infinitely cultured in presences of anti-differentiation factors. In the absence of anti-differentiation factors, stem cells aggregate to embryonic bodies. Cells of embryonic bodies lose their self-renewal and pluripotent properties and start to differentiate. Thereby, embryonic bodies show characteristic features of post-implantation, gastrulation, and early organogenesis [17,18]. Treatment of embryonic bodies with retinoic acid or other stimuli favors the differentiation into the neuroectodermal lineage [19]. Stem cells derived neural progenitors are isolated by cell singularizing and kept either as monolayer culture or neural rosettes. Different pathways regulate progenitors' proliferation during neurogenesis in a stage and region-specific context, determining the outcome of neuron type during final differentiation [20]. Thus, a defined cocktail of signaling molecules must be administered to culture to prime the progenitors' developmental and regiospecific identity for the desired neuron outcome. Thus, stem cell-derived neuron models are showing all hallmarks of ectodermal and neuronal development. Those hallmarks range from Sox2 and nestin-positive progenitors to β-tubulin and Mapt positive newly formed neurons to fully matured Map2 positive neurons showing properly arranged axons, dendrites, and synapsis. Thus, this cell model is adequate to study a gene product's phenotype in all steps of neuron differentiation. However, this model is highly labor-intensive, making screening approaches highly time-intensive.

The priming step needed for stem-cell models can be omitted for neural-stem-cell-derived neurons and cultured neurons. Regio- and developmental identity can be defined by regiospecific isolation of progenitor cells at a distinct developmental stage – typically between E14-E17. Neural stem cells can be isolated from embryonic and adult neural tissue and cultured either as monolayer or neurospheres [21]. For neurosphere-culture, singularized neural stem cells are kept in free-floating conditions to form spheres, which similars embryonic body culture. For monolayer cultures, neural stem cells are cultured on coated culture dishes. In both models, proliferative and self-renewal properties are maintained by supplementing the medium with basic FGF. Post-mitotic neuron differentiation is initiated in the absence of FGF on culture dishes. Neurospheres and monolayer neural stem cell cultures are less labor-intensive. However, they only recapitulate only neuron formation. Neuron maturation and synapse formation are hard to detect [21].

Both models are adequate to study the function of a gene in the context of the transition from progenitor to neuron but not maturation.

For neuron culture systems, neural progenitors are plated on culture dishes subsequent after isolation. Cells are maintained in serum- and growth factor-free medium [22]. The change of neural precursor to neurons is happening directly after seeding. However, neurons can be kept for a long time in culture and show all neuron maturation features as the formation of axons, dendrites, and synapsis. Thus, this model can be applied to study the function of a gene product in neuron maturation but hardly in the context of neuron formation.

3.4 Transcription

In all kingdom of life, the genetic information is stored as double-stranded DNA. DNA is transcribed to RNA, and a small fraction of RNA called messenger RNAs (mRNAs) are translated into proteins [23]. The process of transcription and translation is defined as gene-expression. In multi-cellular organisms, however, not all genes are expressed simultaneously. Several genes are highly spatiotemporal regulated such as developmental genes and cell-type-specific genes. The expression state of genes defines the cellular identity and the property of self-renewal and differentiation. Additionally, physiological changes affect gene-expression to enable the cell to respond to stimuli.

Transcription of protein-coding and regulatory non-coding genes is carried out by an enzyme complex called RNA polymerase II (RNA-Pol II). This complex recruitment to the transcription start site (TSS) of a gene is regulated by *cis* regulatory DNA elements and *trans* acting protein factors (TFs) binding to the *cis* elements. Transcriptional regulatory *cis* elements can be distinguished in promoters located adjacent to the transcription start site and enhancers located more distant to the transcription start site. Although enhancers may be located up to several MB away from the TSS, both are in close vicinity organized in a loop-like structure [24].
TFs can be distinguished in general TFs (GTFs) regulating the expression of all genes and gene-specific TFs regulating the expression only of a subset of genes. Gene-specific TFs act in a sequence-specific manner by binding to defined as the consensus sequence. Depending on the genes they regulate, gene-specific TFs can act as master-regulators for pluripotency and cell fate commitment. For the most TFs, the consensus is between 6-12bp [25]. Due to the small size, the TF consensus sequences occur more in the genome than the TF actually bind, defined as futility theorem [26]. This observation is explained by the fact that TFs are binding to DNA in combination with other TFs synergistically [25]. Whereby GTFs act primarily on promoter regions, gene-specific TFs also bind to enhancer regions [25].

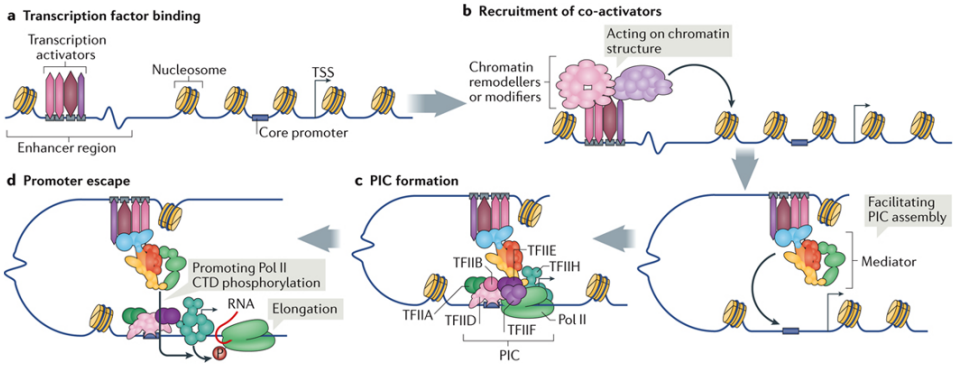


Figure 5: Recruitment of the transcriptional machinery (RNA-Pol II) by the Mediator complex. Depicted from [24].

On enhancer regions, TFs recruit a protein-complex called mediator (Figure 5). This complex comprises 25 subunits organized in a head, a middle, and a tail domain and bridges enhancer and promoter regions to facilitate RNA-Pol II binding at the TSS. Therefore, the tail domain binds to the TF-occupied enhancer, and the head domain initiates preinitiation complex (PIC) formation -also known as basal machinery – on the opposite promoter region. In doing so, the mediator only transiently binds the PIC by recruiting the GTFs: TBP and TFIIB [27, 28]. TBP and TFIIB bind to a TATAWAW motif upstream to the TSS defined as TATA-box and bend the DNA to allow RNA-Pol II holoenzyme binding [29]. After RNA-Pol II binding, the GTFs TFIIE, TFIIE, and TFIIH bind to initiate RNA Pol-II elongation [30]. The whole PIC comprises 6 GTFs and the RNA Polymerase II holoenzyme built by 12 subunits. The entire process is precisely timely regulated by the dynamic assembly of the mediator/PIC complex and post-translational modifications such as the phosphorylation of the C-terminal domain of RNA-Pol II by TFIIH facilitating RNA-Pol II elongation [24].

3.5 Chromatin

In eukaryotes, DNA is linear, segmented in different chromosomes, and organized as chromatin. Thereby, DNA is wrapped around histone proteins' octamers to form the chromatin (Figure 6). Histone proteins are basic proteins with a C-terminal α-helical and a flexible N-terminus extruding from the nucleosome. The main histone proteins: H2A, H2B, H3, and H4, form an equivalent octamer [31]. Each octamer encircles 146 bp of DNA corresponding to 1.65 turns to form the nucleosomes [31, 32]. Nucleosomes are linked to each other in a bead on a string structure – known as euchromatin. Nucleosome packing reduces the length of DNA to facilitate dense packing of eukaryotic DNA into the nucleus.

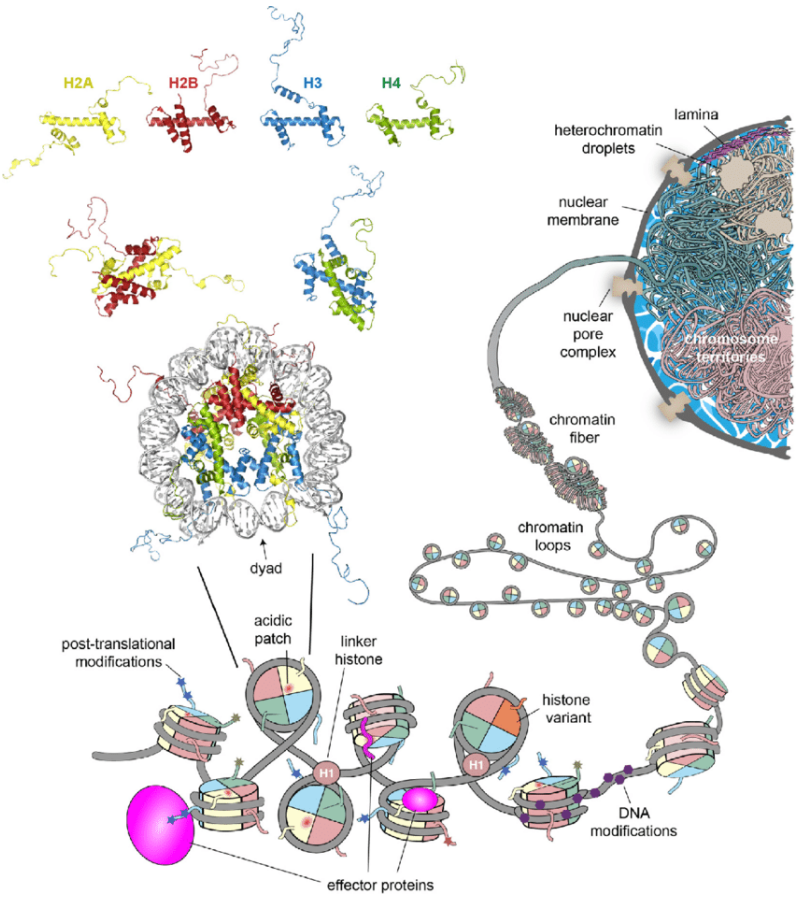


Figure 6: Organization of the chromatin. Figure depicted from [33].

In general, chromatin can be distinguished in compartment A, and compartment B. Compartment A is enriched for the euchromatin. Euchromatin is less-dense packed chromatin and more accessible for transcriptional machinery. Compartment A is localized centrally in the cell nucleus around speckles and paraspeckles- organelles where splicing and RNAs' storage occurs [34]. Compartment B is enriched for heterochromatin. Heterochromatin formation is induced by the linker protein HP1 condensing chromatin to highly packed fibers [35, 36]. Heterochromatin is inaccessible for the transcriptional machinery and silent. Compartment B is located either around the nucleolus – an organelle where rRNA transcription and processing occurs – or the nuclear envelope.

The next level of three-dimensional organization is generated by intrachromosomal and also interchromosomal interactions. This higher-level organization is not static, spatiotemporally regulated, and differs between cell types, thereby giving the cell a nuclear architecture required for cell-specific gene programs.

Intrachromosomal interactions are contacts of genomic regions within the same chromosomes. It is suggested that a loop extrusion mechanism organizes the chromosome into many dynamic loops as locally packed chromosome regions called Topologically As-

sociated Domains (TADs, Figure 7). During loop extrusion, the protein complex cohesin forms a ring shape structure, binds, and moves on chromatin under ATP hydrolysis. This movement along DNA extrudes the DNA resulting in loop-like structures as TADs that promote promoter-enhancer interaction and insulate neighboring transcriptional loci [37]. Looping of enhancers and promoters facilitates that different enhancers regulate the same promoter or the same enhancer regulates different promoters resulting in a fine-tuned expression and increased gene-regulation complexity [38].

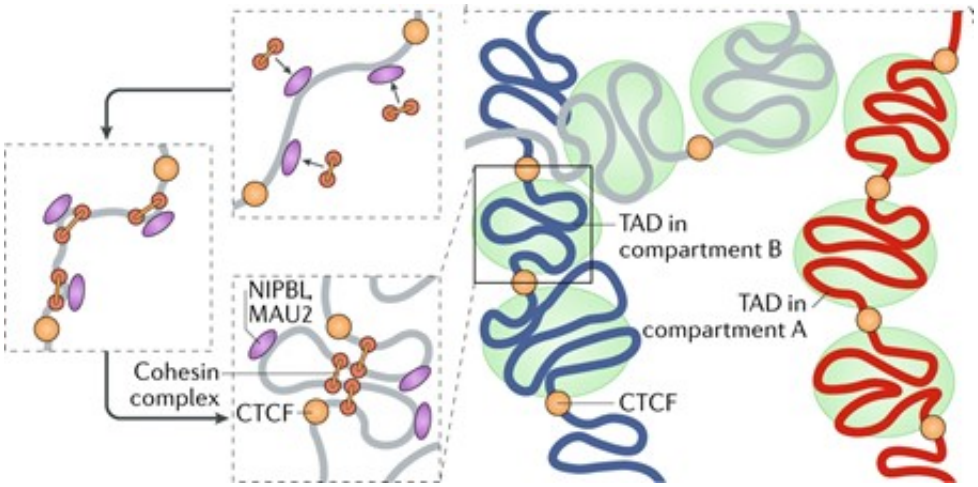


Figure 7: Formation of TADs. Figure modified from [38].

The flexible linker DNA connecting nucleosomes acts as a binding site for gene-regulatory proteins as transcription factors and CTCF-binding factor. CTCF has multiple functions, e.g., it acts as an insulator protein and is required for intrachromosomal contacts ranging from 100kbp to several Mb. In the context of the loop extrusion mechanism, cohesin's movement on DNA is blocked when cohesin anchors to the DNA-binding insulator protein CTCF [37, 39]. Thereby CTCF is involved in the boundary formation between TADs [40].
Additionally, genomic regions from different chromosomes spatial cluster to form inter-chromosomal contacts. Intriguingly, promoter regions of transcriptional active pluripotent factors are spatially clustered by the pluripotent factors Nanog and Oct4 in ESC [41]. In the same cells, repressed developmental genes as the Hox transcription factors are spatially clustered by the polycomb-protein-complex 1 [42]. These observations indicate that the genomic regions are not randomly located in the nucleus merely but rather distinct organized in the nucleus.

The regulatory mechanisms of the formation and resolution of chromosomal contacts are not fully understood. However, chromatin-remodeling and epigenetic marks are playing a pivotal role [43, 44]. Thereby, the regulatory mechanisms are not hierarchical ordered but rather interconnected and implemented in gene-regulatory circuits.

3.5.1 Epigenetics

Epigenetic describes the *epi* information that is not propagated by DNA replication. Rather, the organism uses a code of chemical modifications of the flexible N-termini of histone proteins. In addition to histone modification, DNA is epigenetically modified by methylation.
In contrast to DNA, epigenetic information can be edited and erased. Thereby, the chromatin act as a whiteboard for writers, readers, and erasers. Enzymes adding a chemical modification to a histone tail are considered as writers. Enzymes removing a modification are considered as erasers. Gene-regulatory protein and protein-complexes as chromatin remodelers recognizing specific histone modifications are considered as readers.
Thereby, epigenetic modifications regulate DNA accessibility towards transcriptional machinery and transcription factors and the formation of TADs. The deposition and removal of epigenetic marks are highly dynamic and specifically regulated during development. Like genomic information, epigenetic information can be propagated during replication [45].

DNA methylation: Methylation of DNA is catalyzed by DNMT 1, 3a & b. Methylated DNA is predominantly found at GpC island around the transcription start sites (TSS) or silenced heterochromatin. Hypermethyltated genes are transcriptionally inactivated. Whereby DNMT3a/b are responsible for de novo methylation, DNMT1 methylates hemimethylated CpG island produced during S-phase. Methyl groups are either actively removed by Tet-proteins or passively removed by inhibition of DNMT1 during the cell cycle. ESCs exhibit hypomethylated chromatin due to reduced DNMT enzyme levels and increased Tet protein expression. Methylation is gained again during development [46].
Histone modifications are much more complex and tightly regulated than DNA methylation. There are four canonical histone proteins, and each of them contains several amino-acids at the flexible N-terminus that can be modified. To date, not all modifications are discovered. Especially, the reactive ϵ amino-group of lysines can be easily methylated or acetylated. Methylated lysines are recognized by proteins containing a Tudor or chromodomain. Acetylated lysines are recognized by proteins containing a bromodomain [47]. Thereby, the histone tail: H2A, H2B, H3, H4, the amino-acid, and position are important. The following histone marks are involved in transcriptional activation and repression in euchromatin:

H3K4Me3: The trimethylation of H3K4 is predominately found close to the TSS of genes. H3K4Me3 is deposited by Set1 methyltransferase [48, 49]. H3K4me3 around the TSS is predominantly associated with active genes and is read by the chromatin

remodeler CHD1 to open the chromatin for transcription [50]. The demethylase LSD1 removes the trimethyl mark on H3K4me to silence gene-expression [51].

H3K4Me1: Mll3 and Mll4 catalyze monomethylation of H3K4. Monomethylated H3K4 are predominantly found at promoters and distant regulatory elements as enhancers [44, 52, 53]. H3K4me1 around promoters interfere with H3K4me3 and recruits the histone deacetylase complex of Sin3a to repress transcription [54]. At enhancers, however, H3K4 methylation has a priming and activation function [44, 53].

H3K27Ac: Acetylation of H3K27 is catalyzed by Histone acetyltransferase (HATs) and CBP-p300 [55]. H3K27Ac marks have transcriptional activating functions on promoters and enhancers and are removed by histone deacetylase HDACs, Sin3a, and Sirtuins at telomers [44, 56].

The following histone marks are predominantly found in heterochromatin. In contrast to other histone modifications, heterochromatin marks are not deposited by single deposition events but spread over chromatin. Bounding of chromatin by the insulator protein CTCF blocks heterochromatin spreading [57]. Telomeric and pericentromeric regions are constitutively packed in heterochromatin and are marked by H3K9me2 & 3. Other genomic regions are only transiently packed to heterochromatin, such as temporally silenced developmental genes. These regions are defined as facultative heterochromatin.

H3K27Me3: H3K27me is found at promoter regions, enhancer regions, and the inactivated X-chromosome (Xi). Thereby H3K27Me3 is involved in the formation of heterochromatin and inactivates developmental genes in ESC. [58, 59]. Hence, H3K27Me3 is associated with facultative heterochromatin. The methyltransferase Ezh2 catalyzes trimethylation of H3K27. Ezh2 is one subunit of the multi-protein complex Polycomb Repressive Complex 2 (PRC2). Intriguingly, PRC2 has a high affinity towards H3K27Me3. Thus PRC2 stays bound at chromatin after H3K27Me3 deposition. This observation explains the spreading of H3K27Me3 and the propagation of H3K27Me3 during the cell-cycle [60]. Moreover, PCR2 has a low affinity for methylated DNA. Thus, it is suggested that methylated DNA inhibits PRC2 to demarcate facultative from constitutive heterochromatin [61].

H3K9Me2/3: Di and trimethylated H3K9 is deposited constitutively at centromeric regions and telomeric regions. However, H3K9Me2/3 is also found in gene bodies and intergenic regions. Both marks are involved in the formation of heterochromatin and silence transcription. The trimethylation of H3K9 is catalyzed by Suv39h and dimethylation of H3K9 by G9a histone methyltransferase [62, 63]. H3K9me2/3 marks are recognized by UHFR recruiting DMNT1. Thus, the DNA is methylated in H3K9me2/3 marked heterochromatin [64]. During heterochromatin formation, the linker protein HP1 acts as a reader for H3K9Me2/3 marks to condense H3K9Me2/3 marked chromatin into

highly-dense fibers [36].

3.5.2 Chromatin-remodeling

The chromatin is massively rearranged and reorganized by chromatin remodelers during cell-differentiation and lineage-commitment. Chromatin remodelers use ATP hydrolysis to slide, reject, edit, and assemble nucleosomes - happing during DNA replication [65,66]. Chromatin remodelers are multi-protein complexes with a core ATPase harboring DNA helicase activity (Figure 8). The composition of complexes is not static and varies. Several chromatin remodelers interact and recruit epigenetic- readers, writers, and erasers such as HDAC1 [67]. Thus, chromatin remodeling complexes are versatile tools to manipulate the chromatin structure by repositioning the nucleosomes to enable or prevent the binding of transcription factors, transcriptional machinery, and CTCF, and rewrite epigenetic marks. Chromatin remodelers act not DNA-sequence specific but rather their bound cofactors guide them. Major cofactors are transcription factors, epigenetic-readers, and other bound DNA-factors. However, the exact mechanism remains elusive and might be controlled by still unknown factors. Moreover, ATPase subunits of chromatin-remodeling complexes comprise a DNA- and an RNA-helicase domain [2,68]. That indicates that RNAs might modulate chromatin-remodeling activity.

The main four chromatin remodeling families include:

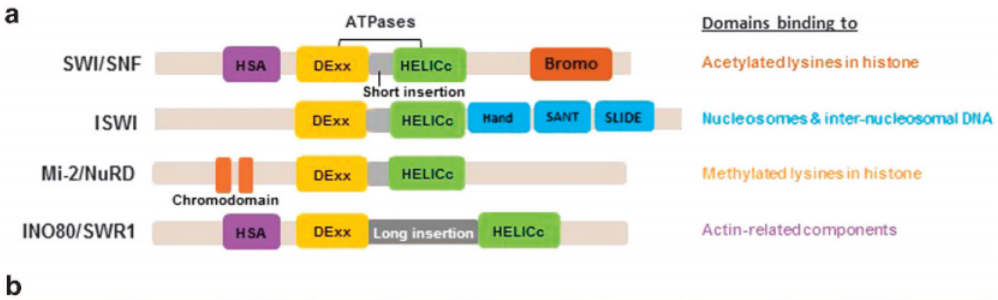
BAF: The BAF complexes are a subfamily of the SWI/SNF remodelers and include the ATPase Brg1 (Smarca4). BAF complexes are large protein complexes that can be differentially composed. The interaction with key transcription factors and other factors define the BAF complex's substrate-specificity. During differentiation and lineage commitment, BAF complex composition changes, resulting in a switched substrate specificity and enabling the BAF complex to execute key developmental task [69, 70]. As epigenetic readers, BAF complexes recognize acetylated and methylated histones at active promoters. Here, BAF complexes eject nucleosomes around the TATA box and TSS to facilitate the binding of the transcriptional machinery [71, 72].

NuRD: NurD complexes can be considered as epigenetic erasers. NurD complexes contain the ATPase CHD3/4 recognizing methylated lysines via the chromodomain and interact with HDAC1 to silence gene expression [67].

ISWI: ISWI -family contains the ATPase Snf2h (Smarca5) and Snf2l (Smarca1). In contrast to other chromatin remodelers, the histone modifications modulate Snf2h's specificity. Especially, the acidic amino acids of H2A and H2B, defined as the acidic patch, favors Snf2h's substrate specificity [73]. Snf2h is involved in ribosomal DNA silencing and DNA repair [74, 75]. Furthermore, Sn2h acts on CTCF bound regions and maintains CTCF binding [43]. Thus, Snf2h regulates TADs formation and chromatin

architecture [76].

INO80: The remodeling enzyme INO80 is involved in DNA-repair,activates the transcription, and targets preferentially YY1 occupied sites [77, 78].



b

	Human SWI/SNF		Human ISWI			Human Mi-2/NuRD	Human INO80		
	BAF	PBAF	NURF	ACF	CHRAC		INO80	SWR1	
								Tip60	SRCAP
ATPases	BRG/BRM	BRG	SNF2L	SNF2H	SNF2L	CHD3/CDH4	hINO80	P400	SRCAP
Other core subunits	BAF45 BAF47 BAF53 BAF57 BAF60 BAF155 BAF170 BAF250 BRD9 BCL7 BCL11	BAF45 BAF47 BAF53 BAF57 BAF60 BAF155 BAF170 BAF180 BAF200 BRG7	BPTF RbAP46 RbAP48	hACF1	BPTF RbAP46 RbAP48	HDAC1/2 MBD2/3 MTA1-3 RbAp46/48 LSD1	Tip49a Tip49b BAF53a Arp5,8 hles2 hles6	Tip49a Tip49b BAF53a TRRAP Tip60	Tip49a Tip49b BAF53a Arp6

Figure 8: **The main four families of chromatin remodeling complexes. A:** schematic view of core ATPases' structures including: HAS (helicase/SANT associated), Dexx (Dead-like helicase superfamily, RNA and DNA helicase activity), HELICc (helicase superfamily c-terminal), chromo (methylated lysine binding), bromo (acetylated lysin binding), and HAND-SANT-SLIDE (chromatin binding) domains. **B:** table of participating core subunits. Figure depicted from [79].

3.5.3 Activation of promoters enhancer activity by eRNAs

Enhancer and promoters genomic regions physically interact and are organized in TADs. Enhancers can be active, poised, and inactive. Thereby, H3K27Me3 marks inactive and poised enhancers, H3K4Me1 marks poised and active enhancers, and H3K27Ac marks active enhancers. During activation, active enhancers are bound by Bet (Bromodomain and Extraterminal) proteins Brd2 and Brd4 [80, 81]. Thereby, Brd2 corporates with CTCF at the promoter-enhancer contact and is supposed to insulate neighboring transcription loci [4]. Brd4 recruits RNA-Pol II complex to induce enhancer RNA (eRNA) expression [82]. eRNAs are not polyadenylated and less stable than other RNA-Pol II transcripts [81]. Co-transcriptionally, eRNAs recruit more Brd4 and the histone H3K27 acetyltransferase CBP-p300 to stabilize Brd4 occupancy and elevate H3K27Ac deposition [83, 84]. Increased H3K27 acetylation is accompanied by increased expression of

the promoter-driven target gene [81]. Besides, eRNAs recruits the mediator complex to enhancer regions [85]. Thus, eRNAs are required for enhancer activation and positively influence gene expression.

3.5.4 Repetitive Genome

More than two-thirds of our genome is repetitive [86]. Repetitive sequences can either be located adjacent to each other, spaced, or inverted. Two kinds of repetitive sequences can be distinguished: satellite DNA and interspersed repeats. Satellite DNA sequences are site-specific in the genome. Satellite DNA is categorized into satellite, minisatellite, and microsatellite DNA. Sattelite DNA includes five bp till 170 bp long sequences that span several kilobases. It is found at centromeres acting as binding sites for the chromosome segregation protein CENP-B during the cell cycle's anaphase. Minisatellites and microsatellites are short repeat units comprising 10-100 bp and 1-10 bp, respectively [87]. Microsatellites are supposed to be introduced by replication slippage of DNA Polymerase [88]. Satellite DNA is also found at telomers introduced by the enzyme telomerase to maintain chromosome integrity [89]. Besides centromere and telomers, satellite DNA is dispersed at subtelomeric and pericentromeric regions [87]. Interspersed repeats are randomly inserted regions by transposable elements. Transposable elements (TE) are saltatory genetic elements. They are distributed in the genome by a cut and paste mechanism mediated by transposons. DNA transposons are flanked by terminal inverted repeats and encode for transposase that cuts the TE acting at the inverted repeats and integrates the cut DNA to new genomic sites. Retrotransposons are working on the RNA level. RNA Polymerase III transcribes the TE of retrotransposons. Retrotransposons can be divided into LTR transposons (long terminal repeat) and non-LTR transposons comprising LINE (long interspersed nuclear element) and SINE (short interspersed nuclear element). TE elements of LTR Retrotransposons encodes for reverse transcriptase and integrase. After reverse transcription of TE-RNA into TE-DNA, integrase integrates the TE-DNA randomly into genomic DNA via the LTRs. LINEs encode for endonuclease nicking genomic DNA sequence-specific to generate a free 3'OH group that acts as a priming site for the reverse transcription reaction. SINE transposons encode for no enzymes and require the activity of LINE transposons to distribute tRNAs, Alu elements, and 7Sl-RNA in the genome [90]. Transposons are suggested as a driving force for rapid evolution. However, most transposons are repressed by eukaryotes. Repetitive elements are involved in chromosome-segregation and chromosome-integrity. Furthermore, they play a role in heterochromatin formation, three-dimensional folding of the genome, and regulation of the gene-expression [91, 92]. Genomic loci encoding for ribosomal RNA (rRNA) are also repetitively organized (Figure 9). Several repeats are distributed within different chromosomes, and each repeat contains multiple rRNA copies to enable the high expression of ribosomal RNAs. Each repeat copy is separated by a repetitive internal genomic spacer (IGS) acting as a promoter region. 28s, 18 s, and 5.8s rRNA are transcribed from the same precursor harboring internal transcription

spacer and external transcription spacer to separate the different rRNA transcripts [93].

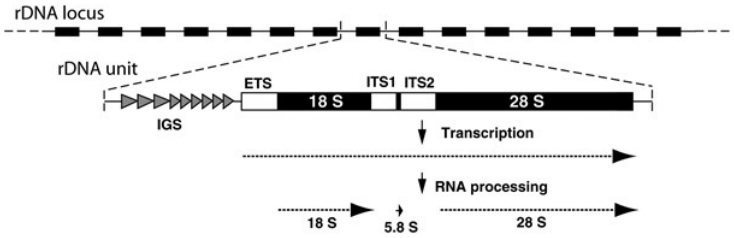


Figure 9: Organization of the rRNA locus: IGS: internal genomic spacer. ITS: internal transcript spacer. Figure depicted from [93]

3.6 LncRNAs

The central dogma of molecular biology teaches us that DNA encodes all the genetic information that, in turn, gets transcribed to mRNA, which gets translated to proteins [23]. Thereby, proteins are considered to exert all enzymatic and regulatory functions and serve as scaffold molecules. However, this protein-based perspective underestimates the role of RNAs. RNAs of the ribosome – a macromolecular complex of RNA and proteins – and tRNAs coupled to amino-acids carry out the main challenging cellular enzymatic reaction, meaning the translation of an mRNA into a protein [94,95]. Furthermore, RNAs are participating in the maturation process of RNA during transcription. Spliceosomal RNA-protein complexes (snRNP) enable the transesterification reactions by binding to the immature RNA via Watson Crick base pairing during splicing [96, 97]. Ongoing genetic studies explored that eukaryote gene-regulation is more regulated in a coding-independent fashion. For example, miRNAs control the expression of a gene on the mRNA level. MiRNAs are annealing to mRNA targets by Watson crick base pairing and recruit the RISC complex to degrade target mRNAs [98, 99].
Regarding the protein-based perspective of cell-biology, we have to keep in mind that only 2% of the mammalian genome is protein-coding [100]. However, much more than those coding regions are considered transcribed, comprising removed introns and the so-far known non-coding RNAs. For a long time, the scientific community considered the majority of DNA as non-transcribed junk. This view was drastically changed by the advent of next-generation sequencing technology, allowing the sequencing and quantifying of total transcriptomes (bulk RNA Seq). Up to 90% of our genome is transcribed. The majority of those novel found transcripts are longer than 200 nt, spliced, capped, and polyadenylated; however, they lack any protein-coding capacity [100]. Based on this, those transcripts are termed long non-coding RNAs (lncRNAs). Controversial, the abundance of most lncRNAs in bulk RNA-seq experiments was lower than the abundance of protein-coding mRNAs, arguing that lncRNAs might be rather unspecific transcribed junk. However, lncRNAs are more tissue and cell-type-specific expressed than protein-coding genes. Single-cell sequencing experiments revealed that cell-type-specific lncRNAs are abundantly expressed, and the low expression in bulk RNA-sequencing was a consequence of pooling different cell types [101]. Furthermore, functional and

biochemical studies have shown that lncRNAs are functional relevant by exerting important gene-regulatory functions on any level.
The mode of action is highly diverse between different lncRNAs and can be categorized by their subcellular localization. LncRNAs localized in the cytoplasm rather affect the expression on a post-transcriptional level, whereas lncRNAs localized in the nucleus rather affect mRNAs' maturation and regulation chromatin structure. Besides, lncRNAs can be distinguished by *cis* and *trans*-acting lncRNAs. Whereby *cis*-acting lncRNAs regulate the expression of genes in the transcription locus's genomic neighborhood, *trans*-acting lncRNAs regulate the expression of more distant genes.

3.6.1 Cytosolic lncRNAs

In the cytosol, *trans* acting lncRNAs regulate mRNAs' stability and protein translation at the ribosome. A special class of lncRNAs 1/2sbsRNAs (1/2-stau-binding sites) harbor an Alu element in the 5' end that annealing to complementary Alu elements in the 3'UTR of mRNAs. The RNA binding protein Staufen recognizes the long RNA duplex and triggers the target mRNA's decay [102].
Competing endogenous RNAs (ceRNAs) are long non-coding RNA that suppresses the miRNA pathway. CeRNAs are either linear or circular due to backward splicing and contain several miRNA binding sites [103].
The lncRNA treRNA forms an RNP (ribonucleic particle) with the proteins hnRNP-K, FXR1, FXR2, PUF60, and SF3B3 to suppress the translation of the epithelial marker E-cadherin at the ribosome during ETM transition in metastatic cancer [104].
Intriguingly, also lncRNAs acting in *cis* exist in the cytosol. BACE-AS is an antisense transcript of BACE and regulates BACE expression in a positive feedback loop [105].

3.6.2 Nuclear lncRNAs

Nuclear lncRNAs affect the maturation of RNA by regulation of splicing, transcription, and epigenetic modification of chromatin. Moreover, nuclear lncRNAs recently were described to affect the three-dimensional organization and nuclear localization of chromatin.

Chromatin associated lncRNAs are considered to function as scaffolds, recruit, or repress specific transcription factors, chromatin remodelers, or chromatin-modifying enzymes at distinct genomic loci. Interaction with chromatin-bound factors frequently targets *trans*-acting lncRNAs to their target sites. Besides forming Watson Crick base pairs between duplexes, purine-rich strands can anneal to a third strand by formation by Hoogsten base-pairing (Figure 10) [106]. Several *trans*-acting lncRNA bind to their genomic sites by RNA-DNA triplex formation, as observed for the lncRNA Meg3, that interacts with Ezh2 of the PRC2 (polycomb repressive complex 2) to suppress TGF-β pathway genes [107].

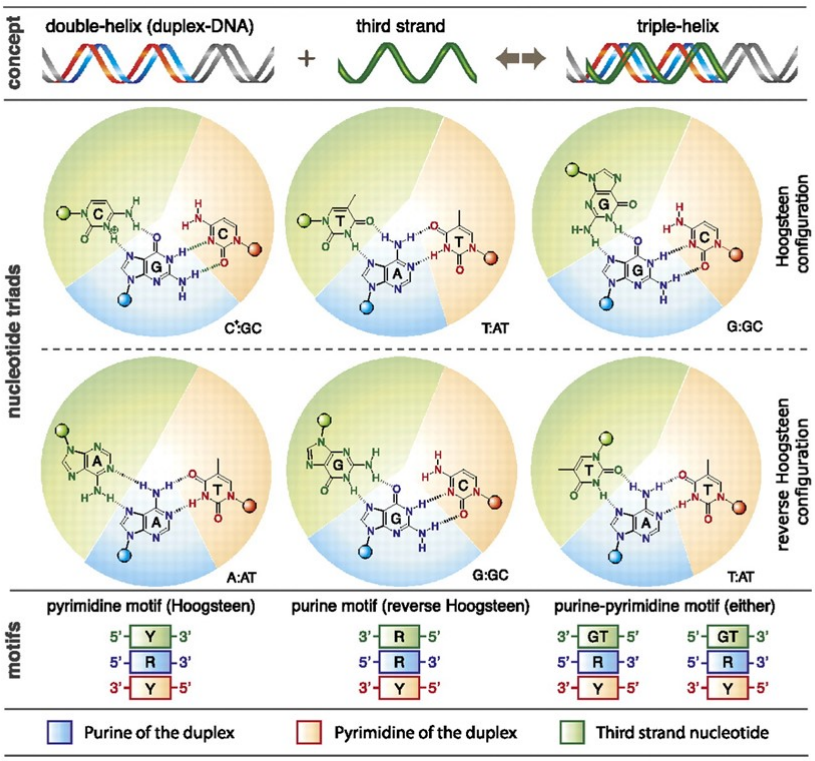


Figure 10: RNA-DNA Triplex formation by Hoogsten base pairing. Figure depicted from [108].

lncRNAs are highly advantageous for the organism to regulate gene expression. Due to their size, they can scaffold multiple protein complexes. Most underestimated, lncRNAs can act in contrast to proteins on the locus where they are transcribed. Every transcription factor, chromatin modifier, or other regulatory proteins is first translated in the cytosol and then translocated to the nucleus. Recent studies confirmed that chromatin-associated lncRNAs rather act nearby the locus where they are transcribed. Thereby chromatin-associated lncRNAs find their targets by proximity and affinity-guided mechanisms (Figure 11). With increasing abundance and affinity to scaffold regulatory proteins and form triplexes with distal genomic DNA, lncRNAs act more distant to their transcription locus. Besides, the length of the lncRNAs plays a role. Longer lncRNAs can scaffold with more and different proteins [109]. Hence, chromatin-associated lncRNAs can't be strictly divided into *cis* and *trans*-acting lncRNA. Moreover, the definition is rather continuous, which becomes more obvious by the lncRNAs Malat, Neat1, Hotair, and Xist.

Malat and Neat1 lncRNAs are highly abundant chromatin lncRNAs and scaffolds with multiple proteins with high affinity. Thereby, Malat and Neat1 affect alternative splicing and formation of stress granules and act primarily in *trans* [34, 110]. Due to RNAseP cleavage at the 3'end, Malat and Neat lack a polyA tail [111,112]. Whereby Neat1 binds to the 5' and 3' region of transcribed genes, Malat1 predominantly binds to the 3'end. Both lncRNAs interacts with serine- and arginine-rich proteins: PSF and SRSF1 and

other splicing and RNA maturation factors. Both lncRNAs share the most interacting proteins, but several interacting proteins are unique for Malat and Neat1, respectively [34]. The binding of Neat1 and Malat to its binding proteins is accompanied by phase separation resulting in membrane-less nuclear speckles that are visible under the microscope. Malat1 is localized in speckles suggested as a nuclear locus where splicing and alternative splicing occurs. Neat1 is found in paraspeckles, speckles that are located next to speckles. Paraspeckles size and numbers increase under stress, coinciding with increased Neat1 expression. Thus, paraspeckles might act as nuclear stress granules to stall transcription and store mRNA in order to decrease cytosolic protein load [110].

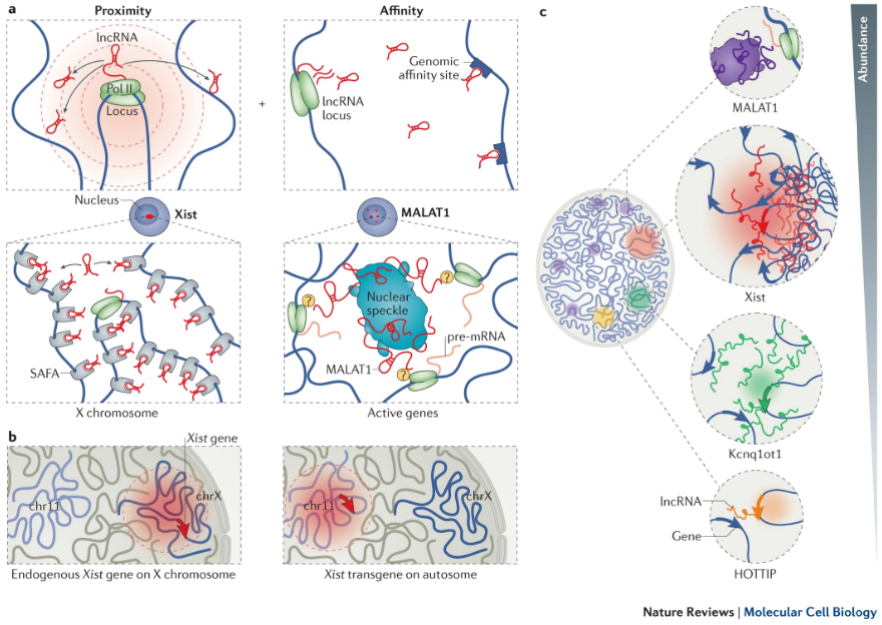


Figure 11: Proximity and affinity guided mechanisms determine the chromatin binding of lncRNAs as HOTTIP, Kcnq1ot1, Xist, and Malat. Pol II : RNA polymerase II. Figure depicted from [109].

Hotair is a lncRNA expressed from the HoxC locus and suppresses the expression genes from the HoxD locus during development [113]. The Hox gene clusters are gene loci encoding for developmental transcription factors required for body patterning [114]. Besides, Hotair is expressed in breast cancer to suppress tumor suppressor genes. Hotair interacts with Ezh2 of the PRC2 complex at the 5' end and with the LSD1-CoRest-Sin3A complex at the 3'end. Hotair is suggested to act primarily in *trans* guiding both complexes to its target sites to suppress the gene expression [115–117]. Although how Hotair is binding to its target sites is not well understood. It is worth to mention that Hox genes are spatially clustered in the nuclei of embryonic stem cells by PRC2 [42]. Thus, the HoxD locus and Ezh2 would be in close vicinity to Hotair's transcription site. Controversially, the knockdown of Ezh2 did not affect Hotair chromatin binding, indicating that Hotair bind to its target genes independent of Ezh2. Furthermore, Hotair binding

sites enrich an 11 nt polypurine motif suggesting that Hotair may form an RNA-DNA triplex [117].

Several lncRNAs also contribute to imprinting. In female embryonic stem cells, both X-chromosomes are active. During the first round of differentiation, one X-chromosome gets silenced by heterochromatin marks' deposition along the entire X-chromosome. In advance to silencing, the lncRNA Xist is expressed from the silenced X-chromosome (Xi) [118]. Xist acts enigmatically only at the Xi chromosome but spreads over the entire Xi chromosome. When Xist is knocked into an endogenous gene locus, the surrounding genes get silenced at a long-distance [119]. Xist harbors several repeat regions acting as binding sites for the SHARP deacetylase complex, PRC2, and hnRNP-U – a protein that binds DNA and RNA. These interactions are supposed to deposit heterochromatin marks. Besides, Xist interacts with cohesion in a repulsive manner to repress loop formation after silencing. Other bound factors, such as Lbr and CTCF, enable the spreading of Xist from one to the adjacent locus. After induction of Xist's expression, Xist acts in the sounding bound regions. Once a locus is silenced, the binding to Lbr enables the attachment of silenced heterochromatin to the nuclear envelope. Then, Xist spread to the next locus by exploiting the X-chromosome's 3-dimensional shape mediated by CTCF. CTCF is predominantly bound at the closure sites of DNA loops, so that novel Xist substrates are in proximity [68, 109, 120, 121].
Chromatin-associated lncRNAs also modulate the 3-dimensional organization of the nucleus. lncRNA Firre (functional intergenic-repeat RNA-element) is transcribed from the X-chromosome. In females, the transcription occurs from a stretch of Xi that escaped inactivation. Firre contains several 156 bp long repeat elements acting as a hnRNP-U binding platform to form several trans chromosomal contacts by a proximity guided mechanism. Loss of Firre resulted in a loss of transchromosomal contact sites [122].
Furthermore, chromatin-associated lncRNA regulates DNA stability. The lncRNA NORAD is activated by DNA damage and maintains chromosomal stability by interacting with RBMX. The interaction of NORAD and RMBX is essential to form the topoisomerase I RNP complex NARC1, which ensures proper chromosome segregation [123]. Moreover, the lncRNA TERC (telomerase RNA component) is expressed in the S-phase during the cell cycle. TERC serves as a primer-template for telomerase elongating telomeres to sustain chromosomal stability [124–126].

3.7 Methods to study lncRNA

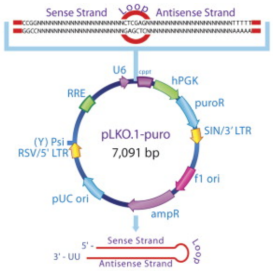


Figure 12: pLKO.1 vector system modified from [127].

3.7.1 shRNA mediated knock-down (KD)

Commonly, knockdown experiments are performed to analyze the loss of function phenotype. The most convenient knockdown approach represents the lentiviral transduction of small hairpin RNA (shRNAs). shRNAs mimicking miRNA are also processed by Drosha and Dicer and loaded onto the RISC complex to degrade the target RNA via Watson Crick base pairing [99]. Thus shRNA annealing to unique RNA sequences to avoid off-target effects can be easily designed by simple Watson Crick rules. Moreover, shRNAs can be packed in lentivirus to transduce cell types that are very sensitive towards transfection reagents (Figure 12). Since the RISC complex's selectivity for the sense or antisense strand can not be accurately predicted, several shRNA are generated in advance. Usually, five hairpins are used per target gene screened for efficacy together with a scrambled sh-RNA control [127].

3.7.2 RNA-affinity purification techniques

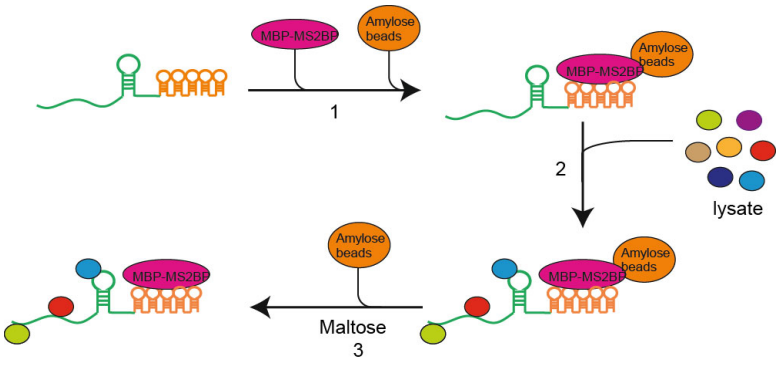


Figure 13: *In vitro* purification of RNA interacting proteins. 1) immobilization of RNA to amylose beads via MS2BP-MBP 2) pull-down of interacting proteins from cell lysate. 3) Elution of RNA protein complexes with maltose. Orange: stem-loop structure of the MS2 tag. Green: RNA of interest, colored dots: proteins.

Although several non-coding RNAs exert catalytical functions, lncRNAs do not. However, lncRNAs act as scaffolds for regulatory proteins and bind to other nucleic acids comprising genomic DNA, mRNAs, and miRNAs. Thus, understanding a lncRNA's mode of action requires determining proteins that interact with the lncRNA in a physiologically relevant manner. Furthermore, it is pivotal to reveal the target genes by isolation of bound DNA or RNA.
The common way to analyze interacting proteins of an RNA of interest comprises recombinant tagging, *in vitro* transcription of the RNA, and the pull-down of interacting proteins from cell lysate (Figure 13). Therefore, aptamers -binding nucleic acids - of phage capsid proteins such as MS2 are suitable affinity tags. The capsid protein MS2BP to pull down the tagged RNA can be recombinantly expressed and purified. Affinity tags, such as the maltose-binding protein (MBP), are tagged to the MS2BP to purify the protein and immobilized lncRNA protein complexes on amylose beads. Therefore, MS2 tagged *in-vitro* transcribed RNA is incubated with the recombinant capsid protein. The MS2 tag binds to MS2BP so that the RNA gets immobilized on amylose beads. Then the cell lysate is added, and interacting proteins bind to the RNA. After extensive washing, protein-RNA complexes can be eluted with maltose and analyzed either by western blot analysis or mass spectrometry [128]. The same method was transferred to cell culture. MS2 tagged RNA and the MS2 binding capsid protein tagged with biotinylation signal are cotransfected to mammalian cells. *In vivo* reconstituted RNA protein complexes are purified on streptavidin-coated beads. Streptavidin is a biotin-binding protein. The purified proteins can again be analyzed by western blot analysis or mass spectrometry [129].

3.7.3 Antisense oligonucleotide purification techniques

Other described purification methods exploit RNAs' property to bind nucleic acids by Watson Crick base-pairing (Figure 14). Antisense nucleotides hybridizing to the lncRNA can be readily designed and allow for endogenous lncRNA purification. By default, antisense oligonucleotides are equipped with a biotin linker to enable lncRNA complexes' purification on streptavidin beads. This method suits not only for the purification of interacting proteins but also for the purification of lncRNA bound RNAs or chromatin sites. It is pivotal for chromatin-bound lncRNAs to cross-link the cells and solubilize the chromatin by DNAse digestion or sonication in advance. Three conventional methods are described in the literature: CHART (capture hybridization analysis of RNA targets), ChIRP (chromatin isolation by RNA purification), and RAP (RNA antisense purification) [117, 120, 130]. These three methods differ primarily in the oligonucleotide design, used crosslinker, the DNA fragmentation, and the number of probe sets. Bound chromatin or RNA target sites can be quantified by quantitative RT-PCR or systematically analyzed by next-generation sequencing. Bound purified proteins are again analyzed by western blot analysis or systematically by mass spectrometry.

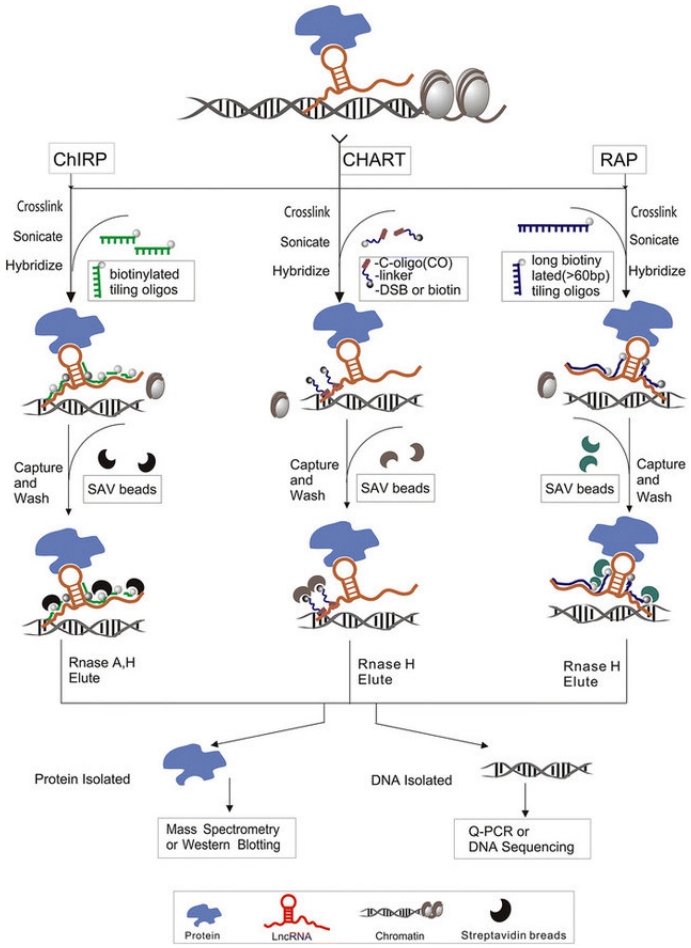


Figure 14: Scheme of antisense purifications: ChIRP, CHART & RAP. Figure depicted from [131]

CHART

The CHART method applies 25 nt long chemical synthesized biotinylated DNA oligonucleotides. Potential accessible binding sites for antisense oligonucleotides are determined by RNAseH sensitivity assay in advance and evaluated for off-target sites by *in silico* Blast anaylsis. Accessible sites showing no homology to other regions are used for probe design. In general, one probe set of different 25mer oligonucleotides are used for the purification of RNA-chromatin complexes from formaldehyde cross-linked and sonicated chromatin [130].

ChIRP

ChIRP method uses 20 nt long chemical synthesized biotinylated DNA oligonucleotides tilling the RNA of interest. By default, the design is that every 100 nt of the corresponding lncRNA is covered by one 20 nt probe showing no homology to other sites. Probes are divided into two probe sets: odd comprising probes with odd numbers and even comprising probes with even numbers. The experiment is performed with the odd

and even probe set separately. LncRNA chromatin complexes are purified from glutaraldehyde cross-linked sonicated chromatin. Only sites or proteins that co-purify from both probe sets are considered lncRNA interacting proteins or lncRNA target sites, respectively [117].

RAP

RAP method uses 90-120 nt long high-affinity biotinylated DNA oligonucleotides synthesized by molecular biological methods. Probes are designed to till the lncRNA of interest with a space of approximately of 15 nt length. Low complexity and repetitive probes are discarded. Probes can be divided into odd and even. This method was shown to succesfully purify chromatin-bound sites or bound RNA target sites of lncRNA. For the isolation of lncRNA bound chromatin sites, cells are cross-linked with di(N-succinimidyl)glutarate (DSG) and formaldehyde. Chromatin is fragmented by sonication and DNAseI digestion to sustain RNA integrity. For the isolation of lncRNA bound RNA sites, RNA-RNA hybrids are cross-linked in cells with psoralen before lysis [132, 133].

3.8 LncRNAs in neurogenesis

While the number of lncRNAs linearly increased, the quotient of brain mass and body mass, defined as the encephalization quotient, increased exponentially during evolution (Figure 15). Particularly, the central nervous system is expressing a high number of lncRNAs. The expression of most of them is restricted to the brain, particular brain regions, and specific neural subtypes. Neuron-specific lncRNAs are genomically located to neurodevelopmental regulators that are frequently co-expressed. Most notably, brain-specific lncRNA display higher sequence conservation between orthologues than lncRNAs expressed in other tissues [1].

lncRNAs act as an additional level to increase the complexity of gene regulation tremendously. lncRNA interferes with post-mitotic differentiation signaling to increase and diversify progenitors' pool on the one side and activates post-mitotic differentiation on the other side [3, 134, 135]. Thereby, lncRNAs accurately regulate the spatiotemporal behavior of self-renewal and differentiation properties of neuronal cells. Hence, the fast evolution of lncRNAs is directly connected to increased brain size and diversity of neuronal subtypes. However, only a small fraction of neuron-specific lncRNAs are investigated in more detail.

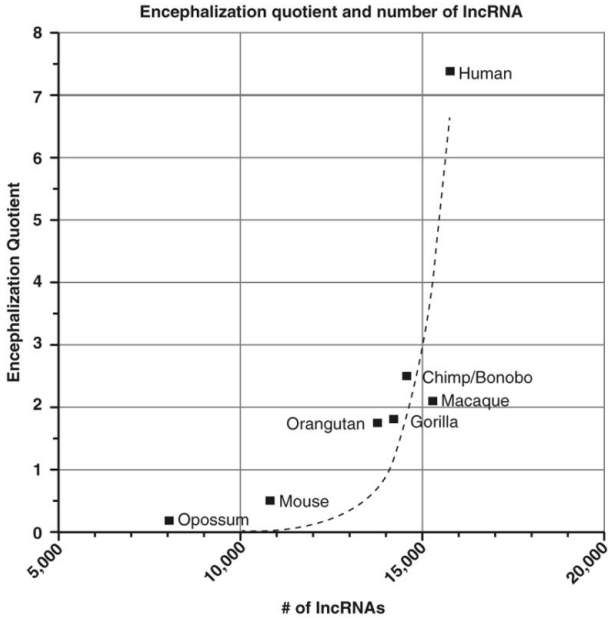


Figure 15: Correlation of encephalization quotient and the number of lncRNAs between taxes. Figure depicted from [1].

The lncRNA Tuna is expressed in embryonic stem cells and maintains pluripotency. KD of Tuna suppresses neuronal commitment. Tuna interacts with hnRNP-K, PTBP1, Ncl, and Nono and is found at the promoter sites of Nanog, Sox2, and Fgf4 [136].
The primate-specific lncRNA lncND acts as a miRNA sponge for mir 143, which suppresses the expression Notch1 and 2. Thereby, lncND maintain the expression of Notch1 and 2 to increase the pool of neural progenitors in primates [134].
The lncRNA RMST is regulated by Re-1 silencing transcription factor (Rest) and is expressed during neuronal differentiation. RMST interacts with Sox2 and is required to bind Sox2 at neurogenic transcription factors [137].
The lncRNA Evf2 interacts with the transcription factors Dlx1 at two conserved enhancers located between Dlx5 and Dlx6 to suppress the BAF chromatin remodeler ATPase subunit Brg1 during forebrain development. The inhibition of the remodeling activity of Brg1 by Evf2 accompanies the silencing of both enhancers [3].

4 Aim and goal of this study

The plethora of un-investigated lncRNAs indicates that we are still beginning to understand the function of lncRNAs in the context of neurogenesis. Hence, this work's primary aim was to identify novel neurogenic lncRNAs and describe the underlying gene-regulatory mechanism.
To select potential lncRNAs contributing to neurogenesis, I focused on lncRNAs whose expression is restricted to neurons to minimize the vast number of potential candidates on the one side and increase the chance to find a lncRNA with a strong KD phenotype

on the other side. Thereby, I assumed that the neuron-specific lncRNAs are required for neuron differentiation, neuron maturation, or proper neuron function. To accomplish this, I selected potential candidates from public available transcriptome datasets. I analyzed their function by lentiviral-transduced shRNA Knock Down (KD) in differentiating neural stem cells combined with immunofluorescence. LncRNAs are studied in more detail if their KDs strongly interfered with neuron formation. For a detailed description of the KD phenotype, RNA Seq experiments are performed to decipher how the lncRNA affects neuron formation and what genes and signaling pathways the lncRNA regulates.
Tissue and cell type-specific expressed lncRNAs often have key roles in gene regulation. On the molecular level, lncRNAs recruit or titrate gene-regulatory proteins, often ubiquitously expressed, to or away from their target genes. Thereby, lncRNAs target genes on the RNA level and chromatin level. I was particularly interested in chromatin-associated lncRNA and what gene-regulatory proteins are binding to selected lncRNA. To decipher this, I intended to establish a novel native affinity *in vivo* RNA-purification method to circumvent folding artifacts produced by *in vitro* transcription. Moreover, I aimed to unravel the genomic binding sites of selected lncRNA and whether selected lncRNAs have a titrating or recruiting effect on their interacting gene-regulatory proteins at their target sites. Therefore, I performed a combination of ChIRP Sequencing and ChIP experiments.
Since lncRNAs evolved rapidly during evolution, their molecular mechanism to affect gene-regulation is highly varying. Deciphering the molecular mechanism of selected lncRNAs biochemically, I aimed to uncover a still unknown RNA-based gene-regulatory mechanism. Linking the molecular to functional data enlightens the significance of studied lncRNAs' regulatory mechanisms for neuron formation. That ultimately improves our knowledge and comprehension of neurodevelopment from a non-coding perspective on the one side. On the other side, the studies shed light on how more complex gene-regulatory circuits evolved to increase nervous system complexity. Taken together, I aimed to:

1. identify a novel neuronal relevant lncRNA.
2. unravel the gene-regulatory mechanism of the identified lncRNA on the molecular level.
3. describe lncRNA's gene regulatory mechanism in the context of neuron formation.

5 Materials & Methods

5.1 Materials

5.1.1 Instruments

Table 4: Instruments

instrument	source	identifier
Mini Trans-Blot Module	Biorad	1703935
Mini-P TET,Comp Sys 10W 1.5 mm	Biorad	1658006
Powerpac HC Power Supply	Biorad	1645052SP
gel documentation system	Biorad	1708195
real time pcr detection system, 384 wells	Biorad	CFX384
sonifier ultrasonic cell disruptors and homogenizers	Branson	SF 4000
focused ultrasonicator	Covaris	M220
Mastercycler nexus gradient	Eppendorf	6331000017
flourescent microscope	Leica	DM 8000
STEPONE PLUS 96-well RT PCR SYS	life technologies	4376599
balance	Mettler-Toledo	30029108
SevenCompact pH-Meter	Mettler-Toledo	30130863
PerfectBlue Gelsystem Mini L	Peqlab	40-1214
PerfectBlue Breitformat Gelsystem	Peqlab	40-2314N
PerfectBlue Breitformat Gelsystem Giessschiene	Peqlab	40-1214-MC
microwave	Roth	HX94.1
benchtop uv transilluminator	UVP	M-15V
thermomixer C	VWR	460-0223

5.1.2 Reagents

Table 5: Reagents

reagent	source	identifier
5(6)-Carboxytetramethylrhodamine succinimidyl ester	Sigma-Aldrich / Merck	21955
5/(6)-carboxyfluorescin succinimidyl ester	Sigma-Aldrich / Merck	21878
Acetic Acid 100% 1l Essigsäure ROTIPURAN	Roth	3738 . 4
Acrylamid Rotiphorese Gel40(37.5:1)	Roth	T802 . 1
Agar	Becton Dickinson	214030
Agarose Peq gold	VWR (Peqlab)	35-1020
Alkaline Phosphatase	NEB	M0290
Ammoniumperoxodisulfat	Roth	9592 , 2
Ampicillin	Roth	K029 . 2
Amylose resin	NEB	E 8021S
β-Mercaptoethanol 100ml	Roth	4227 . 3
BSA, Protease-free heat shock fraction, proteasefree pH 7.0	Sigma-Aldrich / Merck	A3294-10G
C-1-strepdavidin magnetic beads	Thermo Fischer	65002
CHCl3	Sigma-Aldrich / Merck	C2432-1l
D(+)-Maltose Monohydrat	Roth	8951.1
DAPI (4',6-Diamidine-2'-phenylindoledihydrochlorid)	Roche	10 236 276 001
Deoxynucleotide (dNTP) Solution Set	NEB	N0446S
Dexamethasone powder	Sigma-Aldrich / Merck	D4902-25MG
Dextransulfat	Roth	5956.5
D-Glucose	Roth	X997.1
DNA ladder 1Kb Plus (1000μg,1μg/μl)	life-techn Invitrogen	10787 - 026
DTT 1,4 Dithiothreit >99%, p.a. 5g	Roth	6908.1
EDTA	Merck	1.08418 . 1000
Ethanol p.a.	Roth	9065 . 3

Fluorescein RNA Labeling Mix	Sigma-Aldrich / Merck	11685619910
Formaldehyd 37% p.a. 1l	Merck	104003 . 1000
GelRed Nucleic Acid Stain	Biotium /VWR	#41003 / 730-2958
Glycerin	Roth	3783.1
Glycine for Buffers 5kg	Applichem	A1377 , 5000
Goat Serum	Gibco	16210-064
Grace Bio-Labs HybriSlip^{TM} hybridization cover	Sigma-Aldrich / Merck	GBL716024-100EA
HCl saturated	Roth	4625 . 1
Heparin HiTrap-column 5x 1ml	GE healthcare	GE17-0407-01
Hepes for Buffers	Applichem	A1069 , 1000
I-Block Tropix	Applied Biosystems	T 2015
IPTG (Isopropyl-β-D-thiogalactoside)	Roth	CN08.2
Isol RNA Lysis Reagent (200ml) Trizol	5 prime / VWR	2302700 /733 - 1089 VWR
Isopropanol	Roth	9866 . 2
KCL p.a.	Roth	6781 , 1
linear acrylamide		
Methanol p.a.	Roth	CP43.3
Mowiol 4-88 (diazabicyclo-octane)	Roth	0713.2
Na_2HPO_4	Merck	106580 . 5000
NaH_2PO_4	Merck	106346 . 1000
NaCl	Roth	3957. 2
NaHCO3 p.a.	Merck	106329 . 1000
NaOH p.a.	Roth	6771 . 1
Normal Goat Serum	life.tech. Gibco	16210-064
Oligo(dT)12-18 Primer	life technologies	18418-012
Paraformaldehyde, 16% Solution, EM Grade.	VWR	ICNA0219998320
PMSF	Roth	66367 , 2
Precision Plus Protein^{TM} All Blue Prestained Protein Standards	Biorad	#1610373
Protein G agarose	Thermo Fischer	20397
PVDF Immobilon-P Transfer Membran 0,45 μ x 2.5cm x 3,75cm	Millipore	IPVH00010
random hexamer	Thermo Fischer	N8080127
Ribonucleotide set (rNTPs)	NEB	N0450S
Rubidiumchlorid p.a. 10 g (Buffer für comp. DH-5 alpha)	Roth	4471.1
Tris for Buffers	Applichem	A1379 .5000
Triton x 100	Merck	108603 . 1000
Trypton 2 kg Bacto-Trypton	Becton Dickinson	211699
TSA Blocking Reagent	Perkin Elmer	FP1020
TSA Plus Fluorescein Evaluation Kit	Perkin Elmer	NEL741E001KT
Tween 20	Merck	822184 . 0500
UltraPure DNAse/RNAse-Free Distilled Water	Thermo Fischer	10977-049
Urea >99.5 % p.a.	Roth	3941.1
Yeast Extract 2kg	Becton Dickinson	212720
yeast tRNA 25 mg	Thermo	15401011

5.1.3 Biological strains

Table 6: Biological strains

strain	source	identifier
Dh5α	NEB	C 29871
BL21	Haass Department	n.a.
Hek 293T	LGC standards	ATCC®CRL-1126^{TM}
Neuro2A	LGC standards	ATCC®CCL-131^{TM}

5.1.4 Plasmids

Table 7: Plasmids

vector	source	identifier
pcDNA.5-FRT	Thermo Fischer	V601020
pCSFLPe	addgene	31130
pLenti-GFP-Puro	addgene	73582
pLKO.1-Puro-scr	addgene	1864
pLKO.1-Puro-TRC control	addgene	10879
pLV mCherry	addgene	36084
pSico Ef1α-mCherry-2TA-Puro	addgene	31845
pSLIK-Neo	addgene	25735
pSLIK-Venus	addgene	25734

5.1.5 Enzymes

Table 8: Restriction enzymes

vector enzyme	source	identifier
AflII	NEB	R 0520L
Age I, recombinant	NEB	R 0552 S
Age I-HF	NEB	R 3552 S
ApaI	NEB	R0114S
AvrII	NEB	R0174S
BamH I-HF	NEB	R 3136 S
Bgl2	NEB	R0144L
BsiWI-Hf	NEB	R3553S
BspEI	NEB	R0540S
BsrGI-HF	NEB	R3575S
BstE II-HF	NEB	R 3162 S
ClaI	NEB	R0197S
EcoR I-HF	NEB	R 3101 S
EcoRV	NEB	R3195S
Hind III-HF	NEB	R 3101 S
KpnI-HF	NEB	R3142 L
MfeI	NEB	R0589S
MluI-Hf	NEB	R3198S
MscI	NEB	R0534S
Nco-1 HF	NEB	R 3193 S
NgoMIV	NEB	R0564S
Nhe1	NEB	R0131L
Not-I-HF	NEB	R 3189 S
PspOMI	NEB	R0653S
Pst I-HF	NEB	R 3140 S
SacI-HF	NEB	R3156M
Sal I-HF	NEB	R 3138 S
SbfI	NEB	R3642S
SspI-Hf	NEB	R3132S
Xba I, recombinant	NEB	R 0145 S
Xho I, recombinant	NEB	R 0146 S
XmaI	NEB	R0180S

Table 9: Other enzymes

enzyme	source	identifier
5x FIREPol®Master Mix	solis biodyne	04.11.0125
DNAse (RNAse free)	roche / sigma aldrich	4716728001
Fast SYBR®Green Master Mix	Thermo Fischer	4385614
M-MLV Reverse Transcriptase	Thermo Fischer	28025013
murine RNAse Inhibitor	NEB	M0314L
Phusion Hf-DNA Polymerase	NEB	M0530L
Proteinase K (RNAse free)	Thermo Fischer	AM2548
RNase A, DNase and protease-free	Thermo Fischer	EN0531
RNAse H	NEB	M0297L
rSAP shrimp Alkaline Phosphatase 1000Units	NEB	M0371S
SP6 RNA Polymerase	NEB	M0207L
T4 DNA Ligase	NEB	M0202L
T4 DNA Polymerase	NEB	M0203S
T7-RNA polymerase	NEB	M0251S

5.1.6 Molecular biology kits

Table 10: Molecular biology kits

Kit	source	identifier
BCA Protein Assay Reagent	Thermo Fischer	23225
FirstChoice®RLM-RACE Kit with Manual	Thermo Fischer	AM1700M
milliTUBE 1 ml AFA Fiber (24)	Covaris	SKU:520135
Nucleobond Xtra Midi Kit ef (Midi-Plasmid)	Machery Nagel	740420.5
NucleoSpin®Gel and PCR Clean-up	Machery Nagel	740.609.250
NucleoSpin®Plasmid (Mini-Plasmid)	Machery Nagel	740.588.250
Pierce ECL Western Blotting Analysis System (Normales ECL)	Thermo Fischer	32106
Qubit™ dsDNA HS Assay Kit	Thermo Fischer	Q32851
RNeasy Mini Kit	Quiagen	Cat No./ID: 74106
SuperSignal™ West Femto Maximum Sensitivity Substrate	Thermo Fischer	34094
TA Cloning™ Kit, Dual Promoter, with pCR™II Vector, without competent cells	ThermoScientific	K207020
truChIP Chromatin Shearing Kit with Formaldehyde	Covaris	SKU: 520154

5.1.7 Antibodies

Table 11: primary antibodies

antibody	biol. source	source	identifier
Bmi-1	mouse	Millipore Merck	05-637
Brd2	rabbit	Merck	ABE477
Brd4	rabbit	Biomol	A301-985A -M
caspase3 (Asp175)	rabbit	cell signalling	#9664
GFAP	goat	Santa Cruz	sc-6170
Glast	rabbit	Thermo Fischer	# PA5-80012
IgG	rabbit	cell signalling	#2729
LBR	rabbit	abcam	ab122919
Mapt	rabbit	Dako	A002401-2
Nestin	mouse	Millipore	MAB353
Smarca-5	rabbit	Merck	ABE1026
β-tubulin III / Tuj1	mouse	Covance	MMS-435P
BrdU Antibody — Bu20a	mouse	BioRad	MCA2483GA

Table 12: secondary antibodies

antibody	biol. source	source	identifier
DyLight550-αIgG-mouse	donkey	Thermo Fischer	SA5-10167
DyLight550-αIgG-rabbit	donkey	Thermo Fischer	SA5-10039
DyLigth488-αIgG-mouse	donkey	Thermo Fischer	SA5-10166
DyLigth488-αIgG-rabbit	donkey	Thermo Fischer	SA5-10038
IgG-mouse-HRP	donkey	Promega	W 1018
IgG-rabbit-HRP	donkey	Promega	402 B

5.1.8 Cell culture reagents

Table 13: Cell culture reagents

reagent	source	identifier
1.0M HEPES	Thermo Fischer	15630-056
10x MEM	Thermo Fischer	21430-020
B-27®Supplement (50X), serum free	Thermo Fischer	17504001
DMEM, high glucose, GlutaMAX™ Supplement	Thermo Fischer	61965-059
DMEM/F12	Thermo Fischer	11330-057
FCS	PAN Biotech	P40-37 500
Geneticin™ Selective Antibiotic (G418 Sulfate), Powder	ThermoFischer	11811023
GlutaMAX™ Supplement	Thermo Fischer	35050038
HBSS, calcium, magnesium, no phenol red	Thermo Fischer	14025100
Horse serum 500ml	Thermo Fischer	26050088
Hygromycin B	Thermo Fischer	10687010
Insulin-Transferrin-Selenium-Sodium Pyruvate (ITS-A) (100X)	Thermo Fischer	51300-044
Laminin	Roche	11243217001
L-Glutamin	Thermo Fischer	25030081
MEM Amino Acids Solution (50X)	Thermo Fischer	11130-036
MEM Non-Essential Amino Acids Solution (100X)	Thermo Fischer	11140-035
Neurobasal medium	Thermo Fischer	21103-049
Opti-MEM®I Reduced Serum Medium, GlutaMAX™ Supplement	Thermo Fischer	51985026
Penicillin/Streptavidin	Thermo Fischer	15070-063
poly-D-Lysine HBr	Sigma-Aldrich	P7280
Puromycin dihydrochloride from Streptomyces alboniger	Sigma	P8833-25MG
Recombinant Human FGF-basic (154 a.a.)	peprotech	100-18B
Trypsin-EDTA (0.05%), phenol red	Thermo Fischer	25300-062

5.1.9 Oligonucleotides

All purchased from integrated DNA technologies (IDT)

5.1.10 Software

Table 14: Software

software	source
Ape	Davis (2010)
MaxQuant	Cox and Mann (2008) [138]
R 4.0.2	Ihaka (1993) [139]
python 2.7	Rossum (1991)
bowtie 2	Langmead et al. (2012) [140]
Star 2.0	Dobin et al (2013) [141]

Rsem	Li et al (2011) [142]
samtools	Li et al (2009) [143]
MACS1.4	Feng et al (2012) [144]
Meme Suit	Mcleay and Bailey. (2010) [145]
Homer	Heinz et al (2010) [146]
Triplexator	Buske et al. (2012) [106]

Table 15: R-packages

R package	source
rmarkdown	Xie et al (2020)
dplyr	Wickham et al. (2020)
magrittr	Bach and Wickham (2014)
knitr	Xie et al (2020)
ggplot	Wickham et al. (2005)
gplots	Warnes et al. (2020)
pheatmap	Kolde et al. (2004)
biomaRt	Smedley et al (2009) [147]
DESeq2	Anders and Huber (2010) [148]
SummarizedExperiments	Morgen et al (2020)
SummaryTools	Morgen et al (2020)
GenomicRanges	Lawrence et al (2013)
Rsubread	Liao et al. (2019) [149]

5.1.11 Buffers and Media:

- 50x TAE: 2 M Tris, 1 M acetic acid, 50 mM EDTA
- 10x TBE: 0.89 M Tris, 0.89 M boric acid, 20 mM EDTA
- 1 x PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.6 mM KH2PO4
- LB (1l): 10 g Bacto Tryptone, 5 g yeast extract, 5g NaCl
- LB-Agar (1l): 15 g Bacto Agar, 10 g Bacto Tryptone, 5 g yeast extract, 5g NaCl
- 500x Ampicilin: 100 mg/ml

5.2 Methods

5.2.1 Molecular cloning

Agarose Gel

1-1.5g Agarose was dissolved in 100ml TAE buffer in a microwave oven. 5 μl of GelRed was added to melted agarose. The gel was run in an electrophoresis chamber equipped with 1x TAE at 90-135V.

Gel purification of DNA

According to the manufacturer's instruction, amplified or digested DNA was purified via a 1-1.5% agarose gel and the MN PCR column kit. The DNA was eluted in 30 μl nuclease-free H_2O and quantified using a nanodrop.

Preparative Polymerase chain reaction (PCR)

1 pg-10 ng template DNA was amplified in 50-100 μl reaction scales containing 200 μM dNTS, 500 nM forward, 500 nM reverse primer, 1x HF- fidelity buffer, and 1units Phusion polymerase / 50 μl under hot start conditions using a thermocycler. The reaction was preheated at 98°C for 3 min. DNA was denatured at 98°C for 10sec, annealed to primers for 10 sec, and polymerized at 72°C for 30sec / kbp in a total number of 40 cycles. PCR reaction was terminated after a final incubation at 72°C for 10 min. The annealing temperature was calculated using the NEB https://tmcalculator.neb.com/#!/main tool. For the first five cycles, the annealing temperature was decreased by 5°C. Amplified DNA was gel purified.

Overlapping extension PCR

Two PCRs separately synthesized the 5' and the 3' educt with primers featuring an overlap between both educts of 18-20 bp lengths. Both educts were applied in a molar ratio of 1:1. A PCR reaction of 5 cycles produced the overlap without forward and reverse primer. After the first five cycles, forward- and reverse-primer were added to amplify the overlapping product. Amplified DNA was gel purified.

analytical Polymerase chain reaction (PCR)

Analytical PCR was carried out as 20 μl reaction scales in 1x FirePol MasterMix with 10 pmol FW and 10 pmol RV primer. The reaction was preheated at 95°C for 3min. DNA was denatured at 95°C for 10 sec, annealed to primers for 10 sec, and polymerized at 72°C for 60sec / kbp in a total number of 40 cycles. PCR reaction was terminated after a final incubation at 72°C for 10 min. The annealing temperature was calculated using the NEB https://tmcalculator.neb.com/#!/main tool.

Enzymatic restriction digestion

0.3 – 10 μg DNA was digested in a 30 μl reaction scales containing 1x Digestion (typically CutSmart) buffer depending on the restriction enzymes and 10 U restriction enzyme. If required, digested DNA was treated with shrimp alkaline phosphatase *in situ*. The digestion scale was incubated for 2 h at 37°C. Digested DNA was run on a 1-1.5% agarose gel for analytical or preparative purposes.

Ligation

50 ng digested vector was ligated with a 3x fold molar excess of digested insert in a 20 μl reaction containing 1x T4-DNA Ligase buffer and 800 units T4-DNA Ligase. The ligation reaction was performed at RT for 3 h.

Transformation into $RbCl_2$ competent Dh5α or BL21 E.Coli cells

For transformation, 2-9 μl Ligation reaction or 4 μl annealing reaction or 50 ng purified vector was added to on-ice thawed $RbCl_2$ competent Dh5α or BL21 E.Coli cells. Cells were incubated for 30 min on ice and heat-shocked at 42°C for 80 sec. Afterward, cells

were chilled for 5min on ice and treated with 400 μl antibiotic-free LB medium for 60 min at 37°C. Cells were pelleted at 1000 g for 3 min, resuspended in 100 μl antibiotic-free LB, and plated on LB-agar plates containing the appropriate selection antibiotic -typically ampicillin. LB-agar plates were incubated at 37°C overnight.

Mini-Plasmid DNA-preparation

12 ml LB medium supplemented with the appropriate antibiotic was inoculated either with a single colony directly picked from the LB Amp plate or with 10 μl colony suspension used for colony PCR. Cells were grown at 37°C on a shaker. After overnight incubation, cells were harvested at 2700 g and alkaline lysed using the MN column Mini Plasmid preparation kit. DNA was isolated on cationic-exchange chromatography according to the manufacture's instructions. The DNA was eluted in 35 μl nuclease-free H_2O.

Midi-Plasmid DNA-preparation

400 ml LB medium supplemented with the appropriate antibiotic was inoculated with 5 ml of a pre-culture of cells of Plasmid expressing Dh5α cells. The main-culture was incubated at 37°C on a shaker overnight. Cells were harvested at 2700 g, and alkaline lysed. The DNA was purified using the MN Midi-Prep DNA preparation kit according to the manufacturer's instruction. The DNA pellet was resolved in 60-200 μl nuclease-free H_2O.

Cloning of shRNA in pLKO.1-Puro vectors

pLKO.1-TRC cloning vectors were digested with EcoRI and AgeI. Oligonucleotides to synthesize shRNAs were designed according to addgene instruction [127]: 5' CCGG—19-21 bp sense—ctcgag—19-21bp antisense—TTTTTG 3' for the FW oligonucleotide and 5' AATTCAAAAA— 19-21 bp sense —ctcgag—19-21 bp antisense 3' for the RV oligonucleotide. For selecting shRNA targeting regions, we used the online available:
http://sirna.wi.mit.edu/ tool. 100 pmol FW and 100 pmol RV oligo-nucleotide in 50 μl 1x NEB2.1 were heated up to 95°C for 5 min and then slowly cooled down to RT. 2 μl of the annealing scale was ligated into 20 ng digested pLKO.1 vector and transformed into Dh5α. Plasmid DNA was prepared from 1-4 clones per shRNA and screened for correct insert via analytical digestion with EcoRI and NdeI.
Following oligonucleotides were used:

Table 16: shRNA-oligonucleotides

sh-RNA	oligonucleotide	5'5'->3' sequence
Cdr1lncRNA-sh_a	Cdr1lncRNA-sh_a-fw	ccggCATACTCATTCCCTCGAATctcgag ATTCGAGGGAATGAGTATGttttttg
	Cdr1lncRNA-sh_a-rv	aattcaaaaaCATACTCATTCCCTCGAATctcgag ATTCGAGGGAATGAGTATG
Cdr1lncRNA-sh_b/ SH1	Cdr1lncRNA-sh_b-fw	ccggCAATGAATAACAGGTATGActcgag TCATACCTGTTATTCATTGttttttg
	Cdr1lncRNA-sh_b-rv	aattcaaaaaCAATGAATAACAGGTATGActcgag TCATACCTGTTATTCATTG

Cdr1lncRNA-sh_c	Cdr1lncRNA-sh_c-fw	ccggCATGTCTCTTGATAGTCCActcgag TGGACTATCAAGAGACATGttttttg
	Cdr1lncRNA-sh_c-rv	aattcaaaaaCATGTCTCTTGATAGTCCActcgag TGGACTATCAAGAGACATG
Cdr1lncRNA-sh_d / SH2	Cdr1lncRNA-sh_d-fw	ccggCTTCCTGTATGCGGAATATctcgag ATATTCCGCATACAGGAAGttttttg
	Cdr1lncRNA-sh_d-rv	aattcaaaaaCTTCCTGTATGCGGAATATctcgag ATATTCCGCATACAGGAAG
Islr2lncRNA-sh_a / SH1	Islr2lncRNA-sh_a-fw	ccggGGGTATGCATATCCGCAAActcgag TTTGCGGATATGCATACCCttttttg
	Islr2lncRNA-sh_a-rv	aattcaaaaaGGGTATGCATATCCGCAAActcgag TTTGCGGATATGCATACCC
Islr2lncRNA-sh_b	Islr2lncRNA-sh_b-fw	ccggCTCTAGTCATTGGACACATctcgag ATGTGTCCAATGACTAGAGttttttg
	Islr2lncRNA-sh_b-rv	aattcaaaaaCTCTAGTCATTGGACACATctcgag ATGTGTCCAATGACTAGAG
Islr2lncRNA-sh_c	Islr2lncRNA-sh_c-fw	ccggGATGTGGATTCTCTAGTCActcgag TGACTAGAGAATCCACATCttttttg
	Islr2lncRNA-sh_c-rv	aattcaaaaaGATGTGGATTCTCTAGTCActcgag TGACTAGAGAATCCACATC
Islr2lncRNA-sh_d/ SH2	Islr2lncRNA-sh_d-fw	ccggCACATTGAGATGATCCTCActcgag TGAGGATCATCTCAATGTGttttttg
	Islr2lncRNA-sh_d-rv	aattcaaaaaCACATTGAGATGATCCTCActcgag TGAGGATCATCTCAATGTG
mir124a-hg1-sh_a/ SH1	mir124a-hg1-sh_a-fw	ccggGTGGCTGTTATCTCATTGTctcgag ACAATGAGATAACAGCCACttttttg
	mir124a-hg1-sh_a-rv	aattcaaaaaGTGGCTGTTATCTCATTGTctcgag ACAATGAGATAACAGCCAC
mir124a-hg1-sh_b	mir124a-hg1-sh_b-fw	ccggGTTATCTCATTGTCTGTGActcgag TCACAGACAATGAGATAACttttttg
	mir124a-hg1-sh_b-rv	aattcaaaaaGTTATCTCATTGTCTGTGActcgag TCACAGACAATGAGATAAC
mir124a-hg1-sh_c	mir124a-hg1-sh_c-fw	ccggGCAACCAGGATCCTTTAAActcgag TTTAAAGGATCCTGGTTGCttttttg
	mir124a-hg1-sh_c-rv	aattcaaaaaGCAACCAGGATCCTTTAAActcgag TTTAAAGGATCCTGGTTGC
mir124a-hg1-sh_d/ SH2	mir124a-hg1-sh_d-fw	ccggGATCCTTTAAAGGAGAACActcgag TGTTCTCCTTTAAAGGATCttttttg
	mir124a-hg1-sh_d-rv	aattcaaaaaGATCCTTTAAAGGAGAACActcgag TGTTCTCCTTTAAAGGATC
mir124a-hg3-sh_a	mir124a-hg3-sh_a-fw	ccggGCTCCTTTCTCATGGAAATctcgag ATTTCCATGAGAAAGGAGCttttttg
	mir124a-hg3-sh_a-rv	aattcaaaaaGCTCCTTTCTCATGGAAATctcgag ATTTCCATGAGAAAGGAGC
mir124a-hg3-sh_b/ SH1	mir124a-hg3-sh_b-fw	ccggCTGCATCCAAGGTCTAAATctcgag ATTTAGACCTTGGATGCAGttttttg
	mir124a-hg3-sh_b-rv	aattcaaaaaCTGCATCCAAGGTCTAAATctcgag ATTTAGACCTTGGATGCAG
mir124a-hg3-sh_c	mir124a-hg3-sh_c-fw	ccggGTCTAAATCTGGGTGGATActcgag TATCCACCCAGATTTAGACttttttg
	mir124a-hg3-sh_c-rv	aattcaaaaaGTCTAAATCTGGGTGGATActcgag TATCCACCCAGATTTAGAC
mir124a-hg3-sh_d/ SH2	mir124a-hg3-sh_d-fw	ccggGTGATCACTTTGATTCTGActcgag TCAGAATCAAAGTGATCACttttttg
	mir124a-hg3-sh_d-rv	aattcaaaaaGTGATCACTTTGATTCTGActcgag TCAGAATCAAAGTGATCAC
Slitrk3lncRNA (RUS)-sh_a/ SH1	Slitrk3lncRNA-sh_a-fw	ccggGCCAGAAAGACCTGGATATctcgag ATATCCAGGTCTTTCTGGCttttttg

	Slitrk3lncRNA-sh_a-rv	aattcaaaaaGCCAGAAAGACCTGGATATctcgag ATATCCAGGTCTTTCTGGC
Slitrk3lncRNA (RUS)-sh_b	Slitrk3lncRNA-sh_b-fw	ccggGGAGTTTGTAGGCCCATAActcgag TTATGGGCCTACAAACTCCttttttg
	Slitrk3lncRNA-sh_b-rv	aattcaaaaaGGAGTTTGTAGGCCCATAActcgag TTATGGGCCTACAAACTCC
Slitrk3lncRNA (RUS)-sh_c	Slitrk3lncRNA-sh_c-fw	ccggGAGTGATATAGCTCTATCTctcgag AGATAGAGCTATATCACTCttttttg
	Slitrk3lncRNA-sh_c-rv	aattcaaaaaGAGTGATATAGCTCTATCTctcgag AGATAGAGCTATATCACTC
Slitrk3lncRNA (RUS)-sh_d/ SH2	Slitrk3lncRNA-sh_d-fw	ccggGTAATGCCCAAGAGTGATActcgag TATCACTCTTGGGCATTACttttttg
	Slitrk3lncRNA-sh_d-rv	aattcaaaaaGTAATGCCCAAGAGTGATActcgag TATCACTCTTGGGCATTAC
Slitrk3lncRNA (RUS)-sh_e	Slitrk3lncRNA-sh_e-fw	ccggGGTATTTCAAAGCAGCGCTctcgag AGCGCTGCTTTGAAATACCttttttg
	Slitrk3lncRNA-sh_e-rv	aattcaaaaaGGTATTTCAAAGCAGCGCTctcgag AGCGCTGCTTTGAAATACC
Slitrk3-sh_a/ SH1	Slitrk3-sh_a-fw	ccggCCATGCTACTGCGAAGTTActcgag TAACTTCGCAGTAGCATGGttttttg
	Slitrk3-sh_a-rv	aattcaaaaaCCATGCTACTGCGAAGTTActcgag TAACTTCGCAGTAGCATGG
Slitrk3-sh_b	Slitrk3-sh_b-fw	ccggCCAGGCTTTATATCCTGGTctcgag ACCAGGATATAAAGCCTGGttttttg
	Slitrk3-sh_b-rv	aattcaaaaaCCAGGCTTTATATCCTGGTctcgag ACCAGGATATAAAGCCTGG
Slitrk3-sh_c	Slitrk3-sh_c-fw	ccggCTTGGCTTGACTGTCAACTctcgag AGTTGACAGTCAAGCCAAGttttttg
	Slitrk3-sh_c-rv	aattcaaaaaCTTGGCTTGACTGTCAACTctcgag AGTTGACAGTCAAGCCAAG
Slitrk3-sh_d	Slitrk3-sh_d-fw	ccggCCATTGTGCAGATAGATCTctcgag AGATCTATCTGCACAATGGttttttg
	Slitrk3-sh_d-rv	aattcaaaaaCCATTGTGCAGATAGATCTctcgag AGATCTATCTGCACAATGG
Slitrk3-sh_e/SH2	Slitrk3-sh_e-fw	ccggCTTCAGAGGAATTCGATGActcgag TCATCGAATTCCTCTGAAGttttttg
	Slitrk3-sh_e-rv	aattcaaaaaCTTCAGAGGAATTCGATGActcgag TCATCGAATTCCTCTGAAG
Brd2-SH used	Brd2-sh_a-fw	ccggCCACCTGAAATACCTACCActcgag TGGTAGGTATTTCAGGTGGttttttg
	Brd2-sh_a-rv	aattcaaaaaCCACCTGAAATACCTACCActcgag TGGTAGGTATTTCAGGTGG
	Brd2-sh_b-fw	ccggGTTAAACGGAAAGCGGATActcgag TATCCGCTTTCCGTTTAACttttttg
	Brd2-sh_b-rv	aattcaaaaaGTTAAACGGAAAGCGGATActcgag TATCCGCTTTCCGTTTAAC
	Brd2-sh_c-fw	ccggCTGCTGATGTACGGCTTATctcgag ATAAGCCGTACATCAGCAGttttttg
	Brd2-sh_c-rv	aattcaaaaaCTGCTGATGTACGGCTTATctcgag ATAAGCCGTACATCAGCAG
	Brd2-sh_d-fw	ccggGAGTTTCGCTATGCCAAGActcgag TCTTGGCATAGCGAAACTCttttttg
	Brd2-sh_d-rv	aattcaaaaaGAGTTTCGCTATGCCAAGActcgag TCTTGGCATAGCGAAACTC
	Brd2-sh_e-fw	ccggGTTTGCCGGATTATCACAActcgag TTGTGATAATCCGGCAAACttttttg
	Brd2-sh_e-rv	aattcaaaaaGTTTGCCGGATTATCACAActcgag TTGTGATAATCCGGCAAAC

Smarca5-SH used	Smarca5-shfw [150]	ccggGAGGAGGATGAAGAGCTATctcgag ATAGCTCTTCATCCTCCTCtttttg
	Smarca5-shrv	aattcaaaaaGAGGAGGATGAAGAGCTATctcgag ATAGCTCTTCATCCTCCTC
	Lbr-sh_a-fw	ccgggggTggTTaTTaacTTagTctcgag ACTAAGTTAATAACCACCCttttttg
	Lbr-sh_a-rv	aattcaaaaagggTggTTaTTaacTTagTctcgag ACTAAGTTAATAACCACCC
	Lbr-sh_b-fw	ccggggagcTcccTTTaTTaTgactcgag TCATAATAAAGGGAGCTCCttttttg
	Lbr-sh_b-rv	aattcaaaaaggagcTcccTTTaTTaTgactcgag TCATAATAAAGGGAGCTCC
	Lbr-sh_c-fw	ccggcaggagagaagaggTcaaactcgag TTTGACCTCTTCTCTCCTGttttttg
Lbr-SH used	Lbr-sh_c-rv	aattcaaaaacaggagagaagaggTcaaactcgag TTTGACCTCTTCTCTCCTG
	Lbr-sh_d-fw	ccggcTgTcTacgacTTcTTcaTctcgag ATGAAGAAGTCGTAGACAGttttttg
	Lbr-sh_d-rv	aattcaaaaacTgTcTacgacTTcTTcaTctcgag ATGAAGAAGTCGTAGACAG
	Lbr-sh_e-fw	ccggcacTgaTccaaagcTTgcactcgag TGCAAGCTTTGGATCAGTGttttttg
	Lbr-sh_e-rv	aattcaaaaacacTgaTccaaagcTTgca TGCAAGCTTTGGATCAGTG

Construction of pLKO.1-GFP vectors:
GFP was amplified from pLenti-GFP-Puro with the oligo-nucleotides: ttctgg atcccaccatggtgagcaagggcga & cttggtaccctacttgtacagctcgtccatg by PCR to replace puromycin resistance gene via BamHI and KpnI restriction and ligation.

Construction of pLKO.1-GFP-IRES-Neo vectors:
Neomycin/G418 resistance gene was amplified from pSLIK-Neo with the oligo-nucleotides: ttctggatccccggtccaccatgattgaacaa & cttggtacctcagaagaactcgtcaagaaggc by PCR to replace puromycin resistance gene via BamHI and KpnI restriction and ligation.

3' and 5' Rapid Amplification of cDNA Ends (RACE):
For 3'- and 5'-RACE the FirstChoice RLM-RACE kit was used and the experiment was performed according to the manufacturer's instructions.
The following primer sequences were used:

Table 17: RACE-oligonucleotides:

oligonucleotide	5'->3' sequence	position	orientation
5'outer_RUS	CTGAGGCATGCTTTGTGGAGGACA	248..271	RV
5'inner_RUS	CAATCTACACTCCGAACGCCGCA	127..149	RV
3'outer_RUS	GTCTGTCCTCTGTAATGCAAG	235..255	FW

C-terminal His6-tagging of MBP-MS2BP
pMal-MS2BP Plasmid encoding for MS2-binding protein (MS2BP) fused to Maltose-Bining-Protein (MBP) was gifted from the Becker Lab. Additional His_6-tagging at the C-terminus was done in 2 subsequent PCRs. In PCR, MBP-MS2BP was amplied

from pMal vector with the oligonucleotides: CTTGGATCCATCGAGGGTAGGgcttctaa and TTCGTGGTGATGGTGATG ATGgtagatgccggagtttgc. The obtained PCR product was remplified with the oligonucleotides: CTTGGATCCATCGAGGGTAGGgcttctaa and TACGACTCATCTAGAGGCGTacAC. PCR product from 2nd run and pMal-MS2BP were digested with HindIII and BaMHI and ligated.

Construction of pcDNA-5FRT-5xMS2

pcDNA.5-FRT vectors used to generate stable FlpIN Neuro2 A cells were equipped with 5xMS2 stem-loops as followed. 3x MS2 stem-loops were amplified with TACGACTCATCTAGAGGCGTACAC & CATGATTACGGATCCCGTACCCTG from pAdMl3-(MS2)3 gifted from Peter Becker's lab by PCR, digested with BamHI & XbaI, and ligated to BamHI & XbaI linearized pcDNA5-FRT. After transformation into Dh5α, one clone randomly expanded 3xMS2 stem-loops to 5xMS2 stem-loops was selected for Plasmid DNA preparation.

Subcloning lncRNAs into pcDNA-5FRT-5xMS2

Full-length lncRNAs and mutants were cloned into pcDNA.5-FRT-5xMS2 ligation independently [151]. pcDNA.5-FRT-5xMS2 was linearized with AflII and treated with T4-DNA polymerase in the presence of dGTP. lncRNA constructs were amplified by PCR using oligo-nucleotides bearing TTTAAACTTAAG as FW- and AAGCTTAAG as RV-overhang. PCR products were treated with T4-DNA Polymerase in the presence of dCTP. Processed PCR products and vectors were annealed in 2:1 molarity for 5 min and treated with 5 mM EDTA for further 5min. The entire annealing reaction was transformed into Dh5α.
The following primer sequences were used:

Table 18: pcDNA5-FRT-5xMS2-LIC-oligonucleotides

construct	oligonucleotide	5'->3' sequence
full length RUS	LIC-pcDNA-RUS_FW	TTTAAACTTAAG AGAGCATTTGGGCTTAAGCC
	LIC-pcDNA-RUS_RV	AAGCTTAAG GGGAGCTATAGAGCATATGT
5' dom RUS	*LIC-pcDNA-RUS_FW*	*TTTAAACTTAAG AGAGCATTTGGGCTTAAGCC*
	LIC-pcDNARUS-Ex2-rv	AAGCTTAAG ACTTGCATTACAGAGGACAGACGT
$\triangle$5' RUS	LIC-pcDNA-RUS-Ex2-fw	TTTAAACTTAAG ACGTCTGTCCTCTGTAATGCAAGT
	LIC-pcDNA-RUS_RV	*AAGCTTAAG GGGAGCTATAGAGCATATGT*
$\triangle$3' RUS	*LIC-pcDNA-RUS_FW*	*TTTAAACTTAAG AGAGCATTTGGGCTTAAGCC*
	LIC-pcDNA-RUS-Ex5-rv	AAGCTTAAG GTTAGGAGACAGGTCAGTTTAGTC

Construction of bidirectional lentiviral overexpression vectors

1. CMV-mCherry bidirectional vector:
   GFP or the lncRNA of interest were subcloned from pcDNA5.1 vectors with a PCR using the oligo-nucleotides TTTTGCGCTGCTATCGATTGTACGGGCCAG and CCGATCGATCCATAGAGCCCACCGCATCCCCAG. pLV-mCherry and the PCR

product were digested with ClaI and ligated.

2. Ubc-mCherry bidirectional vector:
   mCherry was cloned from pLV mCherry into pLenti-GFP-Puro via BamHI and SalI restriction and ligation. Ubc Promoter was isolated with the oligo nucleotides attatcgataagatctggcctccgcgccgggtttt and cttggatccagctcggtaccaagctt cgtctaaca from pSLIK-Venus via PCR. The PCR product and the overexpression vector were digested with BamHI and ClaI and ligated. The CMV promoter, including and BGH-polyA-signal, was amplified from pcDNA-FRT lacking the EcoRI site with TTTTGCGCTGCTATCGATTGTACGGGCCAG and CCGATCGATCCATAGAGCCCACCGCATCCCCAG and cloned via ClaI restriction into overexpression vector.

3. UbcP-mCherry-2TA-Puro-WPRE bidirectional vector:
   WPRE element was isolated from pSLIK-Venus via PCR with TCAACGCGTtcgccagtcgacaatcaacctctg, and TCAACGCGTaaaggtaccgagctc gaattccag and cloned via MluI restriction site into the Ubc-mCherry-CMV bidirectional vector. 2TA signal and puromycin resistance gene were cloned from pSico Ef1α-mCherry-2TA-Puro into the over-expression vector via PstI and EcoRI sites.

Subcloning lncRNAs into UbcP-mCherry or UbcP-mCherry-2TA-Puro-WPRE bidirectional vector

lncRNA constructs or GFP were either subcloned via NdeI and ApaI restriction sites from pcDNA-5-FRT vectors or via ligation independent cloning methods [151]. For this, sticky ends were generated by linearizing the vector with Bsp120l. Linearized vector was additionally treated with dTTP and T4 DNA polymerase. Inserts were amplified with oligonucleotides bearing AGAGGGCCT as FW - and TTTGCCCGGA as RV-overhang via PCR. Amplification products were treated with T4-DNA polymerase and dATP. Processed Vector and PCR products were annealed in a molecular ratio of 1:2 in the presence of 5 mM EDTA. Annealing products were transformed in $RbCl_2$ competent Dh5α.
Following oligo-nucleotides were used:

Table 19: pLenti-LIC-oligonucleotides

construct	oligonucleotide	5'->3' sequence
full length RUS	LIC_lOE-RUS_FW	AGAGGGCCT AGAGCATTTGGGCTTAAGCC
	LIC_lOE-RUS_RV	*TTTGCCCGGA GGGAGCTATAGAGCATATGT*
△5' RUS	LIC_lOE-RUS-Ex2-fw	AGAGGGCCT ACGTCTGTCCTCTGTAATGCAAGT
	LIC_lOE-RUS_RV	*TTTGCCCGGA GGGAGCTATAGAGCATATGT*
△3' RUS	*LIC_lOE-RUS_FW*	*AGAGGGCCT AGAGCATTTGGGCTTAAGCC*
	LIC_lOE-RUS-Ex5-rv	*TTTGCCCGGA GTTAGGAGACAGGTCAGTTTAGTC*

Construction of pLenti FRT:
pLenti-GFP-Puro was digested with XbaI and BamHI to remove GFP downstream to CMV promoter. FRT site was generated by annealing the oligonucleotides: gatcgaagttcctattccgaagttcctattctctagaaagtataggaacttc & ctaggaagttcctatactttctagagaatag gaacttcggaataggaacttc. For annealing, 100 pmol of each oligonucleotide was heated in 50 μl 1x NEB 2.1 to 95°C for 5 min and slowly cooled down. 2 μl annealing scale was ligated into 20 ng digested vector and transformed in $RbCl_2$ competent Dh5α.

Colony PCR
4-8 colonies per construct were inoculated in 50 μl LB medium containing the appropriate antibiotic. 2 μl of the cell suspension was amplified in a 20 μl reaction scale containing 1x FirePol Mastermix containing 10 pmol forward and reverse primer. The reaction was preheated at 95°C for 3 min. DNA was denatured at 95°C for 10 sec, annealed to primers for 45 sec, and polymerized at 72°C for 60 sec/kbp in a total number of 40 cycles. PCR reaction was terminated after a final incubation at 72°C for 10 min. The annealing temperature was calculated using the NEB Tm calculator https://tmcalculator.neb.com/#!/main tool. For cloning approaches into pcDNA.5-FRT–5xMS2 we used: CGCAAATGGGCGGTAGGCGTG as a forward primer and an appropriate quantitative RT-PCR RV primer. For cloning approaches into lentiviral Ubc-mCherry-2TA-Puro-WPRE-CMV, we used: TAGAAGGCACAGTCGAGG as a reverse primer, and a suitable quantitative RT-PCR FW primer. The total PCR reaction was separated on a 1.5% agarose gel. Colonies showing a PCR product were used to inoculate 12 ml LB medium supplemented with the appropriate selection antibiotic for Mini-Plasmid DNA preparation.

5.2.2 In vitro transcription using Sp6 or T7 RNA-polymerase

1000 ng DNA template encoding for Sp6 or T7 promoter was *in vitro* transcribed with 400 U RNA polymerase in 80 μl 1x RNA reaction Pol buffer, supplemented with 18 mM $MgCl_2$, 3 mM of each rNTP and 80U murine RNAse Inhibitor for 1h at 37°C. DNA template was removed by adding 9 μl 10x DNAse I buffer, and 4 U DNAse I. DNAse Digestion was performed for 30 min at 37°C. *In vitro* transcribed RNA was purified with the Qiagen RNAeasy column kit and eluted in 50 μl nuclease-free H_2O.

5.2.3 Protein purification of MBP-MS2BP-His_6

MBP-MS2BP-His_6 Protein was recombinant purified as described for MBP-MS2BP by LeCuyer et al., 1995 [152]. 10 ng pMal-MS2BP-His_6 Plasmid DNA encoding for MBP-MS2BP-His6 was transformed into Bl21 E.Coli cells. One single colony was picked from the plate to inoculate 5ml LB-Amp as pre-culture. Pre-culture was incubated at 37°C on a shaker overnight. The next day, the entire pre-culture was used to inoculate 800 ml

LB-medium supplemented with ampicillin and 2% Glucose. Cells were grown at 37°C on a shaker to an OD of 0.4. Protein expression was induced with 1 mM IPTG for 3 h at 37°C. Cells were harvested at 28000 g 4°C for 10 min and resuspended in 10 ml lysis buffer per 1 g cell pellet (20 mM Hepes pH 7.9, 200 mM KCl, 1 mM EDTA, 20 μM PMSF). 800ml culture yielded 3 g cell pellet typically. Cells were lysed on a Branson tip-sonicator for 10 min (30 sec pulse, 30 sec break, 70% duty factor). The cell lysate was cleared at 17.000 g, 4°C for 30 min and added on 5 ml preequilibrated Amylose-Agarose. Protein was loaded onto beads for 1 h at 4°C under continuous rotation. Beads were separated on a gravity flow column and resuspended in 20 ml lysis buffer supplemented with 2 M urea. Beads were again incubated for 10 min at 4°C under continuous rotation and separated on a gravity-flow column. Beads were washed at the gravity flow column with a gradient of 8 beads volume lysis buffer and two-volume wash buffer (20 mM Hepes pH 7.9, 20 mM KCl, 1 mM EDTA). Protein was eluted with 20 ml wash buffer supplemented with 10 mM maltose in 1 ml fractions. Fractions were analyzed by SDS-polyacrylamide gel-electrophoresis and Coomasie staining. Fractions containing MBP-MS2BP-His6 protein were pooled and concentrated on an Amicon 3K filter to a volume of approximately 1 ml. 300 μl of concentrated protein solution was subjected to a heparin column pre-equilibrated with HB wash buffer (Hepes pH 7.9, 20 mM KCl). After loading, the column was washed with 5 volumes of HB wash buffer (Hepes pH 7.9, 20 mM KCl). Protein was eluted manually from the Heparin column by a gradient of 20 mM - 200 mM KCl in 6 fractions. Protein typically elutes from 40 mM to 80 mM KCl. Fractions containing MBP-MS2BP-His6 Protein were pooled and concentrated on an Amicon 3 K filter to a volume of approximately 500 μl. Concentrated protein was frozen at -20°C with 10% glycerol. 800 ml culture typically yielded 30-45 mg pure protein.

5.2.4 Cell culture

All mammalian primary cells and stable cell lines were kept in a humidified 5% CO2 incubator at 37°C.

Cultivation of cell lines

HEK293 and Neuro2A cells were cultured in DMEM with Glutamax I supplemented with 8% FCS and 1x Penicillin/Streptomycin (P/S).

Trypzinisation

Cells were washed once in PBS and treated with trypsin till cells dissociate from the culture dish. Trypsin digestion was stopped with the two volumes of 8% FCS containing medium. Cells were pipetted from the plates and pelleted at 200 g for 5 min.

Poly-D-Lysine coating

Steril filtered 0.1 mg/ml poly-D-lysine solution was added to culture dishes. For a six-well plate, we typically used 350 μl, and for a 24-well plate pre-equipped with coverslips, we used 200 μl. After 10 min incubation at room temperature, we removed the poly-D-lysine solution and washed the culture dishes two times with PBS.

Poly-D-Lysine / Laminin coating

Steril filtered 0.1 mg/ml poly-D-lysine containing 10 μg/ml Laminin solution was added to culture dishes. For a six-well plate, we typically used 350 μl, and for a 24-well plate pre-equipped with coverslips, we used 200 μl. After two days of incubation at room temperature in the darkness, we removed the poly-D-lysine / Laminin solution and washed the culture dishes three times with H_2O.

Antibiotic selection of cells:

Following antibiotic concentrations used to select cells:

Table 20: Antibiotic concentrations

antibiotic	concentration
Puromycin	1-2 μg/ml
Hygromycin B	600 μg/ml
G418 / Geniticin	100 μg/ml

Cultivation & differentiation of primary neural stem cells

Primary neural stem cells were isolated from E15-E16 murine cortices. Cortices were dissected from embryonic brains, 5x washed with Hanks Balanced Salt Solution (HBBS), and incubated in Trypsin-EDTA (0.5%) for 15 min. Afterward, cortices were washed 5x with MEM-HS supplemented with 1x L-glutamine, 1x essential amino acids, 1x non-essential amino acids, and 10% Horse serum. Cells were singularized, pelleted at 200 g for 5 min, and seeded in a density of 500.000 cells/ml. Neural stem cells were cultured in DMEM-F12 with 5% FCS, 1x B27 supplement & 20 ng/ml basic-fibroblast growth-factor (bFGF) on poly-D-lysine coated culture dishes. Every second day, the culture medium was supplemented again with 20 ng/ml FGF2. Induction of Differentiation was carried out in Neurobasal medium supplemented with 1x B27 supplement and 0.25x Glutamax.

Isolation and Cultivation of primary cortical neurons

Primary neural stem cells were isolated from E15-E16 murine cortices. Cortices were dissected from embryonic brains, 5x washed with Hanks Balanced Salt Solution (HBBS), and incubated in Trypsin-EDTA (0.5%) for 15 min. Afterward, cortices were washed 5x with MEM-HS supplemented with 1x L-glutamine, 1x essential amino acids, 1x non-essential amino acids & 10% Horse serum. Cells were singularized, pelleted at 200 g

for 5min, and seeded in a density of 50.000 cells/cm^2 in MEM-HS supplemented with 1x L-glutamine, 1x essential amino acids, 1x non-essential amino acids, and10% Horse serum. Cells were plated on poly D-lysine, and laminin-coated culture dishes. 3 h post-plating, the medium was changed to Neurobasal medium supplemented with 1x B27 supplement, and 0.25x Glutamax.

Lentiviral virus production

3 Mio Hek 293T cells were seeded in 8 ml DMEM-GlutMax supplemented with 8% FCS on a 10 cm culture dish. Per virus production, 4 x 10 cm dishes were seeded. The day after, 53 μg Plasmids in a molar ratio of 2:1:1 of lentiviral-vector: psPAX2 (reverse transcriptase & integrase, Integration): pMD2.G (VSV-G, envelope) transfected into 50-70% confluent Hek 293T cells using. Therefore, Plasmids were diluted in 1.2 ml 150 mM NaCl, and 60 μl JetPEI (2.6 μg/ μl) in 1.2 ml 150 mM NaCl diluted. After 5min incubation at room temperature, both solutions were mixed and for further 30 min incubated at room temperature. Per 10 cm dish, 635 μl of the transfection scale was added dropwise. The next day, the medium was replaced with 8.5ml DMEM-GlutMax containing 8% FCS. Two days post-transfection, the medium was harvested, centrifuged at 3000 g for 15 min, and filtered through a 45 μm nylon filter to remove residual Hek 293T cells and cell debris. Lentiviral particles were pelleted at 87.000g for 2 h at 4°C, and the supernatant was carefully decanted. 200 μl TBS5 (50 mM Tris-HCl pH 7.8, 130 mM NaCl, 10 mM KCl, 5 mM $MgCl_2$, 10% BSA) was added to the virus pellet. After 4h incubation at 4°C, the pellet was carefully dissolved and frozen at -20°C. 5 μl or 3 μl of the KD virus were administered to a well of a six-well or 24 well plate, respectively. The double amount of Ubc-mCherry-2TA-Puro-CMV-virus was applied for over-expression experiments.

Generation of 5xMS2 tagged RNA stably overexpressing FlpIn Neuro2A cells

50.000 Wt-Neuro2A cells were seeded and treated with 5 μl pLenti-FRT virus on the next day. 2 days after transduction, cells were selected with Puromycin supplemented DMEM with GlutMax, 8% FCS, and 1x P/S. One Mio Neuro2A-FRT cells were seeded on a 10 cm culture dish. On the next day, cells were transfected with 15 μg Plasmids. We typically used a molar ratio of 1:6 till 1:9 of pcDNA5-lncRNA-5xMS2: pCSFLPe encoding for the Flipase. Plasmids were diluted in 300 μl 150 mM NaCl, and 15 μl JetPEI (2.6 μg/ μl) in 300 μl 150 mM NaCl diluted. After 5 min incubation at room temperature, both solutions were mixed. After further 30 min incubation at room temperature, the solution was added dropwise on Neuro2A-FRT cells. 2 days after transfection, cells were harvested by trypsinization and seeded again on a 10 cm dish in DMEM with GlutMax, 8% FCS, and 1x P/S supplemented with Puromycin and Hygromycin B. The medium was replaced every second day to remove cell debris. 7-10 days post-transfection, the first colonies formed. Colonies were collected and further expanded.

5.2.5 RNA-Isolation

900 μl of Isol-RNA Lysis Reagent was added either to cells from 35mm dishes (6 well plates), chromatin pellets, bead suspensions, eluates, or other RNA mixtures and incubated at room temperature for 10 min. The suspension was transferred to a new 1.5 ml tube and treated with 200 μl $CHCl_3$ for 5min under continuous shaking. The aqueous phase was separated from the organic phase by spinning at 17.000 g and 4°C for 20 min. RNA was precipitated from the aqueous phase with 650 μl isopropanol and 3 μl linear acrylamide at -20°C at least for 1 h. RNA was pelleted at 17.000 g and 4°C for 35 min and washed 2x with 75% EtOH. The RNA pellet was air-dried and solved in 15-35 μl nuclease-free H_2O.

5.2.6 First-strand cDNA synthesis (reverse transcription)

37.5 pmol random hexamers or 0.375 μg Oligo-dT12-18¬ were annealed to 500-1000 ng bulk RNA, or an undefined affinity-purified RNA (7.5 μl) together with 7.5 nmol of each dNTPs in a 9 μl reaction scale. Therefore, RNA was first denatured at 65°C for 5 min and subsequently cooled down to 37°C enabling the hybridization of the Oligo-dT or random-hexamer primer to the RNA. Afterward, 3 μl 5x First-strand buffer, 1.5 μl DTT (100mM), 30 U murine RNAse Inhibitor, and 150 U M-MLV reverse transcriptase were added to the reaction scale. After 45 min incubation at 37°C, the reaction was stopped at 70°C for 10 min. For expression analysis by qPCR, the cDNA was diluted with 185 μl nuclease-free H_2O.

5.2.7 Quantitative RT-PCR analysis

2 μl of cDNA was amplified in a 10 μl 1x Fast SYBR® Green Master Mix in the presence of 2.5 pmol of each primer. Three technical replicates were run per primer-pair. The reaction was run in an AppliedBiosystems StepOneTM or BioRad CFX384 real-time PCR system with the settings: 95°C 3 min for initial denaturing and 40 cycles: 95°C 3 sec, 60°C 30 sec. To measure transcript levels, we designed primers annealing either to two subsequent exons or at least to the border between both exons to avoid amplifying genomic DNA. We used the online tool for primer design: http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi to select for potential primer pairs with a melt temperature of 60°C, GC content between 45% and 60%, and a length between 18-22 nt. We used the online to http://biotools.nubic. northwestern.edu/OligoCalc.html to select for primers forming no hairpins structure and exhibiting no self-annealing properties. Finally, we used the https://www.ncbi.nlm.nih.gov /tools/primer-blast/ tool to predict all potential amplification products. TBP mRNA levels were measured as endogenous controls and for normalization. Quantification was

done using the △△ Ct method:

$$\triangle\triangle\ Ct_{target,sample} = \triangle\ Ct_{target,sample} - \frac{1}{n_{TBP,sample}} \sum_{i=1}^{n_{TBP,sample}} \triangle\ Ct_{TBP,sample}$$

$$\triangle\triangle\triangle\ Ct_{target,sample} = \frac{1}{n_{target,sample}} \sum_{i=1}^{n_{target,sample}} \triangle\triangle\ Ct_{target,sample} - \triangle_{target,control-sample}$$

$$normalizedExpression_{target,sample} = 2^{\triangle\triangle\triangle Ct_{target,sample}}$$

Following oligo-nucleotides were used:

Table 21: quantitative RT-PCR oligonucleotides

Cdr1-fw	CATGGATCCCTTGGAAGACAAA
Cdr1-rv	CAGCAACTGCAAGTCTTCCA
Cdr1lncRNA-fw	ATACAGGAAGCTCTGGCCAG
Cdr1lncRNA-rv	CTGCCTTTCACCACACCATC
Gjd2-fw	CAGTCTTTGTCTGCTGCCTC
Gjd2-rv	ATCATCGTACACCGTCTCCC
Gjd2lncRNA-fw	GAGGAGACGGTCTGGAGTC
Gjd2lncRNA-rv	TGAGCTGTTAATAGGTTCGGTG
Islr2-fw	GCCAGTCGTTGACCACATTT
Islr2-rv	CGCGCGGATTTGGATCATTA
Islr2lncRNA-fw	TGGTGGCCAAGAAAGAATGC
Islr2lncRNA-rv	GGGAGCTCACGTATGGGTTA
mir124a-hg1-fw	ACCCAACGAGGACAGTGAA
mir124a-hg1-rv	CAGCGTTGATCTCTCCAATC
mir124a-hg3-fw	AGCTCTGAGTGCTCGAGATC
mir124a-hg3-rv	GTCTCTCTGGGTGACTCCAC
Slitrk3-fw	TCCTCTGTGAAGCATTTCAGC
Slitrk3-rv	TGCTTGGCTTCATCATTCCTG
Slitr3k3lncRNA (RUS)-fw	GGCGTTCGGAGTGTAGATTG
Slitr3k3lncRNA (RUS)-rv	CTGAGGCATGCTTTGTGGAG
TBP-fw	TCTACCGTGAATCTTGGCTGT
TBP_rv	CAAATCgcTCTTGGCTCCTG
Slitr3k3lncRNA (RUS)-ex4-rv	GTGGCATCCTGAAAGTTGCAAGC
Slitr3k3lncRNA (RUS)-ex5-rv	CTGGCTGCCAATCTTCTCTATG
RUS-TSS-fw	AGAGCATTTGGGCTTAAGCC
RUS-TTS-rv	GGGAGCTATAGAGCATATGT
Nestin-fw	AGGCTGAGAACTCTCGCTTGC
Nestin-rv	GGTGCTGGTCCTCTGGTATCC
-tubulin III-fw	TAGACCCCAGCGGCAACTAT
-tubulin III-rv	GTTCCAGGTTCCAAGTCCACC
Map2-fw	AGTTCAAGTAGTCACAGCTG
Map2-rv	TGTGGCTGTTTGTTCTGAG
GFAP-fw	GCACTCAATACGAGGCAGTG
GFAP-rv	GGCGATAGTCGTTAGCTTCG
Glast-fw	AAGCAACGGAGAAGAGCCTA
Glast-rv	CCTCCCGGTAGCTCATTTTA
GAPDH-fw	CTCCCACTCTTCCACCTTCG
GAPDH-rv	CCACCACCCTGTTGCTGTAG
Malat-fw	GACTAGCATTGGCAGCTGAC

Malat-rv	TCCGTATGGCTCCTTCTTCC
Grin1-fw	CACGGCTCTTGGAAGATACA
Grin1-rv	AAGTGGTCGTTGGGAGTAGG
Pax6-fw	TACCAGTGTCTACCAGCCAAT
Pax6-rv	TGCACGAGTATGAGGAGGTCT
Sox2-fw	GCGGAGTGGAAACTTTTGTCC
Sox2-rv	CGGGAAGCGTGTACTTATCCTT
Mapt-fw	GCCAGGAGTTTGACACAATG
Mapt-rv	TTGGAGTGCTCTTAGCATCG
Brd2-fw	TAGTGATGAAGGCTCTGTGGAA
Brd2-rv	GTACCCATGTCCATAGGCTGTT
Brd4-fw	AGAGACCTCCAACCCTAACAAG
Brd4-rv	TGAAGTCCTGGATACATTCCTG
Smarca5-fw	ACACCGAGATGGAGGAAGTATT
Smarca5-rv	AGGGTGAGGTTGGAGTCTTCT
Arid1b-peak-fw	CCCAGAGTTGGCAGAGCTAC
Arid1b-peak-rv	GGAGACATCGCTTCCACAGT
Dpp9-peak-fw	CTTCCTTGCCTCCACTTCAG
Dpp9-peak-rv	GAGGCAGGCAGATTTCTGAG
pou4f1-peak-fw	TTGCCGAATGGGTCTAGATT
pou4f1-peak-rv	TGCCCATCAATGAATACAGG
rs14-fw	GGCTACACACAAGATGGCGT
rs14-rv	TAGGTTTGGGTGTTATACCCGT

5.2.8 Immunohistochemistry and Immunostaining

Cells were plated on Poly-L-Lysine coated glass plates in a 12-well plate. Cells were fixed in 4% PFA for 20 min at RT, washed once in PBS for 10 min, and blocked with blocking solution (0.3% Triton X-100, 2% donkey serum in PBS) for 30 min. The primary antibody (1:1000) was diluted in 200 μl blocking solution and added for 1.5 h at RT on a shaker. The antibody solution was removed, and the cells washed in PBS three times for 10 min. The cells were incubated with the secondary antibody (1:2000) in 200 μl blocking solution for 1.5 h at RT on a shaker. Afterward, they were washed three times with PBS for 10 min, and nuclei were stained for 15 min using DAPI (2-(4-Amidinophenyl)-6-indolecarbamidine dihydrochloride, 4′,6-Diamidino-2-phenylindole dihydrochloride) (1:1000 in PBS). The cells were washed once more with PBS and mounted with mounting medium containing diazabicyclo-octane (DABCO) as an anti-fading agent. Stained cells were analyzed with a Leica DM8000 fluorescent microscope. 3-5 images were taken per coverslip and quantitatively processed with ImageJ.

5.2.9 BrdU-labeling

Cell culture medium was supplemented with 1 μg/ml BrdU. 24 h after administration, cells were immune-stained with anti-BrdU antibody.

5.2.10 Fluorescent in situ hybridization (FISH)

For Fluorescent *in situ* hybridization, fluorescein-(Fluo)-labeled RNA sense and anti-sense probes were generated for mouse RUS (GM20754). RUS isoform 1 cDNA was amplified from a mouse cDNA library using the oligonucleotides: AGAGCATTTGGGCT-

TAAGCC and GGGAGCTATAGAGCATATGT, and cloned into pTOPO vector using the TA Cloning Kit Dual Promoter according to the manufacturer's instructions. For sense probes, the vector was linearized by XhoI digestion. Sense probes were synthesized using Sp6 RNA polymerase. For antisense probes, the vector was HindIII linearized prior *in vitro* transcription using T7 RNA polymerase. The correct size of the probes was confirmed on an agarose gel. For FISH, neural stem cells were grown on Poly-A-coated coverslips, washed with PBS, and fixed in 4% PFA at room temperature for 10 min. Thereafter cells were washed twice in PBS and incubated with pre-hybridisation buffer (50% formamide, 300 mM NaCl, 30 mM tri-sodium-citrate, 0.02% BSA, 2 mM Vanadyl Complex (NEB), 0.1% Triton, 0.2 mg/ml yeast tRNA) for 1 h at 65°C. RNA probe (1-2 ng/μl) was denatured in hybridization buffer (pre-hybridization buffer + 10% Dextran sulfate) for 10 min at 80°C, added to the cells, and hybridization was performed at 65°C overnight. The following day the cells were washed twice for 30 min with wash buffer 2 (50% formamide, 300 mM NaCl, 30 mM tri-sodium-citrate, 0.1% Triton) and once with wash buffer 3 (300 mM NaCl, 30 mM tri-sodium-citrate, 0.1% Triton) at 65°C. After washing, the blocking of samples was performed with 200 μl FISH blocking buffer (1% Perkin Elmer Blocking reagent in TNT solution (0.1 M Tris-HCl, 0.15 M NaCl, 0.1% Tween)) for 1 h at room temperature. Blocked samples were incubated with Anti-Fluorescein-POD Fab fragments (Roche) in FISH blocking buffer (1:50) overnight at 4°C. The next day cells were washed six times for 5 min in TNT solution, and cells were equilibrated in TSA Plus Amplification Diluent for 10 min. Cells were then incubated with Fluorescein Fluorophore (diluted 1:25 in TSA Amplification Diluent) in the dark for 4 h. Afterward, cells were stained with DAPI (0.5 μg/ml) in TNT solution for 10 min, and images were taken using a Leica DM 8000 fluorescent microscope.

5.2.11 Subcellular fractionation

Subcellular fractionation was adapted from Gagnon et al. [153]: Briefly, cells were harvested in ice-cold PBS and after that lysed in ice-cold hypotonic lysis buffer (HLB; 10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.3% NP-40, 10% Glycerol). Cells were incubated on ice for 10 min, and the supernatant was collected (= cytoplasmic fraction). The pellet was washed three times in HLB, and ice-cold Modified Wuarin-Schibler buffer (MWS; 10 mM Tris pH 7.5, 4 mM EDTA, 0.3 M NaCl, 1 M Urea, 1% NP-40) was added. Following incubation on ice for 15 min with repeated vortexing, the supernatant was collected (= nucleoplasmic fraction), and the pellet (= chromatin fraction) was washed three times with MWS. The RNA in the cytoplasmic and nucleoplasmic fractions was precipitated with ethanol and 3 M sodium acetate and stored at -20°C. Precipitated RNA was pelleted at 17.000 g at 4°C for 30 min and subjected to RNA purification with the chromatin pellet. First-strand cDNA was generated from 7.5 μl RNA solution random Hexamers primers. RUS-, GAPDH-, and Malat levels were quantitatively accessed by quantitative RT-PCR analysis. Cytosol fraction normalization and overall enrichment were calculated as followed:

$$\triangle\triangle\, Ct_{fraction,target} = \triangle\, Ct_{fraction,target} - \frac{1}{n_{ctosol,target}} \sum_{i=1}^{n_{ctosol,target}} \triangle\, Ct_{cytosol,target}$$

$$\tilde{x}_{fraction,target} = \frac{1}{n_{fraction,target}} \sum_{i=1}^{n_{fraction,target}} {}^{\triangle\triangle Ct_{fraction,target}}$$

$$normalizedAbundance_{fraction,target} = 2^{\tilde{x}_{fraction,target}}$$

$$content_{fraction,target} = 100 * \frac{normalizedAbundance}{\sum_{i=1}^{n_{fractions}} normalized_{fraction}}$$

5.2.12 RNA-Seq analysis

RNA-Seq analysis was carried out on two different time points in quadruplicates with one scramble shRNA and two independent shRNA targeting RUS according to standard protocol [154]. 300.000 neural stem cells were seeded on 35 mm culture dishes per time point and replicate & cultured as previously described. One day after seeding, cells were transduced with 5 μl purified virus per well. Two days after plating, the medium was replaced with medium containing 1.4 μg/ml Puromycin. The first samples were harvested five days after seeding with Isol-RNA Lysis Reagent. On day five, differentiation was initiated, and 2nd samples were collected on day 7. Total RNA was isolated, and polyA enriched. After reverse transcription, the cDNA was fragmented, end-repaired, and A tailed. Solexa sequencing adaptors were ligated, and adaptor -modified fragments were enriched by 18 cycles of PCR amplification. Quantity and the sequencing library's size was accessed on a Bioanalyzer before sequencing on an Illumina NExtSeq 500 platform. Sequencing reads from FASTAQ files were aligned using the STAR Aligner version 2.54. The reference genome used for alignment was constructed using the mm10 fasta file and GRCm38.99 transcript table. For quantitative analysis, gene counts and transcripts per million (TPM) were obtained using STAR 2.45 and RSEM, respectively [141, 142].

5.2.13 Preparation of nuclear Extract

Nuclear extract was prepared without dialysis according to Dignam, et al. 1983 [155]: Cell Pellet was resuspended in 5 volumes buffer A (10 mM HEPES (pH 7.9 at 4C), 1.5 mM $MgCl_2$, 10 mM KC1, 0.5 mM DTT, 200U/ml murine RNAse Inhibitor) and incubated for 10 min at 4°C. Cells were homogenized with a Dounce tissue grinder, and cell nuclei were pelleted at 500 g for 10 min at 4°C. Cell nuclei were washed again in 5 volumes buffer A and dissolved in 1 volume buffer C (20 mM HEPES pH 7.9, 25% (v/v) glycerol, 0.42 M KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM

DTT, 200 U/ml murine RNAse Inhibitor) and homogenized again with a Dounce tissue grinder. After gentle rotation at 4°C for 30 min, chromatin was pelleted at 17.000 g and 4°C for 30 min. The supernatant was taken and used for affinity purification. Salt concentration was reduced with 1 volume buffer G (20 mM HEPES pH 7.9, 20% (v/v) glycerol, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT, 200 U/ml murine RNAse Inhibitor). According to the manufacturer's instruction, the protein amount was determined using the BCA method using a BSA standard ranging from 50 μg/ml to 2 mg/ml, 1:200 diluted. Emission at 562 nm was measured on a plate reader.

5.2.14 RNA Affinity Purification from nuclear extract with recombinant MS2BP-MBP

Per Replicate 760 pmol yeast t-RNA and 150 pmol recombinant MS2BP-MBP-His_6 was added to 1 mg nuclear extract isolated from 5xMS2 tagged overexpressing Neuro2A cells. After 2 h gentle rotation at 4°C, RNA-protein complexes were coupled to 50 μl preequilibrated amylose resin. After 2 h gentle rotation at 4°C, Resin was pelleted at 1900 g for 1min at 4°C and three times washed with 900 μl buffer D (20 mM HEPES pH 7.9, 20% (v/v) glycerol, 0.1 M KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT).

5.2.15 Tandem mass spectrometry for quantitative analysis of lncRNA interaction proteins

Interacting proteins were eluted with 50 μg RNAse dissolved in 80 μl buffer D at 37°C for 10 min. The resin was pelleted again at 1900 g for 1min at 4°C, and the supernatant was subjected to Filter Aided Sample Preparation [156]. The amount of protein was estimated by UV light absorption at 280 nm using an extinction coefficient of 1.0. Proteins were loaded on a filter device, reduced with DTT, and alkylated with iodoacetamide. Peptides were eluted by LyC and Trypsin digestion from the filter and applied to a high-performance liquid chromatography column and measured on a Q-Exactive-Orbitrap instrument (Thermo Fisher Scientific) by using a gradient of 0-40% acetonitrile in 150 min. Raw Data were quantified with MaxQuant software4 using the mouse genome as the reference genome. Peptides identified by site, reverse matching peptides & contaminants were removed before further analysis by student's t-test using R.

5.2.16 Western blotting analysis of lncRNA-protein complexes

Interacting proteins were eluted with 10 mM maltose dissolved in 80 μl buffer D at 4°C for 15 min. The resin was pelleted again at 1900 g for 1 min at 4°C, and the supernatant was subjected to western blot analysis. Per western-blot analysis, one 8-12% SDS-polyacrylamide (8-12% Bisacryl-/acrylamide 37.5:1, 0.5 M Tris-HCl pH 8.8, 1% SDS,1% m/v APS, 0.1% v/v TEMED) resolving gel was prepared and equipped with

a stacking (4% Bisacryl-/acrylamide 37.5:1, 62.5 mM Tris-HCl pH 6.8, 1% SDS,1% m/v APS, 0.1% v/v TEMED). 1 μg input protein or 50% of eluate were denatured in one volume 2x Laemmli loading-buffer (126 mM TRIS/HCl pH 6.8 , 20% v/v glycerin, 4% (m/v) SDS, 10% (v/v) β-Mercaptoethanol, 0.02% Bromphenolblau) and loaded on the gel. The gel was run in an electrophoresis chamber equipped with 1x running buffer (25 mM Tris, 0.189 M glycine, 0.1% SDS) at 90V. Proteins were blotted onto PVDF membrane using the "tank-blot" procedure in 1x blotting buffer(20 mM Tris, 0.15 M glycine, 20% (v/v) at 140 mA for 1h. After transfer, the membrane was quickly washed with TBS-T (20 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% Tween 20). The membrane was incubated with specific antibodies after blocking in blocking-buffer (0.2% I-Block in TBS-T) for 1 h to detect proteins. The membrane was then incubated with a secondary antibody coupled to horseradish peroxidase. After every blocking step, the membrane was washed three times with TBS-T for 10 min. Horseradish peroxidase activity was detected using ECL Western Blotting Detection reagent under a BioRad ChemiDoc imager.

5.2.17 RNA immunoprecipitation (RIP)

35 μl Protein G Agarose beads were equilibrated in buffer D (20 mM HEPES pH 7.9, 20% (v/v) glycerol, 0.1 M KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT) and blocked overnight with 1% BSA. Nuclear extract from 5 Mio neural stem cells was treated with 760 pmol yeast t-RNA, 300 μg salmon sperm DNA & 4 μg antibody for 2h at 4°C under gentle rotation. Anti-rabbit IgG was used as a negative control. The pretreated extract was added to blocked Protein G Agarose and further incubated for 2 h at 4°C under gentle rotation. Afterwards Protein G agarose beads were sedimented at 4500 g for 2 min at 4°C and washed 5x with 900 μl buffer E (20 mM HEPES pH 7.9, 20% (v/v) glycerol, 0.3 M KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT). Bead material was resuspended in 800 μl Isol-RNA Lysis Reagent, and RNA was extracted. First-strand cDNA was generated from 7.5 μl RNA solution using random Hexamers primers. RUS levels and TBP levels were quantitatively accessed by quantitative RT-PCR analysis. IgG control normalized enrichment was calculated as followed:

$$\triangle\triangle Ct_{antibody,tartet,replicate} = \triangle Ct_{anitbody,target,replicate} - \frac{1}{n_{IgG,replicate}} \sum_{i=1}^{n_{IgG,target,replicate}} \triangle Ct_{IgG,target,replicate}$$

$$\tilde{x}_{antibody,target,replicate} = \frac{1}{n_{antibody,target,replicate}} \sum_{i=1}^{n_{antibody,targetreplicate}} \triangle\triangle Ct_{antibody,target,replicate}$$

$$normalizedAbundance_{antibody,target,replicate} = 2^{\tilde{x}_{antibody,target,replicate}}$$

5.2.18 Chromatin Isolation by RNA purification (ChIRP)

NSCs from 8x 15cm dishes were harvested by trypsinization and 2x washed with 100 ml PBS. Cells were crosslinked in 100ml 1% Glutaraldehyde for 10 min at RT. Crosslink reaction was quenched 125 mM Glycine for 5 min. Cells were pelleted at 1000 g for 5 min. ChIRP experiment was done according to standard protocol [117]. Crosslinked cells were washed two times in PBS and lysed in 2 ml ChIRP-lysis buffer (50 mM Tris-HCl pH 7.0, 10 mM EDTA, 1% SDS, 1 mM PMSF, 1x protease inhibitor, 100 U/ml SuperaseIn). Chromatin shearing was conducted on a water-bath sonicator for 6 h at 4°C. This procedure typically yielded 150-600 bp sheared chromatin. 2x 10 μl of sheared chromatin was taken as RNA input control, and DNA input control. Sheared chromatin was diluted with 4 ml ChIRP-hybridization buffer (50 mM Tris-HCl pH 7.0, 750 mM NaCl, 15% (m/v) formamide, 1 mM EDTA, 1% SDS ,1 mM PMSF, 1x protease inhibitor, SuperaseIn 100U/ml) and divided into 2 pools. Every 100 nt lncRNA one 20 nt long complementary oligonucleotide was designed for RUS-isoform 1 using the www.singlemoleculefish.com tool and ordered as 3' biotinylated deoxy-oligonucleotide:

Table 22: Odd probe set

probe #	5' ->3' sequence
1	AAAGGCTTGTACGCTGAGTG
3	AGTGAAAGATGAGCACCAGC
5	CTGTAACCACAGTGCAGTGG
7	CCAACCAATCCTCAGAACAT

Table 23: Even probe set

probe #	5' ->3' sequence
2	AGAGTTCCAGTGAGTAGTTT
4	GGAAAGTTTAGTCAGAATTG
6	ACTTTTAGGATTAGTGCACA
8	GGTAGGAGCCATAGTAAAGA

100 pmol odd-biotinylated-probe set containing all probes labeled with odd numbers was added to the first pool, and 100 pmol of the even-biotinylated probe set containing all even labeled probes was added to the second pool. Probes were hybridized at 37°C under continuous rotation. After 4 h hybridization, 100 μl of in ChIRP-Lysis buffer preequilibrated magnetic streptavidin beads were added. Captured RNA-chromatin complexes were coupled to beads for 30 min at 37°C under continuous rotation. Afterward, beads were washed 5x in 1 ml ChIRP-wash buffer (300 mM NaCl, 30 mM tri-sodium-citrate, 0.1% SDS, 1 mM PMSF) for 5 min at 37°C. 10% of washed beads were resuspended in 100 μl PK buffer (10 mM Tris-HCl pH 7.0, 100 mM NaCl, 1 mM EDTA 0.5% SDS) to isolate the captured RNA. Therefore, beads were incubated together with 100 μg Proteinase K at 55°C for 45 min, and 95°C for 10 min to reverse the crosslinks. RNA was purified according to standard protocol. Residual 90% beads were resuspended in 150

μl DNA-elution buffer (50 mM NaHCO3, 1% SDS) and treated with 100 μg RNAseA and 100 U RNAseH to digest captured chromatin complexes from beads. Beads were incubated for 30 min at 37°C and separated to collect the flowthrough. This elution procedure was repeated once. Eluate- and input-DNA diluted in PK buffer-DNA (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA 0.5% SDS) were treated with 300 μg Proteinase K for 45min at 55°C. Precipitated DNA was purified with a PCR clean-up column purification kit according to the manufacturer's protocol. DNA was submitted to ChIP-DNA sequencing to EMBL-GenoCore facility.

5.2.19 Chromatin Immunoprecipitation (ChIP)

Neural stem cells were crosslinked with 1% formaldehyde for 10 min at RT. The Crosslinking reaction was quenched with 300 mM glycine. Cells were washed 2x with PBS. Cell nuclei were prepared with Covaris TrueChIP Kit according to the manufacturer's instruction. Cell nuclei were sonicated in 850 μl ChIP-shearing buffer (10 mM Tris-HCl pH 7.6, 1 mM EDTA, 0.1% SDS) on a M220 Covaris sonicator using the settings: PIP: 75W, duty factor: 10% and CBP: 200 for 8 min. This procedure typically yielded chromatin fragments of 100-700 bp size. Sheared chromatin was cleared at 17.000 g for 5min at 4°C and diluted with 850 μl 2x IP dilution buffer. ChIP experiments were done according to standard protocol [157]. 850 μl of diluted chromatin was used per ChIP experiments and treated with 4 μg antibody overnight. 35 μl Protein G Agarose beads were equilibrated in RIPA buffer and blocked overnight with 1% BSA. Pretreated chromatin was added to blocked beads, and for further 4 h rotated at 4°C. Protein G beads were sedimented at 4500 g and first 3x washed with 900 μl high salt RIPA buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 1% SDS) for 10 min each and finally 2x washed with 900 μl TE(10 mM Tris-HCl pH 8.0 & 1 mM EDTA). Beads were resuspended in 200 μl TE supplemented with 1% SDS treated with 10 μg ProteinaseK for 2 h at 55°C. Crosslinked were reversed overnight at 65°C. Immunoprecipitated DNA was purified with a PCR clean-up column purification kit according to the manufacturer's protocol. DNA was submitted to ChIP-DNA sequencing to the EMBL-Gene Core facility or analyzed by quantitative RT PCR. ChIP Ct values were normalized against input Ct values considering the ChIP and input sample's dilution coefficient to calculate the overall enrichment as followed:

$$enrichment = 2^{\delta Ct_{input} - \delta Ct_{eluate} - log_2(\frac{dilutionFactor_{input}}{dilutionFactor_{eluate}})}$$

5.2.20 ChIP-DNA sequencing

Isolated DNA was ended repaired with T4 DNA polymerase and polynucleotide kinase. An A-base was added to end-repaired DNA fragments. Solexa sequencing adaptors were ligated, and adaptor modified fragments were enriched by 18 cycles of PCR amplification. The sequencing library's quantity and size were accessed on a Bioanalyzer before sequencing on an Illumina NExtSeq 500 platform. Sequencing reads from FASTQ files

were aligned with bowtie2 to mm10. Chip Peaks were called with MACS1.4.

5.2.21 Electromobility shift assay (EMSA) to detect Triplex RNA-DNA and Duplex RNA structures

Fluorescent labeling of oligonucleotides

Oligonucleotides were purchased with a 5' amino modification (5AmMC6), desalted, and dissolved in 1:20 bicarbonate buffer (20 mM $NaHCO_3$/ Na_2CO_3 pH 9.0): PBS. 25 nmol desalted oligonucleotides were together with 75nmol 5(6)-Carboxytetramethylrhodamine-succinimidyl-ester (NHS-TAMRA,) or 5/(6)-Carboxy-fluorescin-succinimidyl ester (NHS-FTIC) overnight. Fluorescent labeled oligonucleotides were precipitated in 0.3 M sodium acetate (pH 5.2) and 70% EtOH. Pellett was washed two times in 70% EtOH, air-dried, and resolved in nuclease-free 100 μl H_2O.
Following oligonucleotides were used:

Table 24: oligonucleotides used by EMSA assay

oligonucleotide	5'->3' sequence
TTS_motif_10repeats_fw	/5AmMC6/TTCCATTCCATTCCATTCCATTCCA TTCCATTCCATTCCATTCCATTCC
TTS_motif_10repeats_rv	GGAATGGAATGGAATGGAATGGAATGGAATGG AATGGAATGGAATGGAA
TFO-RNA	/5AmMC6/CCUUaCCUUaCCUUaCCUUaCCUUaCCUU
GGUAA_repeat	AAGGUAAGGUAAGGUAAGGUAAGGUAAGG

RNA-DNA triplex

2 μM unlabeled DNA-oligonucleotide and 2 μM TAMRA labeled reverse complementary DNA-oligonucleotide were heated to 95°C in 20 mM Tris-acetate pH 7.4 and 20 mM $MgCl_2$ and slowly cooled down. For triplex formation assay, 1 μl of a DNA annealing mix was treated with an increasing amount of FITC labeled triplex-forming RNA-oligonucleotide in 90 mM Tris-acetate pH 5, 150 mM KCl, and 20 mM $MgCl_2$ or 20 mM Tris-acetate pH 7.4, 150 mM KCl, and 20 mM $MgCl_2$ incubated for 2 h at RT.

RNA-RNA duplex

8 μM unlabeled GGUGGx6 RNA and 8 μM FITC labeled triplex-forming RNA-oligonucleotide were heated to 95°C in 20 mM Tris-acetate pH 7.4 and 20 mM $MgCl_2$ and slowly cooled down.

Samples were treated with purple loading dye (NEB) and run on native polyacrylamid (12% Bisacryl-/acrylamide 19:1, 90 mM Tris-acetate pH 5, 20 mM $MgCl_2$) or (12% Bisacryl-/acrylamide 19:1, 20 mM Tris-acetate pH 7.4, and 20 mM $MgCl_2$) in 1x TBE-buffer. Gels were evaluated with a BioRad ChemiDoc imager.

6 Results

6.1 Identification of RUS

6.1.1 Selection of potential candidates

To select lncRNAs expressed primarily in neurons, I focused on two publicly available transcriptome data sets. The first data set I used is the murine brain atlas from the Chu and Barres lab [158]. This atlas is a comprehensive transcriptome analysis of neural cell types isolated by FACS and immune panning methods. I extracted from this data set the expression of lncRNAs in (1) oligodendrocyte precursor cells (OPC), (2) newly formed oligodendrocytes, (3) myelinating oligodendrocytes (MO), (4) microglia, (5) endothelial cells, (6) astrocytes, and (7) neurons having a FPKM (Fragments Per Kilobase Million) value greater than 1.5 in at least one sample. Using the euclidean distance method, I performed a simple clustering approach with z-scores on this data set (see Figure 13). This approach yielded a cluster of about 50 neuron-specific lncRNAs. Besides the murine brain atlas, I used a human transcriptome data set of lncRNA expressed in embryonic stem cells, *in vitro* differentiated -neural precursors, -neurons, and -astrocytes [159]. From this data set, I selected about 30 transcripts whose expression is upregulated during differentiation. I started to evaluate those candidates using online available tools to classify them according to the following features: transcripts that (1) lack any coding capacity, (2) are conserved between mouse and human by synteny and sequence, and (3) are expressed in neural tissues only. For coding capacity, I used the conventional ORF prediction tool (https://www.ncbi.nlm.nih.gov/orffinder/). For conservation, I used the ensemble genome browser (http://www.ensembl.org/index.html) and the online Blast tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Tissue expression was analyzed using the EBI gene-expression atlas (https://www.ebi.ac.uk/gxa/home). By filtering, I ended up our *in silico* analysis with six candidates. Usually, I named the lncRNA according to its neighboring protein-coding gene, if available. From the murine brain atlas data set I selected following five lncRNAs: mir 124-hg1 (lnc00599), mir 124-hg3 (Gm14342), Gjd2lncRNA, Cdr1lncRNA, and Islr2lncRNA [158]. The sixth lncRNA - Slitrk3lncRNA- I selected from the human differentiated neural cell type data set [159]. LncRNAs as mir 124a-hg1 and mir124a-hg3, act as precursor RNAs for the neuron-specific miRNA 124a.

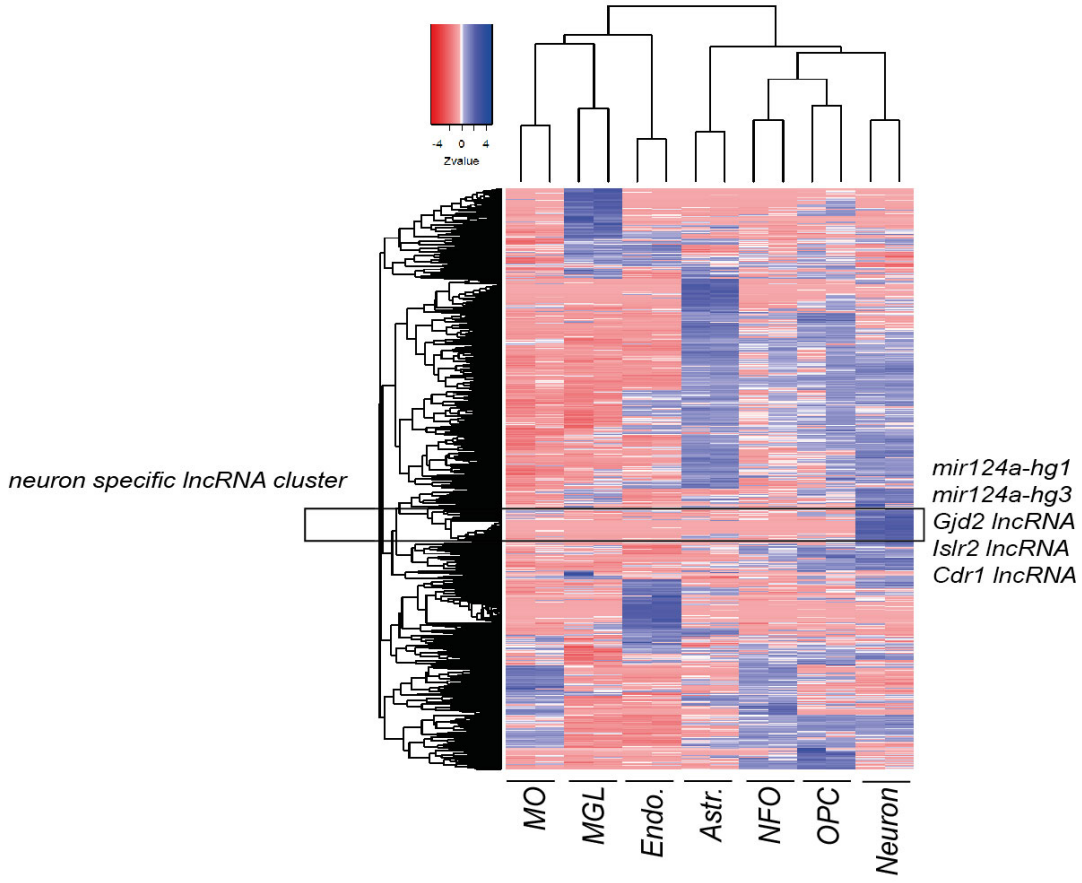


Figure 16: Clustering approach using the euclidean distance method of lncRNAs expressed in neural cell types: MO, MGL, Endo, Astr, NFO, OPC, and Neurons with z-scores yielded a cluster of about 50 lncRNAs specific expressed in Neurons. Epression values extracted from [158].

Next, I verified the neural-specific expression by comparing the expression in different adult (A-) and embryonic (E-) murine tissues using quantitative RT-PCR (Figure 17). For this, the cortex (Cor), cerebellum (Cer), and hippocampus (Hip) from adult and embryonic animals were dissected. Kidney, heart, lung, muscle, skin, and spleen were only dissected from adult animals. If available, I analyzed the expression of the selected lncRNAs and the expression of the corresponding genomic-neighboring protein-coding gene that could be regulated by a *cis*-acting mechanism. For quantitative RT-PCR, I isolated and reversely transcribed RNA to cDNA using Oligo dT primers and first normalized the expression of all lncRNAs against the housekeeping gene TBP. Second, the abundance was normalized to the lncRNA expression in the adult cortex if no neighboring protein-coding gene was available (mir124a-1hglncRNA and mir124a-3hglncRNA). In case a neighboring protein-coding gene exists, the abundance was normalized to the neighboring protein's expression in the adult cortex (Cdr1-, Gjd2-, Islr2- Sitrk3 - lncRNA).

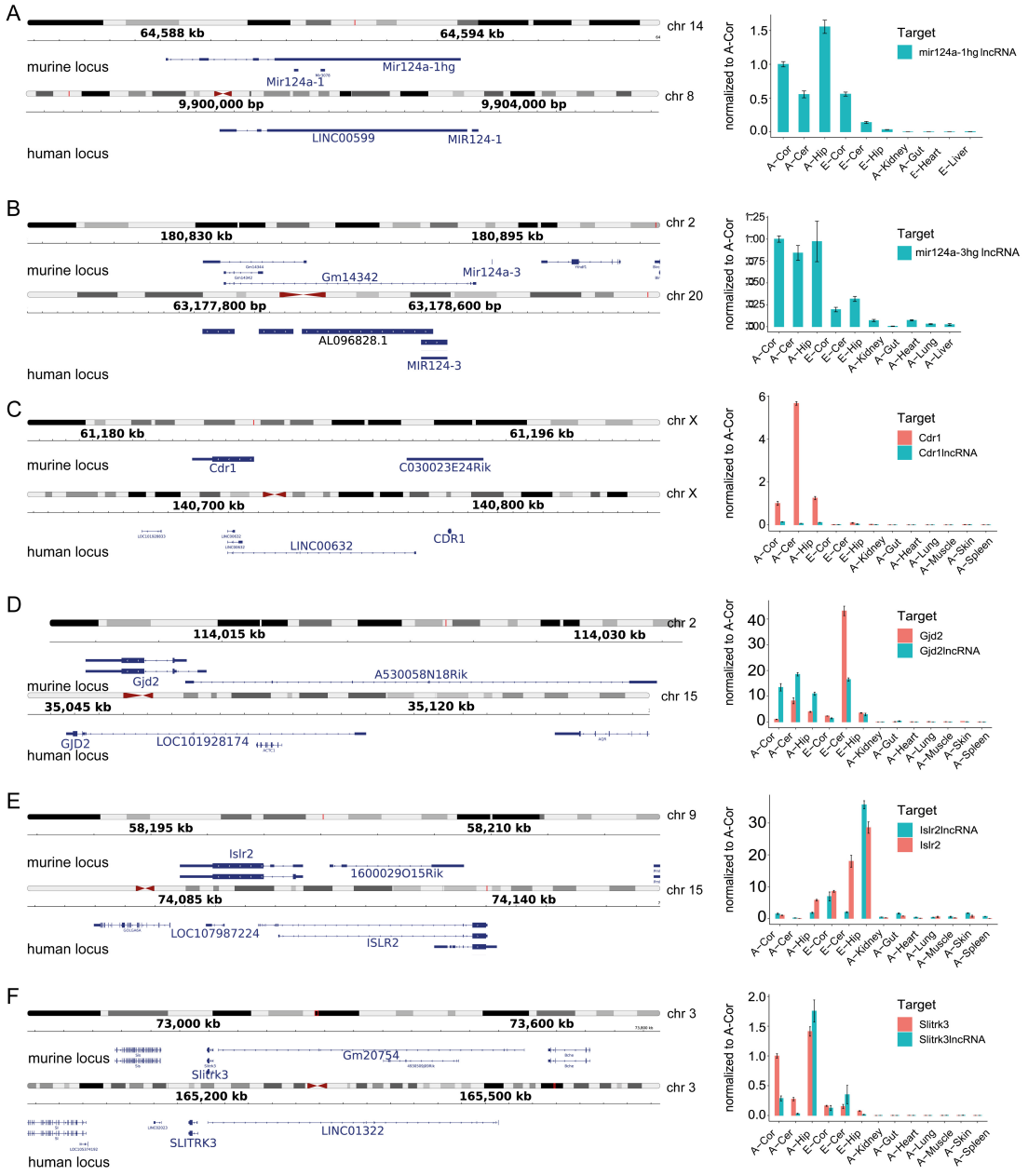


Figure 17: **Genomic location in mouse (top left panel) and human (bottom left panel) and the expression (right panel) of selected lncRNAs:** mir 124a-gh1 **(A)**, Cdr1lncRNA **(B)**, mir124a-hg3 **(C)**, Gjd2lncRNA **(D)**, Islr2lncRNA **(E)**, and Slitrk3lncRNA ((**F**)) in different adult(A-) and embryonic (E-) murine tissues: cortex (Cor), cerrebellum (Cer), hippocampus (Hip), kidney, gut, heart, lung, liver, muscle, skin, and spleen. Black and gray-striped bars represent chromosomes. Red box on chromosome highlights position within the chromosome.

In line with our hypothesis, all lncRNAs were expressed in neural tissues only. Neural tissues dissected from adult animals showed the highest expression for mir124a-hg1 (Figure 17A) and mir124a-hg3 (Figure 17B). Compared to mir124a-hg3, mir124a-hg1 is highest expressed in the adult hippocampus. These results coincide with the observation that mir124a's expression is upregulated during neuronal differentiation [160]. Similar to both transcripts, Cdr1lncRNA (Figure 17B), Gjd2lcRNA (Figure 17D), and Slitrk3lncrNA (Figure 17F) are also neural-specifically expressed and have the highest expression in adult tissues. Gjd2 and Cdr1 lncRNA are most expressed in the adult cerebellum. Slitrk3lncRNA is expressed at the highest level in the adult hippocampus. In contrast to all five lncRNAs, Islr2lncRNA (Figure 17E) is most expressed in embryonic neural tissues and indicates that Islr2lncRNA may have an essential regulatory function in neural stem cells. Although the expression of the neighboring protein-coding gene is also neural-specific, I observed in no case a correlation of expression between lncRNA and adjacent protein-coding genes. Hence, I cannot conclude from the expression data for a *cis*-regulatory mechanism. All selected lncRNA candidates are expressed primarily in neural tissues confirming that our strategy to choose neuron-specific lncRNA was appropriate. Next, we wanted to analyze these lncRNAs on a functional level.

6.1.2 Establishment of a cellular model to study neurogenic lncRNAs

To study the function of the selected lncRNAs, I established neural stem cells (NSC) as a monolayer cell culture. NSC were isolated from murine cortices at stage E15 - E16 and kept in a medium containing basic FGF. Every second day, basic FGF was added to the culture to maintain cell proliferation. FGF was withdrawn when cells were fully confluent to differentiate NSC, typically on day 4 or 5 after seeding.
I systematically investigated our cell model's differentiation behavior by analyzing the expression of neuronal and glial markers. Nestin served as a marker for neuroectodermal cells. Fibrillary-acidic protein (GFAP) and the excitatory amino acid transporter 1 (Glast1) were used as markers for neural-stem and glia cells. β-tubulin III, microtubulin-associated protein tau (Mapt), and microtubulin-associated protein 2 (Map2) served as markers for neurons.

Immunofluorescent was used to visualize nestin, GFAP, β tubulin III, and Mapt gene expression on days 5, 7, and 9 after seeding. During the time course of differentiation, an increase of β-tubulin III and Mapt positive cells and a decrease of Nestin and GFAP positive cells was observed (Figure 18A).
Quantitative RT-PCR analysis was performed with primers against nestin, Glast1, GFAP, β-tubulin III, and Map2 with RNA samples isolated on days 1, 3, 5, 7, and 9 after seeding. A 3.2x fold decrease of Nestin expression was observed over the time course between day 1 and 3 (Figure 18B). A strong decrease was also observed for GFAP (Figure 18C) and Glast1 expression (Figure 18D). GFAP decreased by 3x fold on day

3, and Glast1 by 1.9x fold on day 5. In turn, β-tubulin III (Figure 18E) and Map2 (Figure 18F) expression increased by 3.4x fold and 1.7x fold, respectively.
Both immunofluorescence and quantitative RT-PCR marker gene expression analysis were in line with the differentiation of neural stem cells towards neurons.

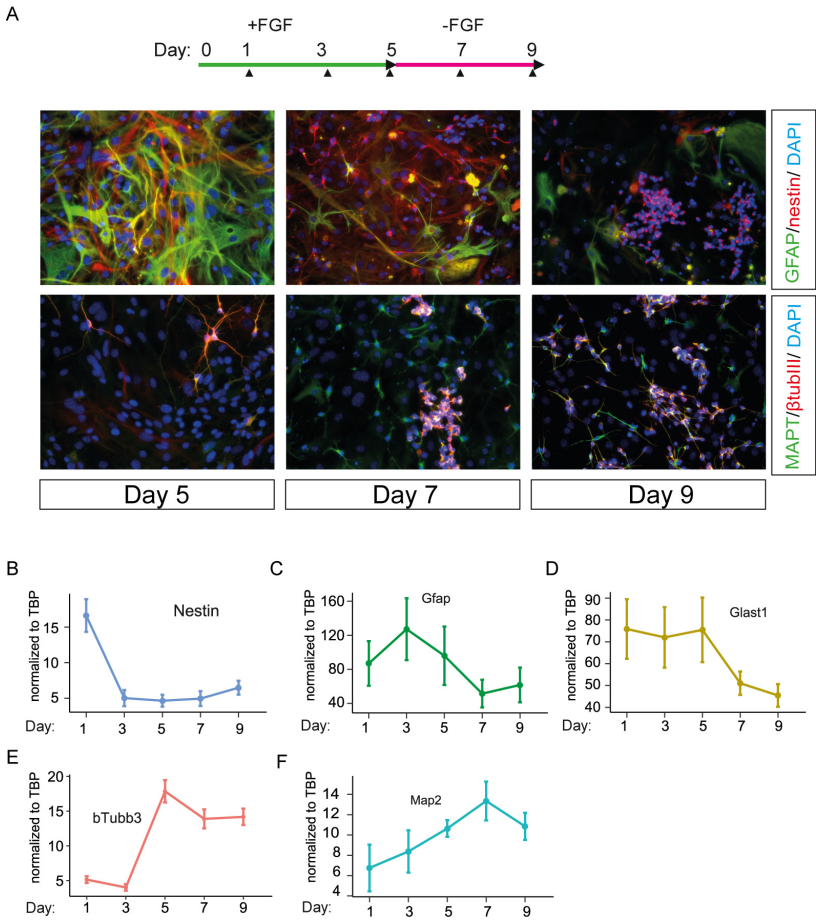


Figure 18: **Characterization of** *in vitro* **differentiation cell model. A:** immunofluorescent analysis against Nestin, GFAP, β-tubulin III , and Mapt. Cell nuclei are stained with DAPI. B-F: quantitative RT-PCR analysis: **B:** Nestin, **C:** GFAP, **D:** Glast1, **E:** β-tubulin III, and **F:** Map2. Expression values are normalized to the Expression of housekeeping gene TBP. The experiment was performed in biological triplicates. Error bars represent the standard error of the mean (SEM).

Next, I analyzed by quantitive RT-PCR whether our selected lncRNA candidates are expressed in the established *ex vivo* differentiation cell model (Figure 19). Except for Gjd2lncRNA (Figure 19B), all lncRNAs significantly changes expression in the time course of differentiation between day 1 to 9. Mir124a -hg1 (see Figure 19D) and mir124a-hg3 (Figure 19E) expression increased by 6.6x fold and 4.2x fold, respectively. Similarly, the abundance of Cdr1lncRNA (Figure 19A) and Slitrk3lncRNA (Figure 19F) increased by 1.6x fold and 25x fold, respectively. In contrast, expression of Islr2lncRNA

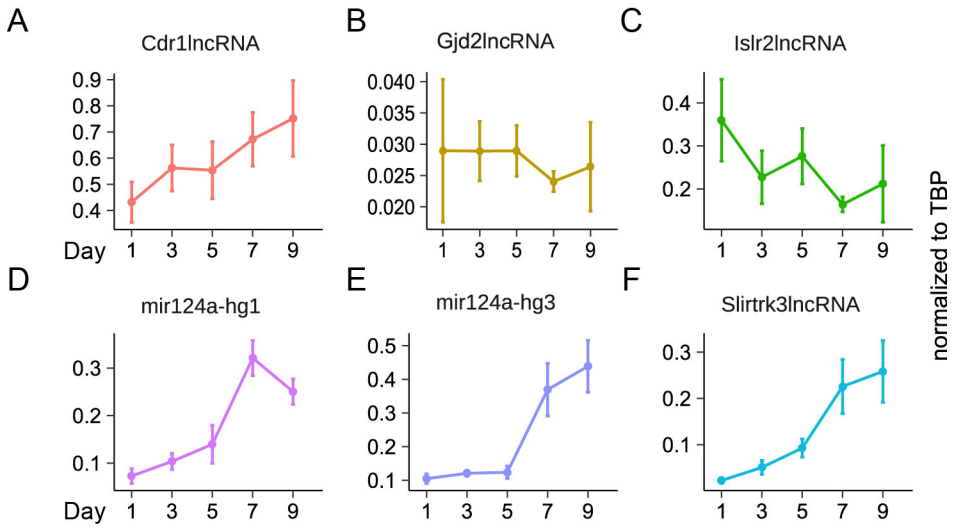


Figure 19: **LncRNA expression in the *in vitro* differentiation cell model.** Quantitative RT-PCR analysis: **A:** Cdr1lncRNA, **B:** Gjd2lncRNA, **C:** Islr2lncRNA, **D:** mir124a-hg1 **E:** mir124a-hg3, and **F:** Slitrk3lncRNA. Expression values are normalized to the expression of housekeeping gene TBP. The experiment was performed in biological triplicates. Error bars represent the standard error of the mean (SEM).

(Figure 19C) decreased by 2.3x fold. These observations are in line with the quantitative RT-analyses in murine tissues (Figure 17). Mir124a-hg1 and -3, Cdr1lncRNA, and Slitrk3lncRNA are at the highest expressed in adult neural tissues, and Islr2lncRNA is most abundant in embryonic neural tissues.
Taken together, the established *ex vivo* model recapitulates the cellular hallmarks of neuronal differentiation and expresses five from six selected candidates. Hence, the NSC model turned out to be a valid model to study neurogenic lncRNAs *ex vivo*.

6.1.3 Knockdown of lncRNA by shRNAs in the *ex vivo* model of neuronal differentiation

In order to investigate the relevance of pre-selected lncRNAs in the development and generation of neurons from neural stem cells, we decided to reduce the expression of the respective lncRNA by lentivirally transduced shRNAs in a knockdown (KD) experiment (Figure 20). The KD of neuronal relevant lncRNAs should reduce the neurons' fraction in our *ex vivo* model. As a control to determine the fraction of neurons without KD, we used a control shRNA (scrambled RNA).
Therefore, shRNAs were cloned upstream to the strong U6 promoter into pLKO.1 vector harboring GFP as a selection marker under the human phosphoglyceratekinase (hPKG) promoter. Lentiviral particles were produced according to standard protocol and administered one day after plating [127]. As previously described, I withdrew basic FGF to start differentiation on day five when cells became 90-100% confluent. Cells were

further kept for four days in culture and stained with antibodies against neural markers such as β-tubulin III to analyze the neurogenic phenotype after KD (Figure 20). First, I tested different shRNAs for their KD efficiency by quantifying the lncRNA expression via quantitative RT-PCR.

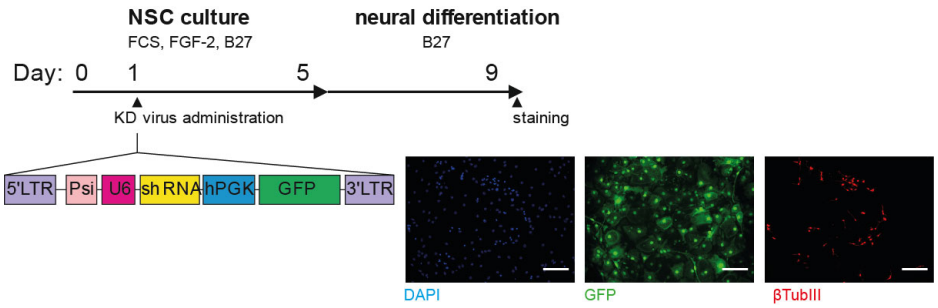


Figure 20: Phenotypic screening of neural stem cells after differentiation and Knockdown of lncRNAs. White bars = 50 μm.

Selection for efficient KD shRNAs

To test the efficiency of shRNAs to knockdown lncRNAs, I produced KD-virus encoding for the Puromycin resistance gene instead of GFP and added the virus one day after seeding. Two days after virus administration, stable KD cells were selected with Puromycin. Seven days after seeding, I analyzed the individual lncRNA expression by quantitative RT-PCR. LncRNA levels were normalized against the housekeeping gene TBP's level and compared against the KD-control's lncRNA level. This screen was performed in 3-4 biological replicate to evaluate KD efficiency statistical by a one-tailed students t-test (Figure 21).

Except for Slirtrk3lncRNA, I tested four different shRNAs for each lncRNA. For Cdr1lncRNA, only shRNA SH-b reduced the expression of Cdr1lncRNA significantly by 52% (Figure 21A). shRNA SH-a, SH-d reduced the expression by 33%. However, the application of shRNA SH-a yielded a higher variation. Hence, I decided to use shRNA SH-b and SH-d for the phenotypic analysis after KD. All four tested Islr2lncRNA targeting shRNAs reduce Islr2lncRNA's expression by 37-64% (Figure 21B). For further analysis, I selected the most effective shRNA SH-a and SH-d, reducing the Islr2lncRNA abundance significantly by 60 and 64%, respectively. For mir124a-hg1, I obtained three efficient shRNAs: SH: -a, -c, and -d, reducing mir124a-hg1 level significantly by 96, 88, and 91%, respectively (Figure 21C). Thus, I selected shRNA SH-a and -d for analyzing the impact on neuronal differentiation. Regarding mir124a-hg3, only one shRNA significantly reduced the level of the mir124a-hg3lncRNA by 77% (Figure 21D). Since shRNA SH-d reduced the level of mir124a-hg3a by 38%, I also used this shRNA to analyze the phenotype after KD. For Slitr3klncRNA, I obtained two efficient shRNAs: SH-a and -d lowering the abundance of Slitrk3lncRNA significantly (Figure 21E) 63% and 56%,

respectively. After this efficiency test, I renamed the two selected shRNAs per lncRNA with No.1 and 2 and used them for the neurogenic phenotype screen.

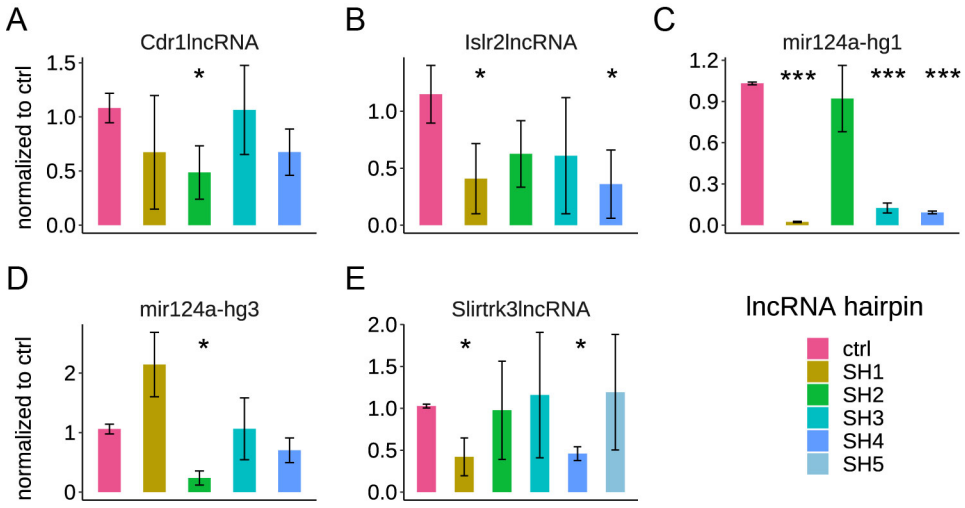


Figure 21: **Sh-RNA selection for efficient lncRNA KD. A:** Cdr1lncRNA, **B:** Islr2lncRNA, **C:** mir124a-hg1, **D:** mir 124a-hg3 and **E:** Slitrk3lncRNA. Experiments were performed in biological triplicates or quadruplicates. Error bars represent standard deviation. One tailed t-test: *: $p < 0.05$, ***: $p < 0.01$.

Phenotypic screening after KD

In order to analyze the impact of lncRNA KD on neuronal differentiation, the fraction of neurons was determined using fluorescent stainings. Microscope images were analyzed for numbers of total cells, transduced cells, neurons, and transduced neurons.
ImageJ image analysis tool is a highly versatile tool to modify and analyze images. ImageJ is written in Java and offers plenty of useful plugins and tools to analyze and quantify microscope images like water-shedding. Water-shedding treats each pixel's brightness value of a greyscale image as topological height and calculates the ridge top's lines, acting as watersheds in nature. Thus, it is used for particle segmentation, and counting [161]. Especially DAPI stained nuclei are ideal particles for water-shedding and, thus, for automatically assessing the cell number in microscope images, which is implemented in the NucleusCounter PlugIn.
Taking advantage of automatic nuclei counting, I calculated the overlay of GFP and DAPI signal to count transduced cells' nuclei, the overlay of β-tubulin III and DAPI signal to count the neuronal nuclei, and the overlay of all three signals to count transduced neuronal nuclei. Therefore, I developed a new algorithm programmed using Java, which quantified, calculated, and compared the three dyes' relative values. Next, I implemented my algorithm in the NucleusCounter Plugin to calculate the number of cell nuclei in the respective overlay images (Figure 22, see also 9.1).

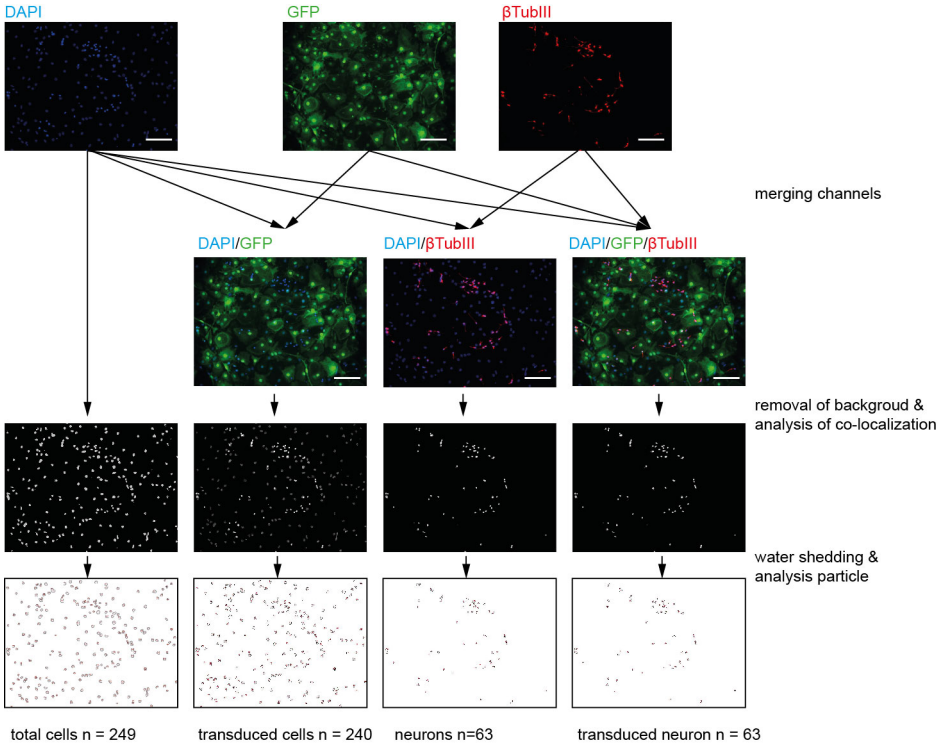


Figure 22: Strategy to develop an ImageJ plugin analyzing the numbre of total cells, transduced cells, neurons, and transduced neurons. White bars = 50 μm.

Reduced levels of neurogenic differentiation after Knockdown of lncRNAs

To search for lncRNA with a neurogenic phenotype after KD, I performed the screen four times as biological replicates and set up three technical replicates per biological replicate. In each biological replicate, the fraction of green fluorescent neurons was determined and normalized to ctrl KD results. Thus, I applied a one-tailed student's t-test for statistical evaluation (Figure 23).

Intriguingly, the KD of all lncRNAs with the two preselected shRNAs resulted in a significant reduction of transduced neurons. After KD of Cdr1lncRNA, I obtained a 28% and 34% reduction of green fluorescent neurons. Islr2lncRNA's KDs decreased them by 48% and 46%. Mir124-hg1's KDs reduced them by 26% and 46%, and mir124hg3's KDs by 37% and 41%. The KD of Slitrk3lncRNA reduced the fraction of green-fluorescent neurons by 54 and 46%.

These results demonstrate that the thoroughly and stringent pre-selection of candidates ensured the success of our screen. All ten tested shRNAs targeting our five selected lncRNAs reduced the fraction of GFP-positive neurons. Although the KD of mir124a-hg1 and -3 resulted in a neurogenic phenotype, I excluded those lncRNAs for further molecular analysis. As host genes for mir124, the KD of mir124a-hg1 and -3 potentially reduced mir124 level. Considering that mir124a is known to inhibit the differentiation from neural stem cells to neurons, the KD phenotypes observed for mir124a-hg1 and -3 could have resulted from the potential reduced level of mir124. The KD of both host

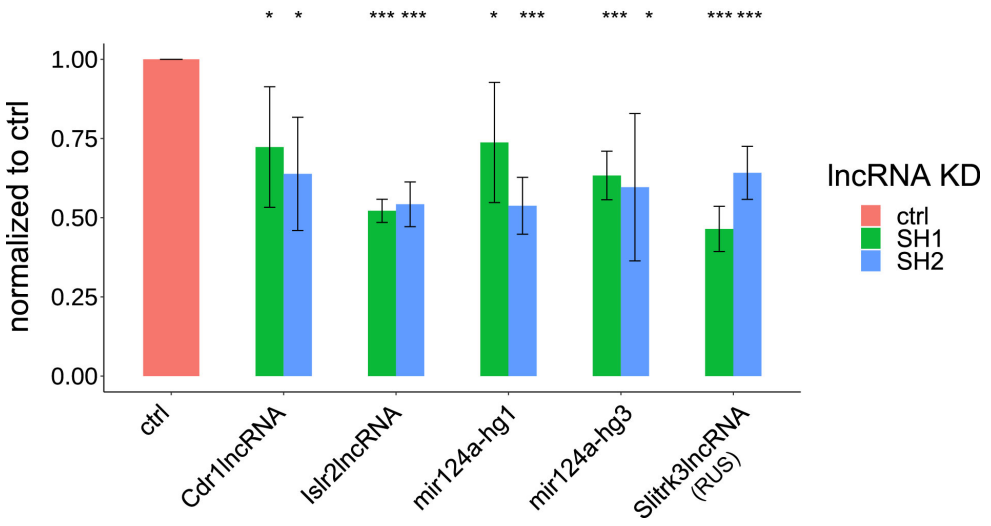


Figure 23: **Result of neurogenic KD screen with GFP KD-virus.** Statistical evaluation of the fraction of green-fluorescent neurons in KD samples compared to scrambled ctrl. Experiments were performed in biological quadruplicates. Error bars represent standard error of the mean. One-tailed t-test: *: $p < 0.05$, ***: $p < 0.01$

genes solely served as positive controls [160]. Since the KD of Cdr1lncRNA resulted only in a weak KD phenotype, I excluded Cdr1lncRNA from further studies. I started to study Slitrk3lncRNA in more detail. After selecting Slitrk3lncRNA, we renamed it with *RUS* (RNA upstream to Slitrk3) for ongoing studies. Studies about Islr2lncRNA were not included and are not part of this dissertation.

Table 25: Summary of selected lncRNAs

lncRNA	data set	highest expressed	expressed in the *ex-vivo* model	KD phenotype
Cdr1lncRNA	[158]	adult cerebellum	yes – increases during differentiation	weak but significant
Gjd2lncRNA	[158]	adult cerrebellum	no	-
Islr2lncRNA	[158]	embryonic hippocampus	yes - decreases during differentiation	strong and highly significant
mir124a-hg1	[158]	adult hippocampus	yes - increases during differentiation	weak but significant
mir124a-hg3	[158]	adult cortex	yes - increases during differentiation	weak but significant
Slitrk3lncRNA (RUS)	[159]	adult hippocampus	yes - increases during differentiation	strong and highly significant

6.2 RUS's function in neurogenesis

6.2.1 Molecular cloning of RUS

RUS annotated as gene-model (GM) 20754 in the ensemble and USCS gene-models is a spliced transcript transcribed from chromosome 3. Located upstream to the promoter of RUS, the protein-coding gene Slitrk3, and downstream to the transcription termination site the protein-coding gene Bche is annotated (Figure 24A).
To isolate the full-length transcript of RUS, I performed RACE (*rapid amplification of cDNA ends*) experiments. Therefore, I purified RNA from the adult hippocampus and designed two reverse primers for the 5'-RACE and one forward primer (3'RACE) for the 3'-RACE. All designed primers are complementary to the annotated exon 2. Performing the 5'- RACE yielded no visible band. For 3'-RACE, the appeared PCR product was cut out and isolated. The isolated DNA sequencing revealed the annotated 3' end (Figure 24). Next, I designed primers annealing to the annotated 5' transcription start site (TSS-FW) and annealing to the annotated and shown 3' end (TTS-RV). Performing a PCR reaction with cDNA isolated from the adult hippocampus yielded one low abundant PCR product at 1.3 kbp and one high abundant at 0.9 kbp (Figure 24C). The sequencing of the 0.9 kbp PCR product revealed an isoform (isoform 1) that missed the annotated exon 4. The sequencing of the low abundant product at 1.3 kbp did not succeed, but according to the PCR product's length, the annotated exon 4 is supposed to be included.
I confirmed our results obtained by 3'-RACE method. Therefore, I designed reverse primers annealing to the start of exon 4 (Ex4-rv) and 5 (Ex5-rv). Performing PCRs with the forward primer (3'RACE) used for the 3' RACE unraveled that RUS exists in at least two isoforms. Using the reverse primer annealing to the start of exon 4 (Ex4-rv), I obtained one high abundant product at 225 bp, arguing that exon 4 is included (Figure 24C). With the primer annealing to the beginning of exon 5 (Ex5-rv), for the RUS isoform missing exon 4, I expected a PCR product of 226 bp in length. Intriguingly, I obtained an additional PCR product at 671 bp relating to the full-length RUS (Figure 24C). This data confirmed that exon 4 is either included or not. Since the isoform - containing all exons - is very low expressed, I argue that RUS must exist in three isoforms: isoform 1 – missing exon4 -, isoform 2 -missing exon 5 -, and the low abundant isoform 3 – containing all annotated exons.
Next, I was interested in whether RUS is alternatively spliced during differentiation in our ex vivo model. To do so, I quantified the level of the isoforms 2 & 3 containing exon 4 and the isoform 1 containing exon 5 but not exon 4 by quantitative RT-PCR, with the 3'RACE primer (3'RACE) as the forward primer and the reverse primers that are annealing to the start of exon 4 (Ex4-rv) and 5 (Ex5-rv). I observed that the expression of both isoforms is upregulated during differentiation. However, isoform-1 is earlier expressed than isoform 2 3, indicating that exon skipping or isoform switching might occur during differentiation (Figure 24D).

6.2.2 KD phenotype of RUS

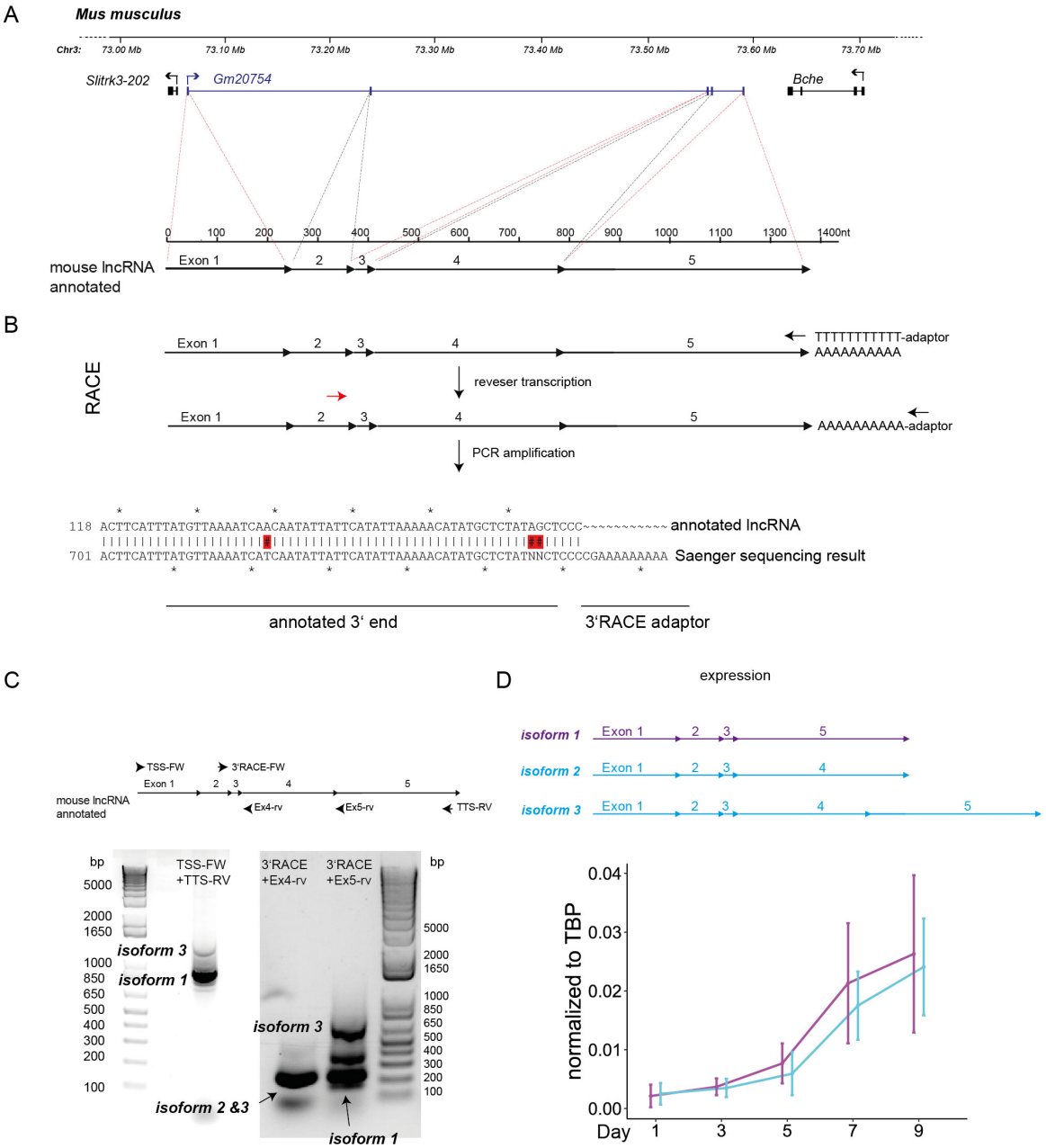


Figure 24: **Genomic organization and expression of murine RUS. A** : Genomic locus around RUS: RUS is annotated as Gm20754 harboring 5 exons. **B:** Scheme and result of the 3'RACE. **C:** Analytical PCR analysis of alternative splicing events of RUS with primers binding to the transcription start site (TSS-FW), transcription termination site (TTS-RV), exon 4 (Ex4-rv), and -5(Ex5-rv). The experiment was performed in biological triplicates. **D:** quantitative RT-PCR analysis results show the expression of RUS isoforms containing exon 4 (isoform 2 & 3) and the isoform 1 containing exon 5 during the differentiation of NSC. Error bars represent the standard error of the mean (SEM).

Knockdown of Slitrk3 did not mimic KD phenotype of RUS in NSC

LncRNAs often regulate the gene-expression in *cis* by affecting only the expression of its neighboring genomic gene. I systematically unraveled whether RUS acts in *cis* by measuring the level of the neighboring protein-coding gene Slitrk3 after control and RUS KD. Besides, I knocked down Slitrk3 as an additional control. To knock-down Slitrk3, we generated four different shRNAs and selected the two most efficient ones.
One day after seeding, NSC were transduced with puromycin resistance gene encoding KD virus particles. KD NSCs were selected with Puromycin 2 days later and differentiated in the absence of basic FGF 4 days later. After two days of differentiation, I isolated the RNA and measured TBP, Slitrk3, and RUS levels. Expression levels were normalized against TBP and compared against the expression in the ctrl KD. The experiment was performed in technical triplicates.
As previous observed (Figure 21), both shRNAs targeting RUS reducing RUS's expression by 46% (Figure 25A). Both shRNAs also reduce Slitrk3 RNA's level by 42% and 41% (Figure 25A). The hairpins targeting Slitrk3 reduce more efficiently the level of Slitrk3 RNA by 52%. (Figure 25B).

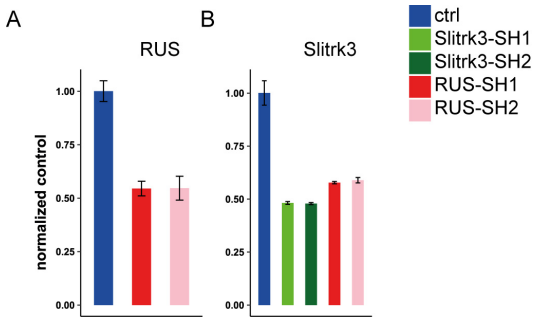


Figure 25: **RUS shRNAs reduce Slitrk3's RNA and RUS levels.** Quantitative RT-PCR analysis of RUS **(A)** and Slitrk3 **(B)** in RUS KD and Slitrk3 KD cells. Experiments were performed in technical triplicates. Error bars represent standard deviation.

Next, I performed KD experiments with GFP encoding KD virus as described before (Figure 23) to see if Slirk3's KD would have the same effect as RUS's KD had on neuronal differentiation. After two days of differentiation, I immuno-stained virus transduced NSC against β-tubulin-III (Figure 26). Experiments were performed in technical triplicates, and therefore, a two-tailed students test was applied. As expected, the knockdown of RUS with SH1 and SH2 significantly reduced the fraction of green-fluorescent neurons by 80% and 71%, respectively (Figure 26). However, Slitrk3 KD using SH1 and SH2 directed against Slitrk3's RNA reduced the fraction of green-fluorescent neurons only by 35% and 20% (Figure 26).

These data show that RUS KD interferes with the expression with Slitrk3. However,

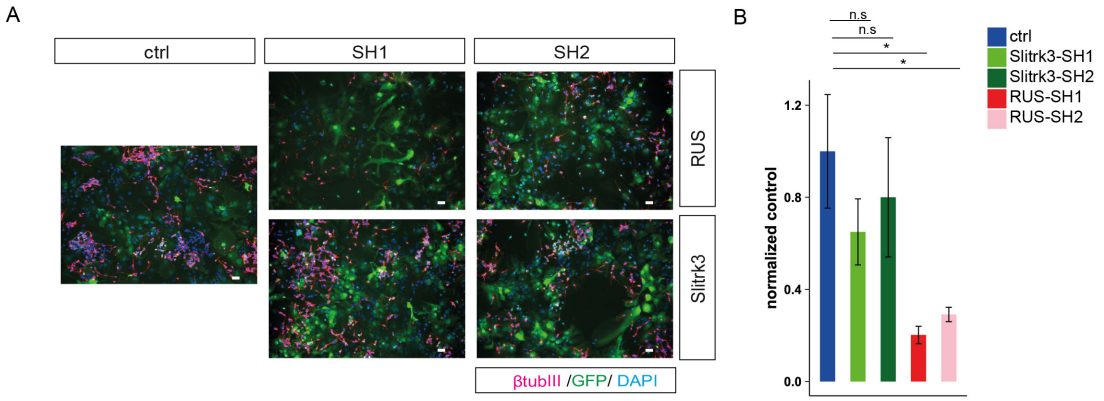


Figure 26: **Knockdown of Slitrk3 did not mimic KD phenotype of RUS in NSC.** KD phenotype of RUS and Slitrk3 in differentiating NSC. **A:** Immunostaining against β-tubulin III (β-tubIII) in RUS and Slitrk3 KD NSC with two efficient hairpins. **B:** Quantification of microscopy images of RUS KD and Slitrk3 KD cells. White bars = 25 μm. Cell nuclei were stained with DAPI. Experiments were performed in technical triplicates. Error bars represent standard deviation. Two-tailed t-test: *: $p < 0.05$, ***: $p < 0.01$.

as knocking down Slitrk3 RNA was not as effective as knocking down RUS on neuronal differentiation, I assumed that RUS acts primarily in *trans*.

Knockdown of RUS significantly decreased fraction of β-tubulin III and Mapt positive cells and proliferation after neuronal differentiation

To evaluate the phenotype in NSC in more detail, I used viruses harboring the puromycin resistance gene. Puromycin selection allowed us on the one side to select for KD cells and on the other side to immune-stain against two neural markers: β-tubulin III and Mapt in parallel. In a parallel experiment, I isolated RNA for quantitative RT-PCR analysis. Again, I added the KD virus encoding for the ctrl shRNA, RUS SH1, and SH2 one day after seeding. Two days later, I conditioned basic-FGF containing medium with Puromycin. Two days after selection, I withdrew FGF and changed to Neurobasal medium. Immunostaining and RNA isolation was performed after the following two days of differentiation (day 7). In addition to day 7, RNA isolation was performed before differentiation (day 5). For RNA isolation, I set up technical quadruplicates. Immunostaining experiments were performed with technical triplicates. As a control for KD-efficiency, I measured the RUS and TBP level by quantitative-RT PCR, normalized against TBP, and compared the relative abundance against the ctrl KD. Immunostaining and quantitative RT-PCR were statistically evaluated by a two-tailed student's t-test. SH1 and SH2 significantly reduced the level of RUS by 41% and 26% on day 5 (Figure 29A), and 58% and 43% on day 7, respectively (Figure 27B, Figure 30A).

These results fit the previous observation that SH1-RNA more reliable reduced the level of RUS than SH2. The reason for that might be that SH1 targets exon 5 included in the high abundant RUS isoform 1 and isoform3. SH2 binds exon 4 included in isoform 2 and 3. According to quantitative RT-PCR analysis, the fraction of Mapt and β-tubulin III positive cells in immunostained cells are reduced in both KD conditions. I observed Mapt staining's background that I avoid during the quantification of microscopy images by increasing the respective channel's threshold. I found 24.6% Mapt positive cells in the ctrl KD (Figure 27A & C). In correlation to KD efficiency, SH1 reduced the fraction of Mapt positive cells to 11.5% and SH2 to 17.2%. 24.2% of the ctrl KD cells, 8.8% of the SH1-KD cells, and 13.4% of the SH2 KD cells were β-tubulin III positive ((Figure 27A& D). Respectively, the fraction of Mapt and β-tubulin III positive cells dropped from 18.7% in the ctrl KD to 7.6% and 11.6% in SH1 and SH2 condition (Figure 27A & E).

Besides, I performed BrdU experiments to label proliferating cells during the final day of the two-day Puromycin selection. 24h after starting the BrdU pulse, I fixed the cells without chase and labeled them for BrdU incorporation. Experiments were done as technical triplicates and statistically evaluated by a two-tailed student's t-test. BrdU incorporation in ctrl KD cells was less than expected (2.4% BrdU positive cells in the ctrl KD; Figure 27A & F), maybe as a result of low BrdU absorption of NSC. Nevertheless, I observed a significant reduction in both RUS KD conditions (SH1: 0.15% and SH2: 0.35% BrdU positive; Figure 27A & F).

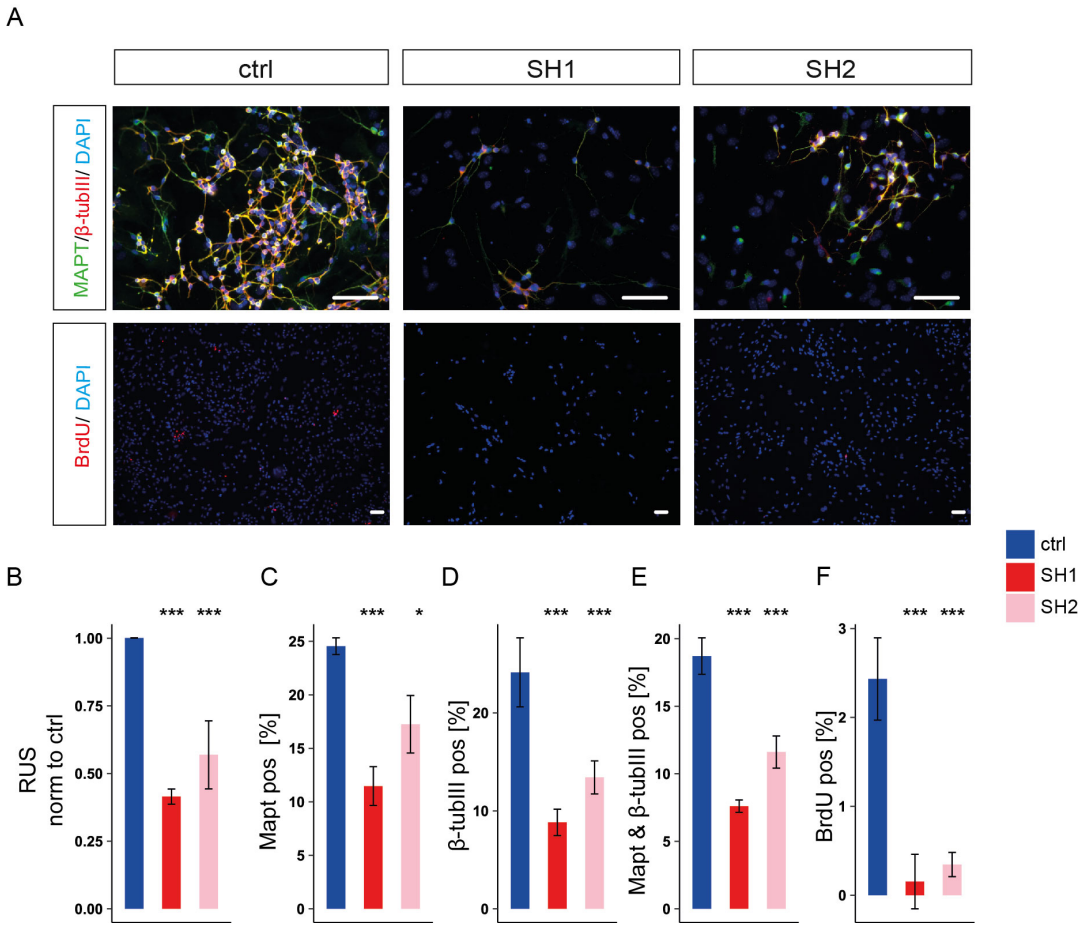


Figure 27: **Knockdown of RUS significantly decreased the fraction of ß-tubIII and Mapt positive cells and proliferation after neuronal differentiation. A:** Immunostaining against β-tubulin III (β-tubIII) and Mapt in RUS KD NSC on day 7. Immunostaining against BrdU in BrdU chased RUS KD NSC on day 5. White bars = 25 μm. **B:** Quantitive RT-PCR analysis of RUS in RUS KD cells on day 7. **C-F:** Quantification of microscopy images for Mapt positive cells **(C)**, β-tubulin III positive cells **(D)**, Mapt and β-tubulin III positive cells **(E)**, and BrdU positive cells **(F)**. Experiments were performed in technical quadruplicates. Error bars represent standard deviation. Two-tailed t-test: *: $p < 0.05$, ***: $p < 0.01$.

KD of RUS induced apoptosis in cortical neuron cell culture

Besides our neural stem cell model, I analyzed RUS KD's effect in cortical neuron culture. I transduced KD-virus 6 h after seeding. Since cortical neurons are not dividing, I used KD virus encoding for GFP as a transduction marker. By performing the experiments, I noticed that cortical neurons did not tolerate well the KD of RUS. Thus, I terminated cortical neuron culture experiments typically on day 4. By quantitative RT-PCR, I measured the level of RUS, Slitrk3, the neural stem cell and Glia marker GFAP, different neuroectodermal like Nestin, Sox2, and Pax6, and neuronal markers: β-tubulin III, Mapt, Map2, and Grin1 (Figure 28A). Expression values were normalized to the expression values of the housekeeper gene TBP. Normalized values in the RUS SH1 and

SH2 KD-condition were compared with the ctrl KDs' normalized values. Experiments were performed in technical triplicates and statistically evaluated by a two-tailed t-test. Like in differentiated NSC, SH1 and SH2 significantly reduced RUS levels by 51% and 42% in cortical neurons, respectively. The level of the neighboring gene Slitrk3 was not significantly affected by the KD experiments. SH1 reduced the neuroectodermal marker Nestin level by 45% and the stem cell marker Sox2 by 45%. SH2 reduced both transcripts by 83% and 69%, respectively. Although the neuronal markers' expression: β-tubulinIII, Map2, and Grin1 were not affected by KD of RUS, I observed a significant decrease in Mapt expression. SH1 reduced the level of Mapt by 42% and SH2 by 36%. Additionally, GFAP expression was decreased by 45% in the SH1- and 46% in the SH2 condition.

I confirmed the decrease in Mapt expression after KD of RUS by immune-staining against Mapt in 4 days *in vitro* cultured neurons (Figure 28B & C). Immunostaining experiments were again performed as technical triplicates. For statistical analysis, I applied a two-tailed student's t-test. 78% of ctrl KD cells were Mapt positive. The KD of RUS with SH1 and SH2 significantly reduced the fraction of Mapt positive cells to 39% and 43%. I also observed that after KD of RUS, Mapt positive cells show fewer branched axons than ctrl neurons (Figure 28B inlays). According to the expression data, staining against Nestin revealed a tremendous drop of 23.8% positive Nestin cells in the control condition to 4% and 10% after SH1 and SH2 mediated RUS KD, respectively (Figure 28B & D). Immunostaining against GFAP yielded no significant reduction between ctrl and RUS KD samples (Figure 28B & E). Because the KD of RUS is not well tolerated by the cortical neurons and differentiating NSC, I asked whether RUS KD cells undergo increased apoptosis. To address this question, I immunostained cortical neurons after RUS KD against the cleaved fragment of Caspase 3 (Figure 28B & F). Caspase 3 plays a central role in apoptosis. Cleavage of Caspase 3 activates the enzymatic activity, and thus cleaved Caspase 3 is the most prominent hallmark of the intrinsic and extrinsic apoptotic pathways in the cells [162]. Immunostaining against cleaved Caspase 3 unraveled a significant increase of apoptotic cells from 2% in the ctrl KD to 12% and 15 % in the SH1 and SH2 condition, respectively.

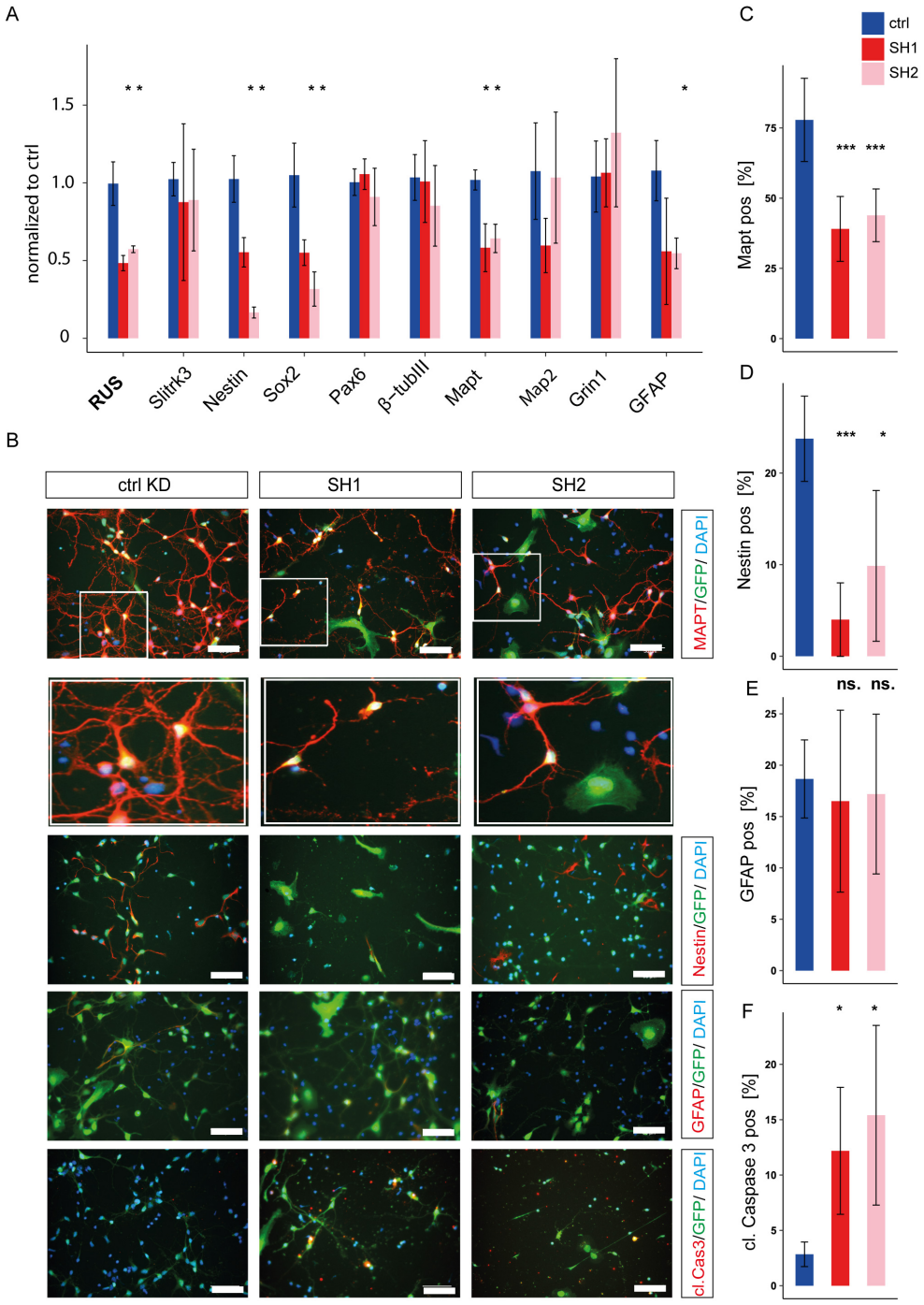


Figure 28: **KD of RUS induced apoptosis in cortical neuron cell culture. A:** Quantitive RT-PCR analysis of RUS KD cortical neurons of RUS, Slitrk3, and neuronal lineage marker: Nestin, Sox2, Pax6, β-tubulin III, Mapt, Map2, Grin1, and GFAP. **B:** Immunostaining against Mapt, Nestin, GFAP, and cleaved Caspase 3 in RUS KD cortical neurons. The white inlays are magnified in the lower white box. **C-F:** Quantification of microscopy images for Mapt positive cells **(C)**, Nestin positive cells **(D)**, GFAP positive cells (**E**), and cl.Caspase 3 positive cells **(F)**. White bars = 25 μm. Cell nuclei were stained with DAPI. Experiments were performed in technical triplicates or quadruplicates. Error bars represent standard deviation. Two-tailed t-test: *: $p < 0.05$, ***: $p < 0.01$.

Altogether, these data suggest that RUS interferes with the proliferative and differentiating behavior of NSC. Isolated neural progenitor cells are not proliferating under neuron culture conditions. However, they divide symmetrically or asymmetrically into neurons. In this manner, cell division produces two neurons or one neuron and one further matured neural progenitor cell [9]. Nestin and Sox2 are markers for neuroectodermal markers and are expressed in neuroepithelial progenitors. GFAP is a marker for radial glia progenitors. All three progenitor markers were reduced after RUS's KD. In-line, Mapt, a marker expressed in mitotic and postmitotic neurons, was strongly reduced. These suggest that RUS's KD interfered with the dividing behavior required for neuron differentiation and locked the progenitor cells in their stage. This locked stage potentially favors the apoptotic program as observed by cleaved Caspase 3 staining. Increased apoptosis would have lead to the ultimate loss of the progenitor cells. However, a loss of progenitor cells should have also been manifested by a loss of neurons. β-tubulin III is also a marker for mitotic neurons but is predominantly expressed in postmitotic cells. Map2 is also a postmitotic marker. Both markers' expression was not affected by RUS's KD, suggesting that RUS not affected neuron maturation. However, cells were analyzed four days after seeding, at a time point, when neurons were not fully matured and express both markers at a low level. Moreover, β-tubulin III positive cells were reduced by RUS's KD in our neural stem cell model.

RNA-sequencing analysis after KD of RUS in differentiated NSC

To explain the KD phenotype observed by staining, I performed RNA Sequencing (RNA-Seq) experiments of RUS KD cells. Because cortical neurons did not tolerate the KD of RUS, I experimented with *in vitro* differentiated NSC. I used RNA samples collected from four ctrl-, SH1-, and SH2-shRNA treated cells each on day 5 (Figure 29) and day 7 (Figure 30). As controlled by quantitative RT-PCR, SH1 and SH2 KD samples showed reduced RUS expression (Figure 29A Figure 30A). The EMBL-Gene Core facility carried out the preparation of sequencing libraries and sequencing on a NextGen 500. Approximately 30 Mio Sequencing reads per replicate were aligned against mm10 mouse reference genome using the STAR aligner. Transcripts per Million (TPMs) were calculated using the RSEM software and statistically evaluated using the R-package DESeq2 by applying the Wald-test [142, 148]. Therefore, I pooled both KD conditions as one sample and compared them to the ctrl KD.

On day 5, 34,029 expressed genes were identified. 3,207 (9.4%) genes were downregulated, and 2,603 genes were upregulated (7.6%) by KD. Clustering of z-scores of all deregulated genes using the euclidean distance method (Figure 29B) clustered -as expected- all ctrl replicates together. I observed that replicating 1 of SH2 does not gather to other knockdown samples and considered this sample an outlier (Figure 29B & C: SH2-1). Replicate 2-4 of SH2 formed a cluster, and all replicates of SH1 are clus-

tering together. Next, I performed gene-ontology enrichment analysis to see whether deregulated genes are involved in the same signaling pathway or the same biological process using the online tool: http://geneontology.org// [163]. I performed the analysis with significantly downregulated (Figure 29C) and upregulated (Figure 29D) genes after RUS's KD separately. From deregulated genes associating with the same biological process, I selected 30 representative genes together with RUS itself and clustered their z-scores using the Euclidean distance method (Figure 29E). Thereby, all ctrl-KD samples and RUS KD samples clustered together, indicating that RUS's KD affected those genes' expression similarly. Genes encoding Wnt-proteins- Dlx1 and Hes1-, and other genes required for neuron formation are downregulated (Figure 29E).
Furthermore, genes inhibiting neuron differentiation and keeping neural stem cells dividing like Notch 1, Notch3, Fgfr3, and the cell cycle genes Cdc25c, Ccna2, and E2F7 were also downregulated. Coinciding to downregulated cell cycle genes, genes belonging to RNA metabolisms as the RNA splicing factors Srsf 11 and Srsf 12 were also less abundant in the KD-conditions.
Genes as Caspase 4 and 6 and Bcl proteins that are part of the apoptotic pathway were upregulated after RUS KD. Intriguingly, RUS annotated as Gm20754 was only subtle downregulated. Subtle changes can be a consequence of sequencing depth and the low expression of RUS on day 5.

I performed the same analysis of samples collected on day 5 for samples that I harvested on day 7. On day 7, a reduced RUS level after KD was more apparent, both by qPCR and RNA-sequencing (Figure 30A & E). From 35.724 identified genes, 3,617 genes (10%) were upregulated, and 2,656 (7.4%) genes were downregulated. The clustering of z-scores yielded distinct clusters (Figure 30B). Each cluster contains all 4 four replicates of the same condition. Besides, the clusters of the two different SH RNAs directed both against RUS are grouping.
Gene ontology enrichment analysis again revealed that genes involved in the cell cycle and RNA metabolisms were downregulated after RUS KD (Figure 30C & E). More precisely, the cell cycle genes Cdc25c, CCna2, Cenpf, Myc and genes encoding for members of the exosome complex such as Exosc7 and Exosc11, as well as ribosomal proteins such as Rsp7 and Rsp 15 were reduced. Genes involved in neurogenesis such as Rest and Notch1, and, again, genes of the apoptotic pathway as Caspase 6 and Caspase 9, were upregulated on day 7 (Figure 30D & E).

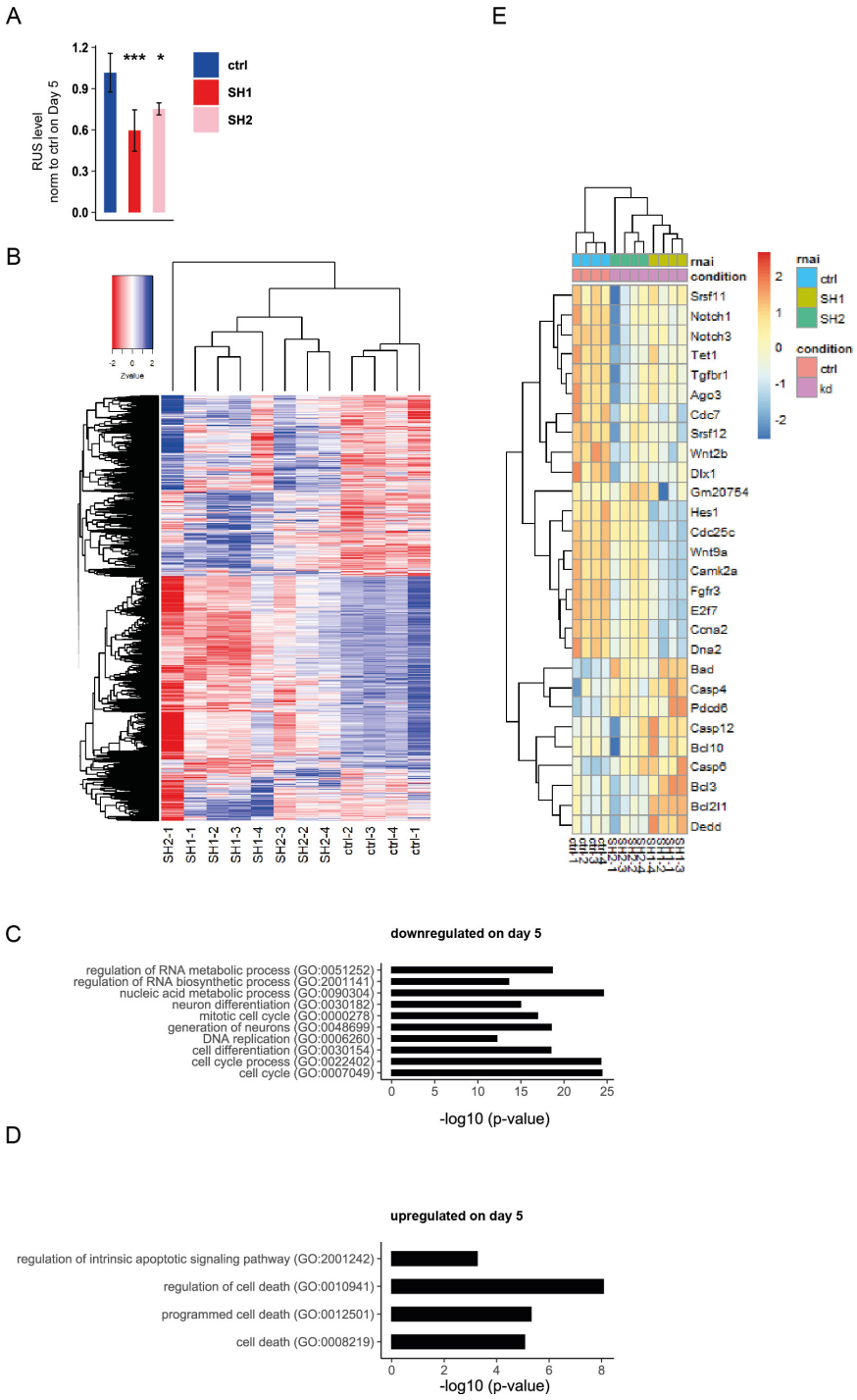


Figure 29: **RNA-Seq analysis of NSC after KD of RUS collected on day 5**. The experiment was performed with technical quadruplicates **A:** Quantitative RT-PCR analysis of RUS. Error bars represent standard deviation. Two-tailed t-test: *: $p < 0.05$, *** : $p < 0.01$. **B:** Euclidean distance method: clustering of z-scores of all significantly deregulated genes. **C:** Gene ontology enrichment analysis of downregulated genes represented by -log10 values of the adjusted p-values. **D:** Gene ontology enrichment analysis of upregulated genes represented by -log10 values of the adjusted p-values. **E:** Euclidean distance method: grouping of the z-scores of 30 representatives significantly deregulated genes together with RUS (Gm20754).

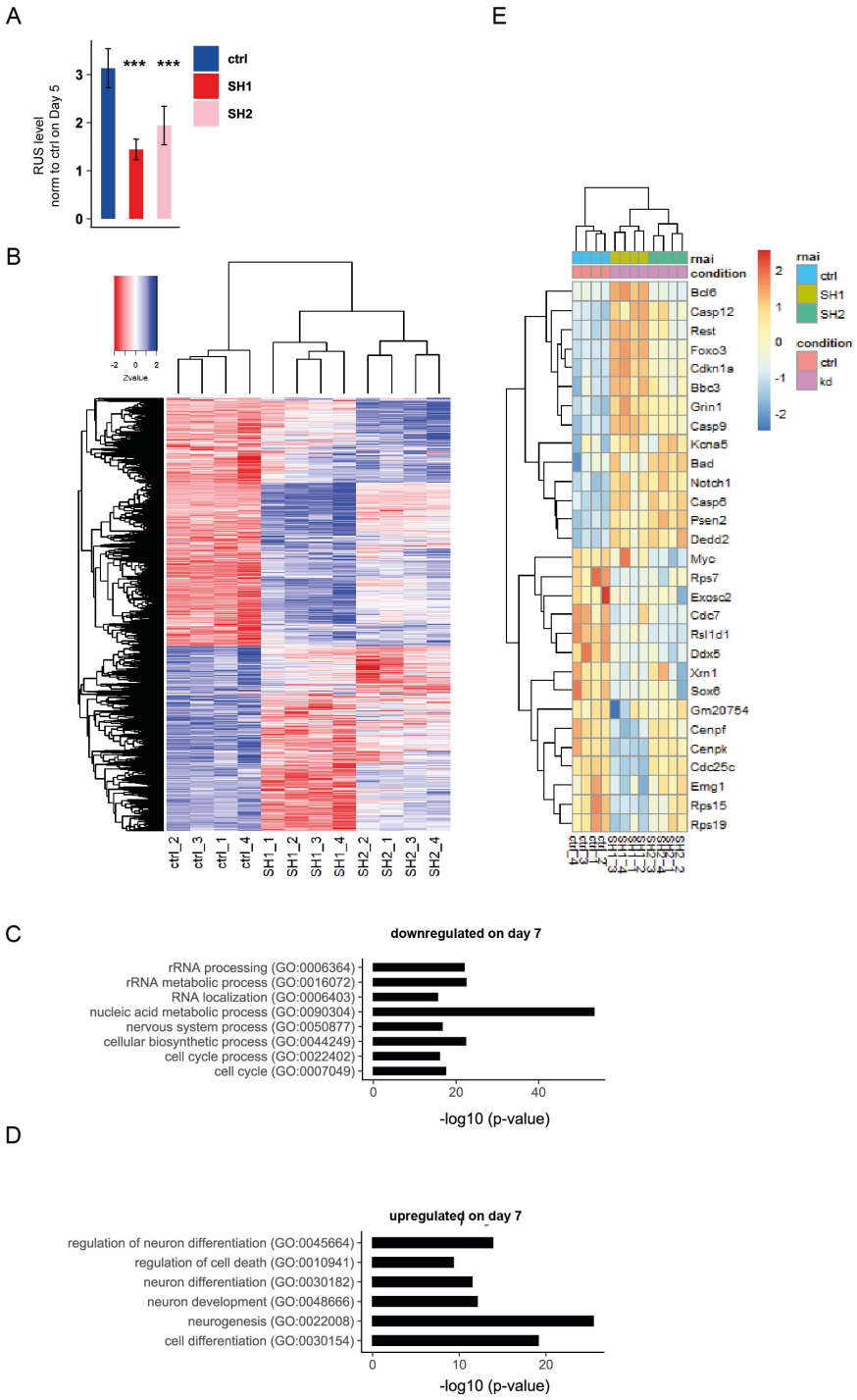


Figure 30: **RNA-Seq analysis of NSC after KD of RUS collected on day 7.** The experiment was performed with technical quadruplicates **A:** quantitative RT-PCR analysis of RUS compared to expression in ctrl KD on day 5. Error bars represent standard deviation. Two-tailed t-test: *** : $p < 0.01$. **B:** Euclidean distance method: clustering of z-scores of all deregulated genes. **C:** Gene ontology enrichment analysis of downregulated genes represented by -log10 values of the adjusted p-values. **D:** Gene ontology enrichment analysis of upregulated genes represented by -log10 values of the adjusted p-values **E:** Euclidean distance method: grouping of the z-scores of 30 representatives significantly deregulated genes together with RUS (Gm20754).

More careful consideration of the expression values of the neuronal markers β-tubulin III (Figure 31A), Mapt(Figure 31B), and Map2 (Figure 31C) revealed β-tubulin III and Mapt were not effectively reduced on the RNA-level by both KD conditions. On day 5, SH1 reduced significantly β-tubulin III expression and SH2 Mapt expression. On day 7, β-tublulin expression increased in the ctrl and RUS KD conditions to an equal level. During the time-course of differentiation, Mapt expression remained unaffected in the control condition. However, depletion of RUS with SH1 increased, and SH2 reduced Mapt expression on day 7 significantly. Only Map2, a late expressed neuronal marker (6.1.2), was significantly reduced by both RUS KD's conditions on day 5. During the time-course of differentiation, Map2 expression increased in all conditions but remained significantly downregulated by both RUS KD conditions on day 7.
Controversial, the number of β-tubulin III and Mapt positive cells was significantly reduced by both RUS KD conditions (6.2.2). Thus, we performed western-blot analyses against β-tubulin III (Figure 31D) and Mapt (Figure 31E) in ctrl and RUS KD cells (SH1 & SH2) NSC on day 7. Besides, we performed western blot analysis against β-actin to verify similar protein-loading. Both RUS KD conditions reduced the protein level of β-tubulin III (by 29-36%) and Mapt (by 26-27%) and confirmed our data obtained by microscopy analysis.

In summary, the RNA-Seq experiments confirmed our initial observation that the RUS KD reduced cell proliferation as manifested by the downregulation of genes belonging to the cell cycle and genes belonging to RNA metabolism. Expression of genes involved in the neuronal fate commitment was first reduced on day 5 and increased later on day 7. Intriguingly, those genes as Notch1 and Rest are required for the neuronal cell fate commitment but block the final differentiation. Both genes are more abundant in progenitor than in neurons. During differentiation, the expression of both genes drops. Upregulation of those markers at a later stage agrees with our previous suggestion that RUS KD interferes with the further differentiation behavior and locks the KD-cells in their differentiation stage. RUS KD differently changed the neuronal markers' RNA level β -tubulin III and Mapt. Only Map2 RNA-level was affected equally by both RUS KD conditions. However, the protein level of β -tubulin and Mapt was strongly reduced by both RUS KD's conditions. These observations might be a consequence of reduced ribosomal protein levels produced by RUS KD. The locked stage's consequence is that cells undergo more apoptosis, in agreement with our observation of upregulated apoptosis as seen in cleaved Caspase 3 stainings (6.2.2, Figure 28) and RNA-seq analysis.

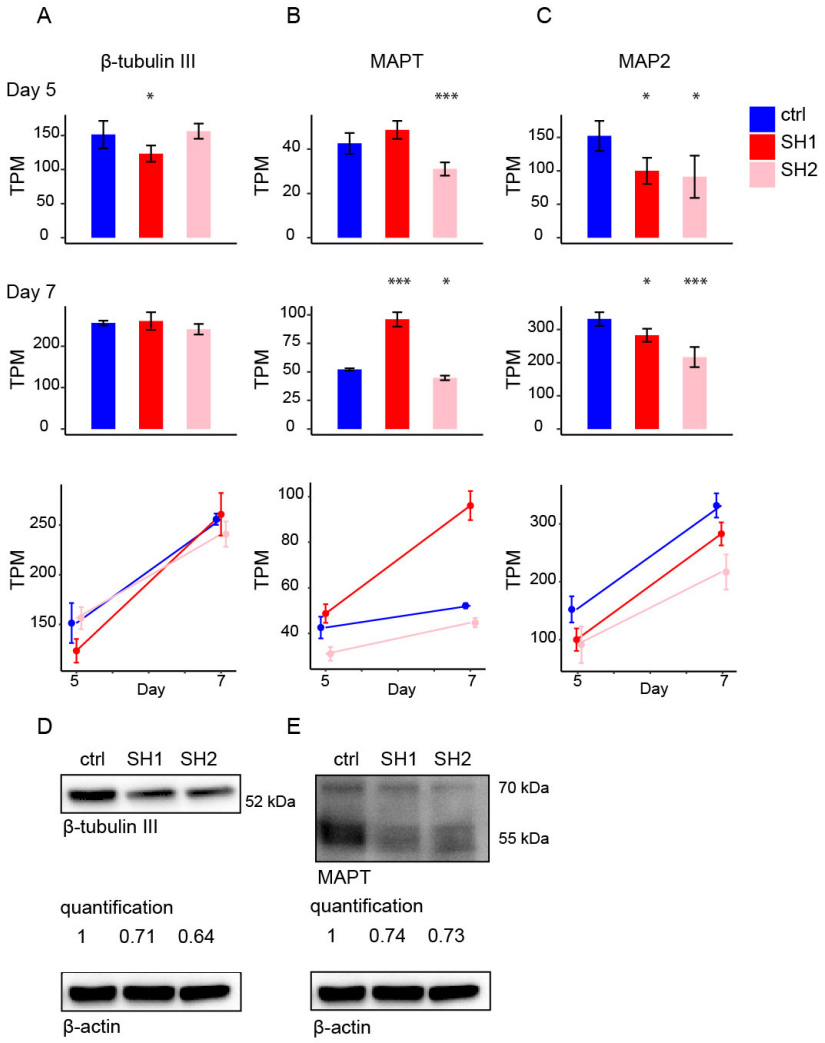


Figure 31: **Expression of neuronal marker genes in RUS KD cells. A-C:** RNA Seq values (TPM) of the neuronal marker genes β-tubulin III **(A)**, Mapt **(B)**, and Map2 **(C)** in ctrl and RUS KD (SH1 SH2) NSC collected on day 5 and 7. The experiment was performed with technical quadruplicates. Two-tailed t-test: *: $p < 0.05$, *** : $p < 0.01$. Western blot analysis and quantification against β-tubulin III **(D)**, Mapt **(E)**, and β-actin **(D+E)** in ctrl and RUS KD (SH1 SH2) NSC collected on day 7.

6.2.3 Rescue Experiments

RUS harbors a remarkable high sequence conserved domain at the 5' end between mouse and human. I was particularly interested in whether this domain is functionally relevant. To address this question, I performed a rescue experiment of the KD-phenotype in differentiating NSC by the overexpression of different RUS constructs. Since isoform 1 is the most abundant isoform of RUS, I decided to overexpress the full-length isoform 1. I overexpressed 5x MS2 aptamers as a negative control of noncoding RNA. Besides, two deletion mutants of RUS were generated. The first deletion construct lacks the conserved 5' domain in exon 1 and 2 called $\triangle$ 5' to test for this domain's functionality. The second deletion construct lacks the non-conserved exon 5 called $\triangle$3' as another control. Since the transfection of plasmids is toxic for neural stem-cells, I decided to overexpress non-coding constructs delivered by lentiviruses. However, all available lentiviral overexpression vectors are inappropriate to overexpress non-coding RNAs. Before I started to overexpress RUS constructs, I developed an appropriate lentiviral vector overexpressing the lncRNA and a selection marker. As I used them for rescue experiments, I named them rescue-vectors.

Design of lentiviral rescue vectors for overexpression of lncRNA constructs

Non-replicating lentiviruses are versatile carriers to deliver genetic material into host cells. Once delivered, the genetic information encoded as RNA gets integrated into the genome by the viral particle's reverse transcriptase and integrase activity [164]. This technology is frequently used to over-express proteins or shRNAs, as described above. The co-expression of selection markers like resistance genes against antibiotics enables to select for transduced cells. All required genetic elements like promoters and the coding sequences must be transcribed into one long RNA molecule in virus-producing cells. Genetic factors such as polyadenylation signals in the sense direction that stop the transcription must be avoided. These signals interfere with virus production. Frequently, the gene of interest and selection marker expressed from different promoters are encoded on the RNA genome in sense direction. Since a polyadenylation signal is missing in between, the transcription from the upstream promoter results in a chimeric transcript. This system is adequate to overexpress proteins since chimeric RNAs are not interfering with protein translation. However, for lncRNAs, the system is not appropriate since chimeric RNAs disrupt the molecular mechanism and cellular localization of lncRNAs. To circumvent this problem, I decided to generate a lentiviral rescue vector overexpressing the lncRNA of interest and a selection marker in a bidirectional manner. Therefore, the lncRNA of interest, including a polyadenylation signal, is cloned in the antisense direction to the selection marker. Adding the polyadenylation signal in antisense direction prevents the preliminary stop of transcription of the RNA genome in virus-producing cells. I tested different promoters and arranged genetic elements like WPRE differently

to evaluate the effect on the host cells' transcription efficiency.

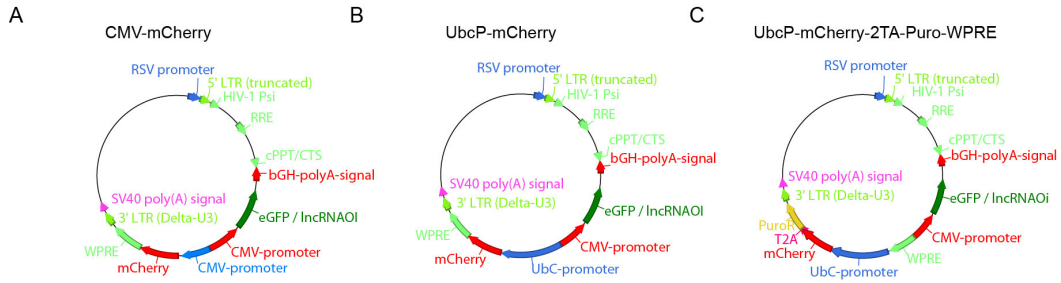


Figure 32: **Maps of generated bidirectional lncRNA over-expression vectors. A:** CMV-mCherry-. **B:** UbcP-mCherry-. **C:** UbcP-mCherry-2TA-Puro-WPRE. The RNA genome is transcribed under the control of the RSV promoter in virus-producing cells. Transcription of the virus genome is terminated by SV40-polyA signal. LTRs at the 5' and 3' end are required to integrate reverse-transcribed RNA into the host genome by viral integrase. The Psi signal is needed for the packing of the RNA genome into viral particles. WPRE, RRE cPPPT/CTS repetitive elements are acting as enhancing elements. lncRNAOI: lncRNA of interest.

For all generated rescue vectors, the lncRNA expression is driven by the strong CMV enhancer-promoter. To control the strength of expression during virus optimization, I replaced the lncRNA with GFP to visualize and document expression's efficiency by green fluorescence. Instead of a selection marker, I used mCherry to visualize the sense strand's expression by red fluorescence during optimization. All tested viruses were produced according to the standard protocol using a molar fraction of 2:1:1 of the bidirectional vector (CMV:mCherry; UbcP:mCherry; UbcP-mCherry-2TA-Puro-WPRE) to both packing plasmids. To test the vector's efficiency, I seeded 50.000 N2A cells on a 12 cm coverslip and treated them 16 h later with 3 μl purified virus. After 48-72 h, I analyzed the number of green and red fluorescent cells.
My first vector expressed both mCherry as well as GFP from the CMV promoter (Figure 32A). Both CMV promoters, one in the antisense direction, are located adjacent to each other. However, the corresponding virus's application resulted only in a small fraction $< 1\%$ of transduced cells.
Next, I replaced the CMV promoter inducing the selection marker's expression from the Ubc Promoter (Figure 32B). The promotor's replacement increased the transduction efficiency to 6.5%, which is still inadequate for lncRNA over-expression.

For a third vector, I shifted the WPRE element (specific DNA sequence from the Woodchuck Hepatitis virus to enhance expression [165]) that is typically located next to 3' end between the Ubc promoter and CMV Promotor (Figure 32C). Additionally, I fused the puromycin resistance gene to mCherry. Thereby, I cloned a 2TA signal between both genes to separate both proteins during protein translation. The application of this fusion protein as a selection marker enables me to visualize the expression of

the selection by red fluorescence on the one hand. On the other side, I can select for transduced cells with Puromycin. This strategy tremendously increased the transduction efficiency to 37%. As a result of these prescreening experiments, I used this vector for the following rescue-experiments. Besides, I further increased the virus amount by 3x fold to obtain full transduction efficiency.

Restoration of β-tubulin III, Mapt, and Map2 positive cells by RUS full- length isoform1 and 3' deletion constructs

Table 26: Expression of RUS in rescue cells

OE (overexpression)	KD	Expression	Mean	STD
5x MS2	ctrl KD	RUS	1.00206762	0.07762898
5x MS2	RUS KD	RUS	0.52997807	0.07315238
full-lengh RUS	ctrl KD	RUS	29.727474	1.63105491
full-lengh RUS	RUS KD	RUS	27.6330485	1.74535342
$\triangle$5' RUS	ctrl KD	RUS	10.3564519	0.79909346
$\triangle$5' RUS	RUS KD	RUS	9.70279782	0.42628779
$\triangle$3' RUS	ctrl KD	RUS	23.5575496	0.62087668
$\triangle$3' RUS	RUS KD	RUS	21.7701478	0.7988117

After I tested lentiviral-overexpression rescue plasmids, I performed rescue experiments of the KD phenotype using KD-viruses harboring the Neomycin resistance gene. That allowed me to select overexpressing KD cells using a medium supplemented with Puromycin and G418. For RUS KD, I used the non-isoform 1 targeting RUS SH2. Similar to other experiments, the KD virus was administered one day after plating. I figured out that the rescue-virus must be added 12 h later to rescue the KD-phenotype efficiently (Figure 33A). The cell density was low at this stage, ensuring that all cells get infected by the lncRNA overexpressing virus. As overexpression constructs to test and control a rescue, I applied the 5xMS2 aptamers as a negative control, full-length isoform1, the $\triangle$5' -, and the $\triangle$3' constructs (Figure 33B). I performed each overexpression construct experiment in ctrl and RUS KD conditions. I selected KD cells overexpressing the rescue vectors three days after plating by switching to NSC medium containing Puromycin and G418 for two days. After selection, I withdraw basic FGF and differentiate cells for two days.

As controlled by quantitative RT-PCR, RUS KD in cells overexpressing the negative control construct (5xMS2aptamers) resulted in a significant reduction in RUS's level of 48% (Figure 33C, Table 26). Overexpression of the RUS's constructs resulted in a 9-30x fold of RUS's level (Table 26).
Next, I analyzed the expression of the neuronal marker β-tubulin III (Figure 33D & G) and Mapt (Figure 33E & H) in RUS overexpressing and 5xMS2 overexpressing KD cells by immunofluorescence. Because RUS's KD did not effectively reduce both neu-

Table 27: number of β-tubulin III, Mapt and Map2 positive cells in rescue experiments

overexpression:	**5xMS2**		**full-length isoform 1**	
KD:	**ctrl**	**SH2**	**ctrl**	**SH2**
staining				
β-tubulin III	28.9(+/- 4.0)%	8.2(+/- 1.8)%	27.9(+/- 6.5)%	24.3(+/- 3.4)%
Mapt	17.5(+/- 6.1)%	2.2(+/- 1.2)%	18.2(+/- 0.3)%	14.1(+/- 3.3)%
Map2	22.7(+/- 12.2)%	7.2(+/- 0.7)%	17.3(+/- 5.9)%	14.6(+/- 4.3)%
overexpression	**△ 5'**		**△ 3'**	
KD:	**ctrl**	**SH2**	**ctrl**	**SH2**
staining				
β-tubulin III	30.4(+/- 7.0)%	10.2(+/- 3.7)%	29.7(+/- 5.7)%	25.3(+/- 5.9)%
Mapt	7.3(+/- 2.7)%	3.5(+/- 1.8)%	17.3(+/- 2.0)%	10.7(+/- 1.5)%
Map2	14.7(+/- 3.8)%	7.1(+/- 2.3)%	25.3(+/- 6.8)%	13.2(+/- 2.4)%

ronal markers in NSC on the RNA-, but on the protein-level, I additionally analyzed the expression of the neuronal marker Map2, which RUS KD significantly reduced on the RNA level (Figure 33F & I). Besides, I removed the mCherry signal during image processing since all cells were red fluorescent during the recording of microscope images. For statistical analysis, I applied a two-tailed student's t-test.

As expected, RUS's KD reduced the fraction of β-tubulin III (Figure 33D & G; Table 27), Mapt (Figure 33E & H; Table 27), and Map2 (Figure 33F & I; Table 27) positive cells significantly in cells overexpressing the negative control (MS2-tag). Intriguingly, the overexpression of the full-length isoform-1 restored the fraction of β-tubulin III, Mapt, and Map2 positive cells after SlitrklncRNA KD. However, the overexpression of the $\triangle$5' construct in RUS KD cells did not restore the fraction of β-tubulin III, Mapt, and Map2 positive cells.

Like the full-length construct, $\triangle$3' construct's overexpression also restored the fraction of β-tubulin III, Mapt, and Map2 positive after RUS's KD.

These data confirm that the neurogenic effect I observed is RUS dependent. Additionally, these data showed that the conserved 5' domain of RUS is indispensable for the biological function of RUS and suggests that the 5' conserved domain may act as an interaction platform for gene-regulatory proteins.

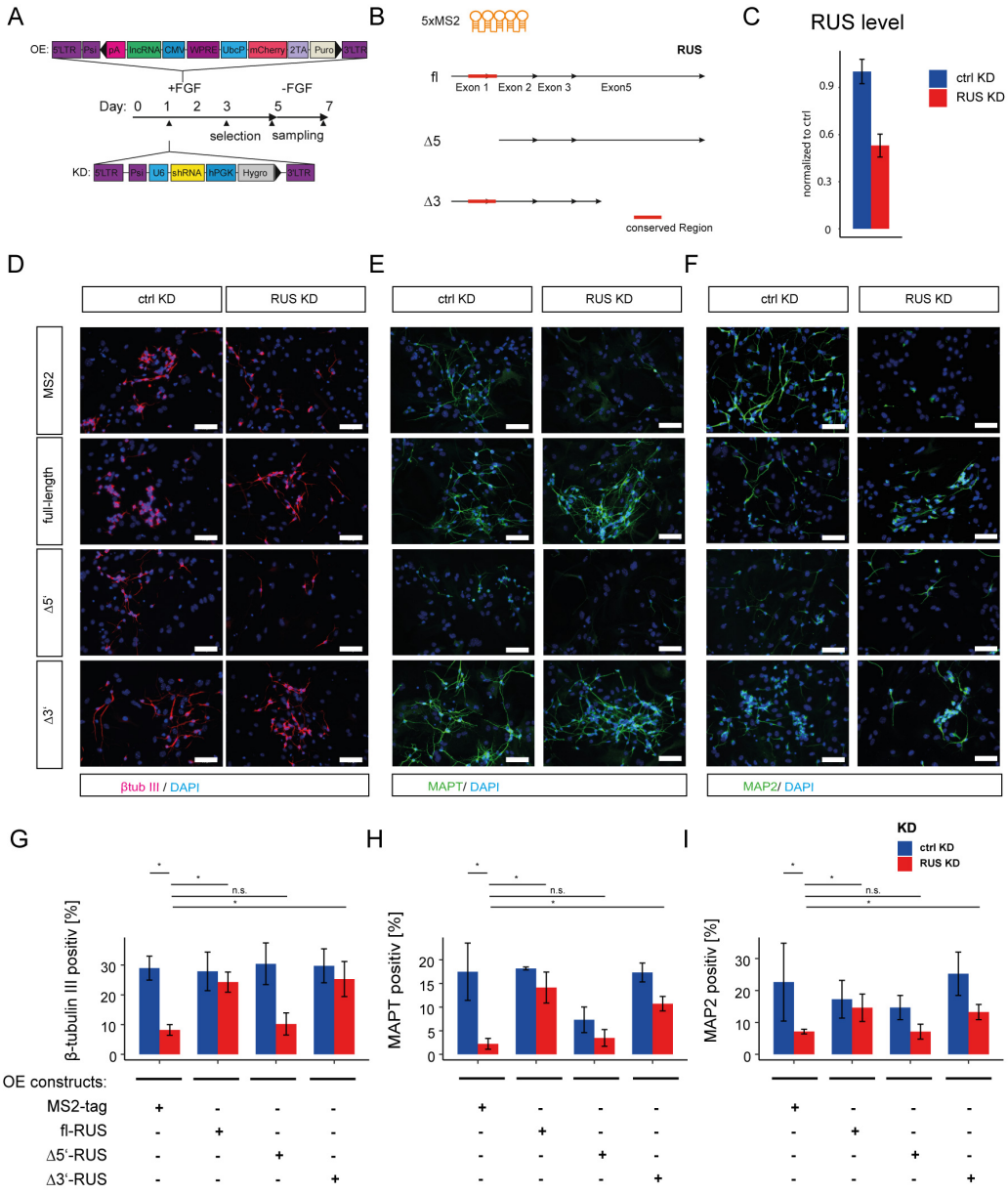


Figure 33: **β-tubulin III, Mapt and Map2 positive cells in rescue experiments. A:** Timeline of experiment & schematic presentation of used constructs. **B:** Overexpression constructs: full-length RUS isoform 1, △5' construct, and △3' construct. **C:** Quantitative RT-PCR analysis of RUS level in KD overexpressing cells of one representative experiment (Error bars: standard deviation of technical triplicates). **D-F:** Immunofluorescence of β-tubulin III (D; magenta), Mapt (E; green), and Map2 (F; green). Cell nuclei were stained with DAPI (blue). White bars = 25 μm. **G-I:** quantification of fluorescence. Experiments were performed in biological quadruplicates. Error bars represent standard deviation. Two-tailed t-test: *: $p < 0.05$, n.s.= not significant. During image processing: mCherry signal was omitted.

6.3 Molecular Mechanism

6.3.1 Nuclear localization of RUS

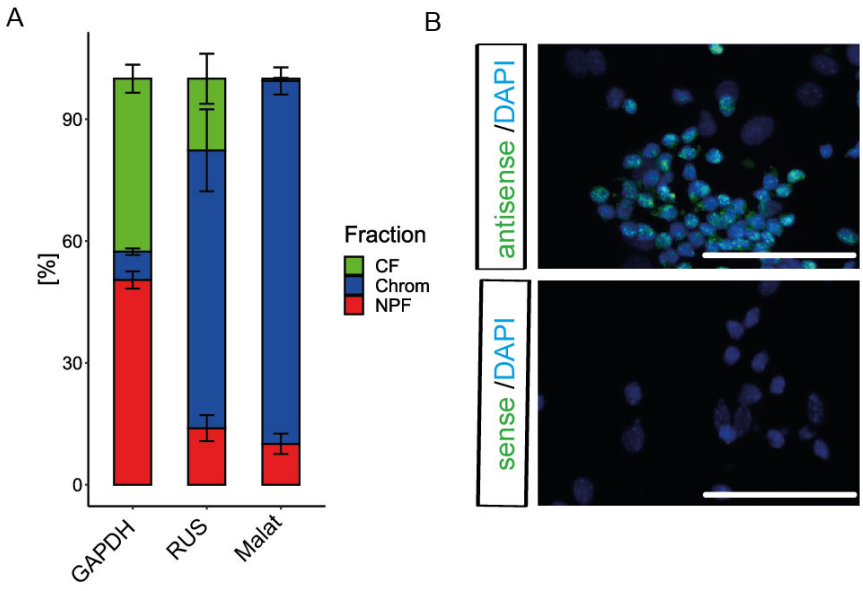


Figure 34: **RUS is predominantly located in the nucleus. A:** quantitative RT-PCR analysis of GAPDH, SlitrklncRNA, and Malat in the cytoplasm (CF), nucleoplasm (NPF), and chromatin (Chrom) of 2-days differentiated NSC. Experiments were performed in biological triplicates. Error bars show standard deviation. **B:** FISH staining in 2-days differentiated neural stem cells with fluorescent-labeled RUS antisense and sense probe. White bars = 25 μm. Cell nuclei were stained with DAPI.

To better understand how RUS is acting on the molecular level, I was interested in the cellular localization of RUS. lncRNAs are located either in the nucleus acting on chromatin or in the cytoplasm regulating post-transcriptional events. To address this question, I performed subcellular fractionation experiments with NSC differentiated for two days. I isolated and transcribed the RNA reversely from the three different isolated fractions: cytoplasm (CF), nucleoplasm (NPF), and chromatin (Chrom). Using quantitative RT-PCR, I measured the level of RUS, GAPDH as a control transcript for lncRNAs residing in the cytoplasm, and Malat as a control transcript lncRNAs residing in the chromatin (Figure 34A). I normalized the corresponding $\triangle$Ct values against the $\triangle$Ct value in the cytosolic fraction to calculate each transcript's normalized enrichment. Next, I summarized the normalized enrichment in each fraction to compute the percentage share as cellular distribution. I experimented with 3 biological replicates. As expected, GAPDH was more located in the nucleoplasm (50.4%) and cytoplasm (42.6%) and nearly absent at the chromatin (7%). Malat was predominantly enriched in the chromatin fraction (89.4%), less abundant in the nucleoplasm (10%), and absent in the cytoplasm (0.6%). Similar to Malat, RUS was highly abundant in the chromatin (68.4%) and less abundant in the cytoplasm (17.7%) and nucleoplasm (13.9%). I con-

firmed the observation that RUS resides in the nucleus by FISH staining in two days differentiated neural stem cells (Figure 34B). I generated fluorescent full-length antisense probe annealing to RUS and a fluorescent sense probe as a negative control. I observed intense staining with the antisense probe. As expected from the subcellular fractionation experiments, the signal was predominant in the nucleus. These observations demonstrate that RUS is located in the nucleus.

6.3.2 A novel *in vivo* RNA affinity purification method revealed the interaction of RUS's 5' region to Brd2, Brd4, Smarca5, and Lbr

Since lncRNAs lack any catalytical activity, it is pivotal to know which gene-regulatory proteins interac with a lncRNA to understand how a lncRNA acts. Several techniques are using *in vitro* transcribed RNA to purify interacting proteins from cell lysates [128]. However, *in vitro* transcribed RNAs do not necessarily resemble RNAs' native structure in the cell due to folding artifacts. Additionally, RNAs are chemically modified during maturation in the cell [166]. Thus, I aimed to establish a novel native affinity purification method to isolate intact lncRNA protein complexes from cells.
Similar to other techniques, our purification strategy using the MS2 aptamer / MS2 binding protein system of 5xMS2 tagged RNAs in cells [128, 129]. As a cellular system, we used the stable murine neuroblastoma cell line Neuro2A cells. After inserting an FRT site upstream to CMV promoter into the Neuro2 cell genome, I flipped-in the pcDN5-FRT plasmid encoding for the 5xMS2 tagged RNA mediated by the flippase enzyme (Figure 35). After flipping, the hygromycin resistance gene lacking a promoter gets expressed from the genomic CMV promoter. Thus, Neuro2A cells integrated the pcDNA5-FRT plasmid can be selected by Hygromycin B conditioning of the medium. Since the integration is happening very rarely, I changed the Hygromycin B supplemented medium several times to remove cell debris. I collected and expanded all grown colonies, typically after 14 days.
To isolate MS2-tagged RNA protein complexes from FlpIN N2A cells, I used the fusion protein consisting of MS2 binding protein and maltose-binding protein (Figure 35A). Therefore, I recombinantly expressed and purified the protein from BL21 E.Coli [128]. The Maltose binding protein allowed for the immobilization of lncRNA-protein complexes on amylose beads. This system was successfully applied to purify interacting proteins of *in vitro* tagged RNA [128]. Using this system to purify 5x MS2 tagged lncRNA protein complexes from cells - as we did – enables us to elute entire lncRNA protein complexes using maltose. Thus our approach differs from other *in vivo* purification strategies [129]. However, the elution mediated by maltose also liberates the high excess of the fusion protein, interfering with the sequencing and quantification process of the interacting proteins by tandem-LC MS analysis. For this purpose, we decided to elute the lncRNA interacting proteins by RNAseA treatment of the beads.

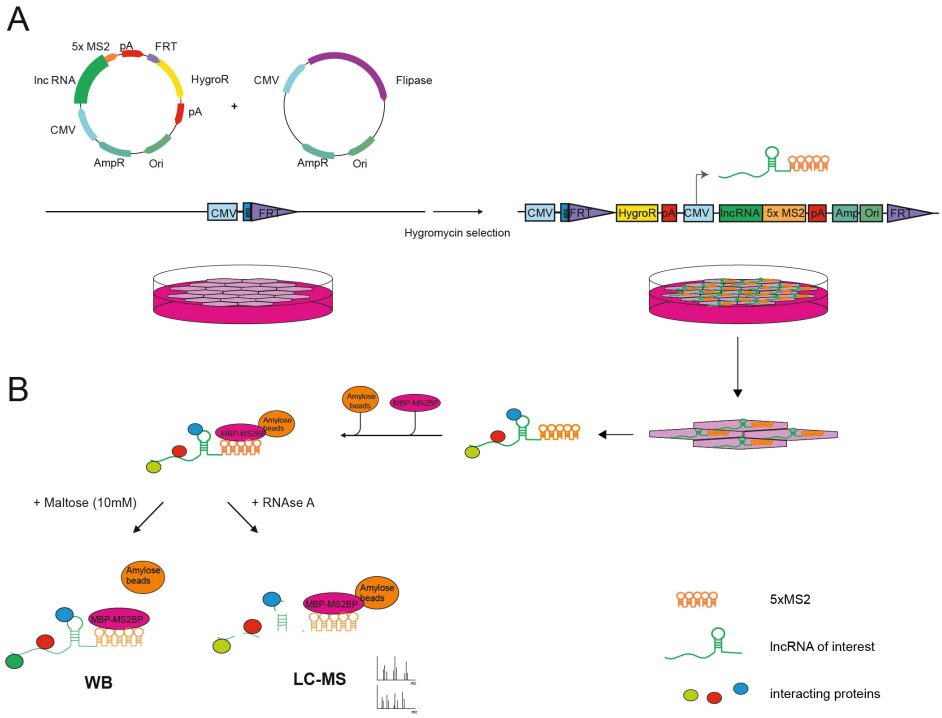


Figure 35: **5x MS2 tagged RNA affinity purification from FlpIn Neuro2A cells. A:** scheme to establish FlpIn cells. **B:** affinity purification strategy. MBP-MS2BP: fusion protein of MS2 binding protein binding and maltose-binding protein. MS2 binding protein is binding to the 5x MS2 tag. Maltose-binding protein allowing for the immobilization of RNA protein complexes on amylose beads. Elution of intact RNA-protein complexes by maltose. Elution of interacting proteins by RNAseA.

To identify relevant interacting proteins, I reasoned to focus on the 5' conserved domain. As observed above in the rescue experiments, I identified this region as functional necessary. To decipher specifically those proteins that interact with this domain, I created three different constructs tagged by 5xMS: The full-length isoform 1 of RUS (nt: 1-912 of RUS; Figure 36), the 5' deletion construct (△5', nt: 233-912 of RUS; Figure 36), and the 5' conserved region itself (nt: 1-232 of RUS; Figure 36). I cloned all tagged versions of RUS into pcDNA.5-FRT vector and generated FlpIn N2As cells. Constructs' overexpression in FlpIn N2As was measured by quantitative RT-PCR and compared with endogenous RUS expression in Wt N2As. Since Wt N2As express RUS at low abundance, all constructs were tremendously overexpressed with a fold enrichment of 4000-6000 (Figure 36).

Because RUS was highly abundant in the nucleus, I isolated the nuclear fraction of full-length isoform 1 and modified RUS overexpressing FlpIn Neuro2A cells before affinity purification.

I set up the experiment in 5 biological replicates per construct. After RNAse A elution, we analyzed the eluted proteins of all three constructs by mass spectrome-

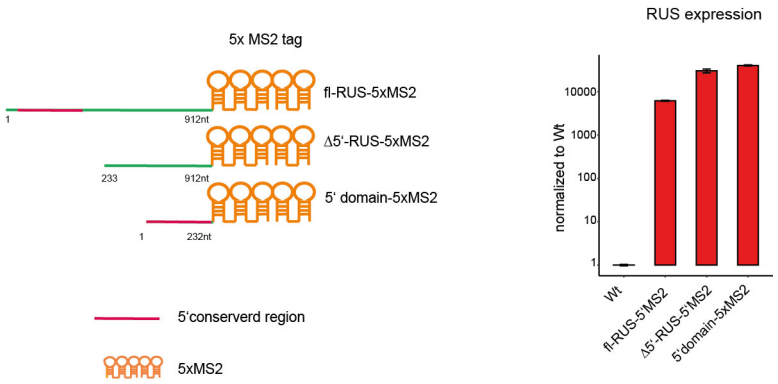


Figure 36: Overexpression of RUS constructs in FlpIN N2As analyzed by quantitative RT. RUS expression values were normalized to TBP and compared against the expression in Wt N2As.

try. First, proteins were digested with Trypsin and LysC. Peptides were alkylated with iodoacetamide and desalted. Sample preparation and the measurement on a Q Exactive orbitrap (Thermo Fischer) were carried out by the – in-house – research group of Prof. Lichtenthaler. We quantified the amount of proteins/ protein groups by the label-free quantification (LFQ) method to maintain the sequencing depth [167] using the mouse genome as the reference genome. We removed contaminants and proteins identified only by post-translational modifications or proteins identified by reverse matching peptides from quantified protein groups. From 3,342 identified and quantified proteins, I selected 1,725 nuclear proteins and continued with 1,028 proteins having an LFQ value greater than 24 in at least one sample. I logarithmized the determined LFQ values greater than zero and imputed zero LFQ values with 0.1 for statistical analysis.

To dissect the proteins bound only at the conserved domain at the 5' end, I statistically selected proteins purified by the full-length construct but not or to a lesser extent by the $\triangle$5' construct. The 5' conserved domain itself should also purify the same proteins. Thus, I also statistically selected protein purified by the 5' domain but not or to a lesser extent by the $\triangle$5' construct and analyzed the overlap between both selected protein sets. Therefore, I calculated the fold-change and the p-value by a student's t-test of each protein between the full-length and $\triangle$5' construct (each protein visualized as one dot in the volcano plot, Figure 37). Forty-six proteins had a fold-change higher than two and a p-value less than 0.05 (red dots, Figure 37, Table 28).

I performed the same analysis between the 5' conserved region and $\triangle$5' construct, which yielded 18 proteins with a fold-change higher than two and a p-value less than 0.05 (red dots, Figure 38, Table 29).

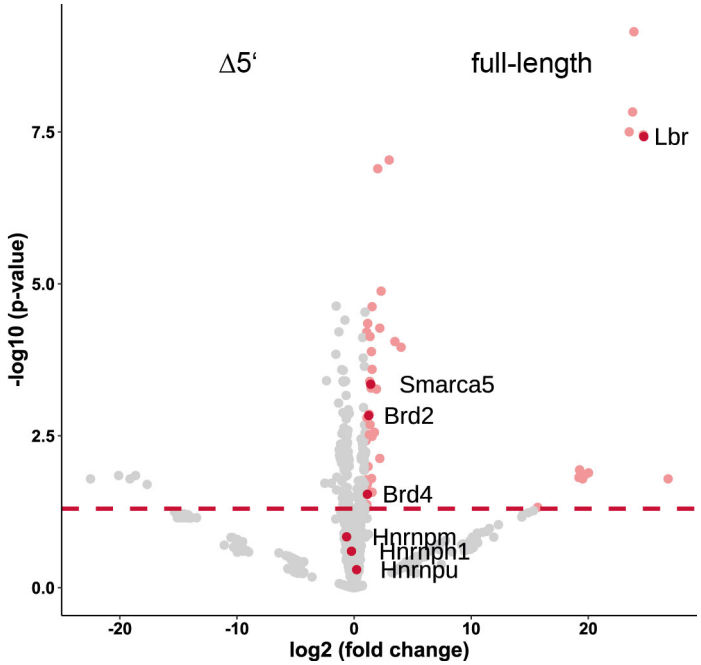


Figure 37: **Deciphering the proteins binding to the 5' conserved domain of RUS by comparing the interactome of the full-length and △5' construct.** The log2 fold-change (x-axis) and -log10 p-value (student's t-test, y-axis) were calculated for each protein (dot) with logarithmized and imputed LFQ values in the full-length and △5' eluates and visualized as a volcano plot. Red dots represent protein with a fold-change > 2 and a p-value < 0.05.

Table 28: Proteins enriched comparing the full-length and △5' construct's interactome with an fold-change > 2 and p-value < 0.05

protein group	description	log_fold_change1	logpvalue
Wdr43	WD repeat-containing protein 43	1.73300427	2.55757368
Wdr12	Ribosome biogenesis protein WDR12	1.52715724	2.4935369
Utp3	Something about silencing protein 10	23.4998825	7.49936712
Utp18	U3 small nucleolar RNA-associated protein 18 homolog	1.09964838	1.36912628
Tor1aip1	Torsin-1A-interacting protein 1	23.7729432	7.82965584
Tmem33	Transmembrane protein 33	20.0166216	1.89173023
Tbl3	Transducin beta-like protein 3	1.28459693	2.85329341
Supt5h	Transcription elongation factor SPT5	1.19655495	1.99550755
Smarca5	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	1.41418927	3.34817875
Rrs1	Ribosome biogenesis regulatory protein homolog	1.52487042	3.5921225
Rrp9	U3 small nucleolar RNA-interacting protein 2	1.33515577	3.39641974
Rcl1	RNA 3-terminal phosphate cyclase-like protein	1.36835167	2.68327022
Rbm19	Probable RNA-binding protein 19	1.47394195	3.88810397
Rangap1	Ran GTPase-activating protein 1	1.11377392	1.68864391
Ppan	Suppressor of SWI4 1 homolog	1.19137745	1.74723283
Polr2f	DNA-directed RNA polymerases I, II, and III subunit RPABC2	19.706409	1.8559407
Polr2a	DNA-directed RNA polymerase II subunit RPB1	1.02696427	2.41831271

Phb2	Prohibitin-2	19.5221531	1.79135766
Phb	Prohibitin	24.6886254	7.45004059
Pdcd11	Protein RRP5 homolog	1.4919023	1.79503368
Nup93	Nuclear pore complex protein Nup93	3.48196635	4.05106174
Nup54	Nuclear pore complex protein Nup54	3.00273802	7.03797519
Nup50	Nuclear pore complex protein Nup50	1.37112528	4.13403571
Nup160	Nuclear pore complex protein Nup160	26.8147836	1.79213858
Nup133	Nuclear pore complex protein Nup133	19.3810955	1.8704386
Nup107	Nuclear pore complex protein Nup107	23.8883235	9.1488687
Noc2l	Nucleolar complex protein 2 homolog	19.2679417	1.9394465
Nat10	N-acetyltransferase 10	1.29518029	2.51802357
Lrpprc	Leucine-rich PPR motif-containing protein, mitochondrial	2.20547877	2.12914653
Lmnb1	Lamin-B1	4.00985597	3.95649578
Lmna	Prelamin-A/C;Lamin-A/C	1.16422961	4.34805241
Lbr	Lamin-B receptor	24.7376995	7.42520393
Hspa9	Stress-70 protein, mitochondrial	2.02670347	6.89739241
Hist1h4a	Histone H4	1.91882215	3.26748875
Fbll1	rRNA/tRNA 2-O-methyltransferase fibrillarin-like protein 1	1.53257921	1.57323417
Emg1	Ribosomal RNA small subunit methyltransferase NEP1	1.06499277	4.20401704
Emd	Emerin	15.6872185	1.32204178
Dld	Dihydrolipoyl dehydrogenase, mitochondrial	2.31365588	4.88050931
Champ1	Chromosome alignment-maintaining phosphoprotein 1	1.45062754	3.28674024
Ccdc86	Coiled-coil domain-containing protein 86	1.53112012	4.62369876
C1qbp	Complement component 1 Q subcomponent-binding protein, mitochondrial	2.20398232	4.26948365
Brix1	Ribosome biogenesis protein BRX1 homolog	1.09033543	2.79762508
Brd4	Bromodomain-containing protein 4	1.13327321	1.53976185
Brd2	Bromodomain-containing protein 2	1.24635405	2.83275391
Bop1	Ribosome biogenesis protein BOP1	1.05987938	1.77526289
Bmi1	Polycomb complex protein BMI-1	19.1883784	1.81578511

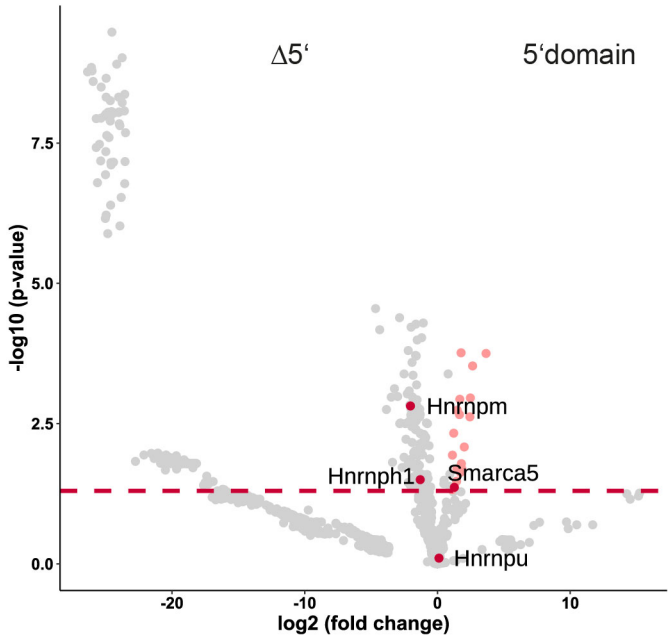


Figure 38: **Deciphering the proteins binding to the 5' conserved domain of RUS by comparing the interactome of the 5'domain and △5' construct.** The log2 fold-change (x-axis) and -log10 p-value (student's t-test, y-axis) were calculated for each protein (dot) with logarithmized and imputed LFQ values in the 5'domain and △5' eluates and visualized as a volcano plot. Red dots represent protein with a fold-change > 2 and a p-value < 0.05.

Table 29: Proteins enriched comparing the 5'domain and △5' construct's interactome with an fold-change > 2 and p-value < 0.05

protein group	description	log_fold_change1	logpvalue
Top2b	DNA topoisomerase 2-beta	1.09571832	1.32429916
Rrp1b	Ribosomal RNA processing protein 1 homolog B	1.13267257	1.93818706
Nat10	N-acetyltransferase 10	1.22883079	2.32960609
Prkaa1	5-AMP-activated protein kinase catalytic subunit alpha-1	1.43486672	1.47533167
Ube2i	SUMO-conjugating enzyme UBC9	1.46668361	1.69281533
Nop56	Nucleolar protein 56	1.48208475	2.73867546
Ddx18	ATP-dependent RNA helicase DDX18	1.64815152	2.66156183
Banf1	Barrier-to-autointegration factor;Barrier-to-autointegration factor, N-terminally processed	1.67917634	2.93228205
Hist1h1c	Histone H1.2	1.73320644	1.62985347
Fbl	rRNA 2-O-methyltransferase fibrillarin	1.77283128	3.76437302
Ubtf	Nucleolar transcription factor 1	1.79888627	1.78634351
Hist1h1a	Histone H1.1	1.85111376	1.68700287
Hmgn2	Non-histone chromosomal protein HMG-17	2.01784039	2.08264932
Hist1h4a	Histone H4	2.44053986	2.62010884
Hist1h1b	Histone H1.5	2.46802032	2.95631424
H1f0	Histone H1.0;Histone H1.0, N-terminally processed	2.64394014	3.52935743
Dld	Dihydrolipoyl dehydrogenase, mitochondrial	3.66559246	3.75224417
Smarca5	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	1.28007836	1.36237385

However, only the four proteins overlapped between the enrichment-set: NAT10 (RNA-cytidine-acyltransferase), Dld (dihydrolipoyl-dehydrogenase), Hist1h4 (Histone 4), and Smarca5 (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5). The interaction of proteins to RNA is defined by the RNA sequence and often depends on the RNA's secondary structure. The small fraction of overlapping proteins may result from disturbed folding between the 5' conserved domain only and the 5' conserved domain in the full-length RUS. Potential disordered folding change may be produced by altering the RNA's maturation by post-transcriptional modifications or by depleted long-range base pairing. Thus, I focused on the four overlapping and the other 42 proteins enriched for the full-length construct. Thereby I focused on candidates with known neurogenic phenotype or transcriptional regulatory activity. Particular attention was drawn to proteins regulating the chromatin structure and epigenome.

The first protein that attracted attention was the ATP chromatin remodeler Smarca5 or Snf2h. Smarca5 was enriched in both comparisons: full-length RUS against $\triangle$5' construct and 5' domain against $\triangle$5' construct by a fold change of 2.6 and 2.4, respectively. Chromatin remodelers reposition nucleosomes around regulatory DNA elements via its DNA-helicase activity at the N-terminus. Thereby, nucleosome-repositioning is either pivotal to activate or prevent progenitor cells' transition to differentiated cells. Avoiding this transition is required to maintain the pool of progenitor cells as described for Smarca5 for neural progenitors [168, 169]. Smarca5 owns similar to Smarca4, a DEAD-box helicase (DEAH domain), which is located between the Helicase ATP binding domain and the Helicase domain. The DEAH domain usually unwinds RNA. Therefore, Smarca5 is binding on the one side to chromatin and on the other side to RNA.

Two other exciting proteins are the Bromodomain and Extra-Terminal motif (BET) proteins Brd2 and Brd4. Both proteins are enriched between full-length RUS and $\triangle$5' construct with a fold change of 2.4x and 2.2x, respectively. Comparing the interactome of 5' domain against $\triangle$5' construct yielded no significant change. BET-proteins binding via their bromodomain at acetylated H4 tails. Both occupying promoter and enhancer regions. Mainly, Brd4 activates transcription by interaction to pTEF-b that switches RNA Pol II in the elongation state [80]. Similar to Smarca5, Brd2 and Brd4 have pivotal functions in neural precursor cells [170, 171].

The last protein that drew my attention is the lamin B-receptor (Lbr). Intriguingly, this protein was enriched only in the full-length RUS interactome. By imputing the zero values with 0.1 in all other samples, we obtained a log-2 fold change value of 24.7. Lbr sits in the nuclear envelope protein and acts as a receptor for the nuclear matrix protein lamin B. Additionally, Lbr possesses a sterol reductase activity. However, the most exciting part of Lbr is the Tudor domain at its N-terminus. Via the Tudor domain, Lbr tethers heterochromatin close to the nuclear envelope to silence the gene-expression. Thereby, the Tudor Domain recognizes heterochromatin via methylated H4K20.

Proteins potentially bind to the 3' end of RUS would have a fold-change lower than 0.5 and a p-value lower than 0.05 in comparing the conserved 5' region and the $\triangle$5'

construct. The selection of those proteins yielded Hnrnpm and Hnrnph1 (Figure 38). Both proteins are binding preferential to polypyrimidine: poly uridine and poly cytosine tracts [172, 173]. Poly-uridine tracts are highly abundant in the 3'region of RUS. Thus, both proteins confirm the robustness and specificity of our affinity purification.

6.3.3 Western Blot analysis revealed the binding of RUS to Brd2 4 depends on their ET domain

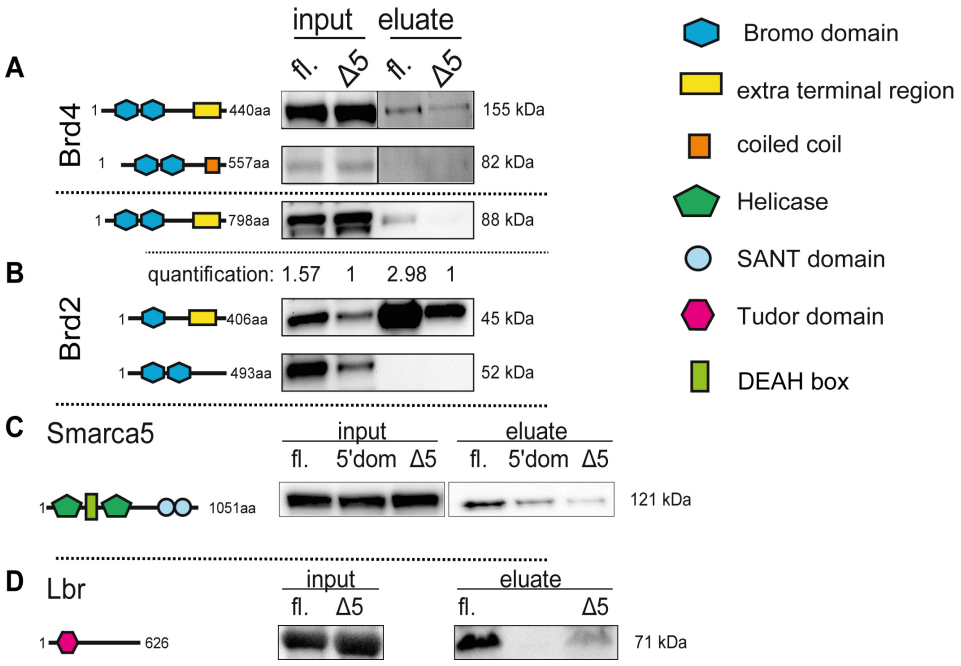


Figure 39: **RUS's binding to Brd2 4 depends on their ET domain.** Western blot analysis in input and eluate of full-length RUS, △5' construct and 5' domain. **A:** Against Brd4. **B:** Against Brd2 (quantification values are normalized to △5' construct sample's value for input and eluate separately). **C:** Against Smarca5. **D:** Against Lbr.

We validated the observed interaction by mass spectrometry by western blot analysis with antibodies against the interacting proteins Brd2, Brd4, Smarca5, and Lbr (Figure 39A-C). Therefore, we affinity-purified lncRNA protein complexes from FlpIn N2As expressing the full-length RUS, the △5' construct, and specifically for Smarca5 the 5' conserved region. Since RNAseA elution also liberates other untagged RNA species bound potentially to amylose beads during purification, I eluted protein-lncRNA complexes with maltose to ensure the protein specificity towards the tagged RNA. Eluates were loaded together with 1% of the input material on the western blot gel to compare the eluate and input fractions' protein levels.
Western blot analysis against Brd4 displayed two isoforms of Brd4 in the input samples appearing at 155 kDa and 82 kDa (Figure 39A). The 155 kDa Brd4's isoform consisting of two bromodomains and an extra terminal domain at the C terminus also appeared in

both eluates and was more enriched in the full-length sample.
Western blot analysis against Brd2 displayed three isoforms in both inputs appearing at 88 kDa, 52 kDa, and 48 kDa (Figure 39B). The isoforms at 88 kDa and 48 kDa, consisting of an extra-terminal domain, two or one bromodomains, respectively, also appeared in both eluates' fractions. Both isoforms were more abundant in the full-length sample. However, the input of the full-length sample also showed enrichment for the 48 kDa isoform. Thus, we quantified the band intensities and compared the full-length sample's intensity against 5' depletion sample's intensity for the input and eluate fraction separately. The 48 kDa isoform enriched 1.57 fold in the full-length sample's input and 2.98 fold in the full-length sample's eluate. This observation confirms that Brd2's 48 kDa isoform was more abundant in the full-length sample's eluate.
Intriguingly, the common feature of all Brd2 and Brd4 isoforms bound by the full-length construct represents the extra-terminal domain. Thus, the BET proteins' interaction (Brd2 and Brd4) to RUS may depend on the C-terminal extra-terminal domain.
Western blot analysis against Smarca5 showed only on isoform in input and eluate (Figure 39C). Full-length sample's eluate and 5' domain sample's eluate enriched more for Smarca5 than the $\triangle$5' depletion sample's eluate, whereby the strongest enrichment was observed in the full-length sample's eluate.
Similar to Smarca5, only one isoform appeared in the input and eluate by western blotting against Lbr. Consecutively, the full-length sample's eluate showed the strongest enrichment for Lbr.

6.3.4 RIP experiments reciprocally confirmed the binding of Brd2, Brd4, Smarca5, and Lbr to RUS in differentiating NSC

To test whether RUS interacts with Brd2, Brd4, Smarca5, and Lbr in NSC, I validated those interactions reciprocally by RNA-immunoprecipitation experiments (RIP). Therefore I tested whether the purification of Brd2, Brd4, Smarca5, and Lbr copurified RUS (Figure 40). Experimentally, I isolated the nuclear extract of 2 days differentiated neural stem cells and immune purified RNA with the antibodies against Brd2, Brd4, Smarca5, and Lbr under native conditions. Additionally, I used antibodies against rabbit IgG and Sox2 as negative controls. Bound RNA was isolated from the beads and analyzed by quantitive RT-PCR for the abundance of RUS and TBP as a negative control transcript. Transcript levels were compared to the abundance in the IgG control. I experimented with biological triplicates. As expected, the TBP fold change of all tested antibodies over the IgG control was between 0.8 and 1.7. The reason for this might that TBP was very low abundant in all samples. Expect Sox2 showing only 1.4x fold change of RUS, the tested antibodies against interacting proteins yielded a high and significant RUS enrichment compared to IgG control. For Brd2, I observed a 4.4x fold-change, for Brd4 a 4.3x fold-change, for Lbr a 3.75x fold-change, and Smarca5 a fold-change of 3.55x. These results confirmed that RUS interacts with Brd2, Brd4, Smarca5, and Lbr

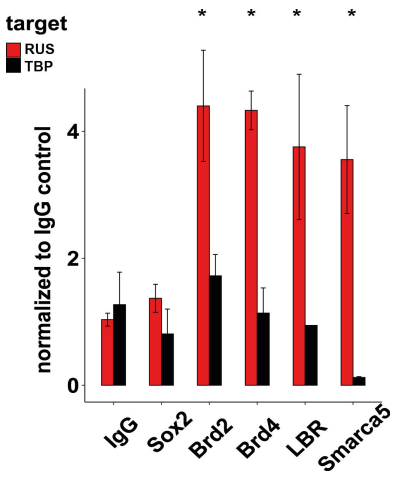


Figure 40: **Binding of RUS to Brd2 & 4, Smarca5 and Lbr can be reciprocally confirmed**. RNA immunoprecipitation of RUS and TBP mRNA using antibodies against rabbit IgG, Sox2, Brd2, Brd4, Lbr, and Smarca5 from 2 days differentiated NSC. Measured RNA levels were compared against IgG control. Experiments were done with three biological replicates. Error bars represent standard deviation. Two-tailed t-test: *: $p < 0.05$.

in differentiating NSC.

Taken together, we were able to validate all interactions observed by tandem LC-MS analysis with Western Blot analysis and reciprocally by RIP experiments. These results demonstrate that our affinity purification of *in vivo* folded tagged RUS or other lncRNAs - that is not described elsewhere - was thorough.
The specific binding of Brd2, Brd4, Smarca5, and Lbr to the conserved 5' region argues that RUS impacts chromatin structure [43, 76, 80, 83]. Intriguingly, the interaction of Lbr and Smarca5 to lncRNAs is already described in the literature for the lncRNA Xist. The lncRNA Xist interacts with Lbr to spread over X_i chromosome, enabling chromosome-wide silencing. The interaction between Xist and Smarca5 maintains the repressed state by H3K27 trimethylation at the silenced X chromosome [68]. Besides, lncRNA interaction to Smarca4 that structurally similars that to Smarca5 is already described for the lncRNA Evf2. Evf2 is binding to an enhancer between Dlx5 and 6 and represses Smarca4 by acting as a competing substrate [3].
Whether RUS is recruiting or titrating those factors to or away from distinct chromatin regions remains to be determined. Vice versa, RUS may target one protein and removes the other at its target genes. Furthermore, RUS may act as a scaffold using the one factor for chromatin binding to recruit or repress the other. Specific attention has to be given to Brd2 and Smarca5. Brd2 cooperates with the insulation factor CTCF, whose deposition is regulated by Smarca5 [43, 76].

6.3.5 Chromatin Isolation by RNA-purification (ChIRP) to determine genomic binding regions of RUS

To reveal the chromatin binding sites of RUS, I applied the Chromatin Isolation by RNA-purification (ChIRP) method. I designed eight 20 nt complementary DNA- probes targeting the RUS isoform 1. Probes were designed according to standard protocol using the online tool www.singlemoleculefish.com. The used program yielded oligonucleotides that are binding at a distance of approximately every 100 nt to RUS isoform 1. The oligonucleotide probes were numbered starting from the 5'end and ordered as 3'biotin-TEG (triethyleneglycolyl-glyceryl) modified deoxyoligonucleotides. Even-numbered probes and odd-numbered probes were pooled and separately used for the ChIRP experiment to sort out unspecific bound chromatin regions during the analysis. RNA-chromatin complexes were isolated from sheared glutaraldehyde crosslinked chromatin that I isolated from 2 days *in vitro* differentiated Wt NSC. After pull-down and extensive washing, 10% of bead material and 1% input was used to measure the Xist, Malat TBP, and RUS RNA level by quantitative RT-PCR analysis. By normalizing the RNA level in bead samples with RNA level in the input sample, I confirmed that RUS was enriched explicitly by both probe sets with an average yield of 25% (Figure 41A). From the residual 90% bead material and 1% input chromatin, DNA was isolated. The EMBL-Gene Core facility carried out the preparation and sequencing of sequencing libraries on a NextGen 500 of captured- and input chromatin. 13 Mio. reads of the even-probes sample and 26 Mio. reads of the odd-probes sample were aligned against mm10 mouse reference genome using bowtie2. Peaks were called with MACS 14 using input chromatin as control. MACS 14 called 36,128 peaks in the odd probes sample and 27,742 peaks in the even probes sample. As peaks detected only in the even or in the odd probe set were considered RUS unspecific, they were filtered out. Furthermore, peaks detected simultaneously in both probe sets were controlled according to their intensity. I considered only peaks that showed similar intensity in both probe sets as specific RUS bound regions. To unravel those regions, I used the R packages GenomicRanges to extract all summits from the even and odd probes peaks that are less than 100 nt distant. Next, I calculated the reads occupancy 500 nt around the mean summit in the even and odd sample using the R subread package to calculate a weighted difference:

$$weightedDifference = \frac{even + odd}{odd} - \frac{even + odd}{even}$$

From 346 coinciding peaks between both samples, 41 peaks (Table 30) showed a weighted difference less than +3 and more than -3. Intriguingly, two peaks overlap to the genomic region of exon 1 and exon 2 of RUS (Figure 41B). Both regions were potentially copurified from the nascent RUS transcript. I excluded those peaks for further analysis and selected 20 top-peaks with a mean coverage higher than 200 reads. I used homer software to annotate those peaks (Table 31, Figure 41C). Fourteen of those peaks are intergenic located, five peaks are located in introns, and one peak is located in the promoter of the potassium channel Kcna5 (Figure 41D).

Table 30: Genomic location and annotation of all 41 RUS target sites

Chr	Start	End	Annotation	Distance to TSS	Gene Name	Gene Type
chr1	12644711	12644862	Intergenic	-22777	Gm17644	ncRNA
chr1	109644159	109644446	Intergenic	-338598	Cdh7	protein-coding
chr1	150128427	150128611	Intergenic	28488	Ptgs2	protein-coding
chr1	180204487	180204715	Intergenic	-8581	Coq8a	protein-coding
chr10	85211616	85211839	Intergenic	-26673	Cry1	protein-coding
chr10	115633121	115633270	Intergenic	-45415	Lgr5	protein-coding
chr10	120466028	120466192	intron (NM_001347170, intron 2 of 5)	10359	Hmga2	protein-coding
chr10	121378183	121378468	intron (NM_029364, intron 6 of 13)	13235	Gns	protein-coding
chr13	58595200	58595433	intron (NM_022317, intron 1 of 17)	15561	Slc28a3	protein-coding
chr14	67356884	67357035	intron (NR_045961, intron 3 of 5)	40783	4930438E09Rik	ncRNA
chr14	90109416	90109619	Intergenic	-17374	4930474H20Rik	ncRNA
chr14	104336827	104337125	Intergenic	131023	Pou4f1	protein-coding
chr15	38361215	38361400	Intergenic	-60596	Klf10	protein-coding
chr16	81874791	81874966	Intergenic	674181	Ncam2	protein-coding
chr17	4626269	4626422	Intergenic	-8635	Arid1b	protein-coding
chr17	56199969	56200120	intron (NM_001360284, intron 10 of 20)	-16124	Mydgf	protein-coding
chr17	63256696	63256886	intron (NM_015794, intron 6 of 8)	56182	4930405O22Rik	ncRNA
chr17	67480569	67480779	Intergenic	-126183	Ptprm	protein-coding
chr17	86947870	86948152	promoter-TSS (NM_029121)	-124	Atp6v1e2	protein-coding

chr18	32442601	32442786	Intergenic	-64409	A830052D11Rik	ncRNA
chr18	40308136	40308305	intron (NM_026135, intron 1 of 1)	49859	Kctd16	protein-coding
chr18	54371167	54371344	Intergenic	-51040	Redrum	ncRNA
chr19	34119278	34119477	intron (NM_023903, intron 9 of 9)	18434	Lipm	protein-coding
chr2	39594697	39594948	Intergenic	-368473	Ppp6c	protein-coding
chr2	60402721	60402876	Intergenic	-8727	Gm13580	ncRNA
chr3	73066368	73066647	non-coding (NR_040557, exon 1 of 5)	156	Gm20754	ncRNA
chr3	73594430	73594714	non-coding (NR_040557, exon 5 of 5)	-112921	4930509J09Rik	ncRNA
chr3	85074303	85074493	Intergenic	122187	Fbxw7	protein-coding
chr3	86860448	86860617	intron (NM_001195500, intron 2 of 11)	60352	Dclk2	protein-coding
chr3	101687440	101687671	Intergenic	-82848	Atp1a1	protein-coding
chr4	25372102	25372298	Intergenic	-90379	Ufl1	protein-coding
chr4	98547743	98548033	intron (NM_001005787, intron 1 of 15)	1254	Patj	protein-coding
chr6	73119912	73120149	intron (NM_001164669, intron 39 of 76)	101601	Dnah6	protein-coding
chr6	126535671	126535834	promoter-TSS (NM_145983)	-197	Kcna5	protein-coding
chr6	126613636	126613842	Intergenic	32062	Kcna1	protein-coding
chr8	43196655	43196809	Intergenic	13610	Triml2	protein-coding
chr9	24911390	24911559	Intergenic	-62734	E130101E03Rik	ncRNA
chr9	80456462	80456635	intron (NM_022016, intron 1 of 16)	8890	Impg1	protein-coding
chrX	59771099	59771317	Intergenic	-185636	Fgf13	protein-coding
chrX	90795323	90795532	Intergenic	-96715	Gm44	protein-coding
chrX	142068707	142068870	Intergenic	128148	Gucy2f	protein-coding

Table 31: Genomic location and annotation of the top 20 RUS peaks

Chr	Start	End	Annotation	Distance to TSS	Gene Name	Gene Type
chr1	12644271	12645271	Intergenic	-22792	Gm17644	ncRNA
chr1	109643746	109644746	Intergenic	-338654	Cdh7	protein-coding
chr1	150128027	150129027	Intergenic	28496	Ptgs2	protein-coding

chr10	85211294	85212294	Intergenic	-26740	Cry1	protein-coding
chr10	120465674	120466674	intron (NM_001347170, intron 2 of 5)	10295	Hmga2	protein-coding
chr14	90108945	90109945	Intergenic	-17446	4930474H20Rik	ncRNA
chr16	81874406	81875406	Intergenic	674209	Ncam2	protein-coding
chr17	4625876	4626876	Intergenic	-8604	Arid1b	protein-coding
chr17	56199521	56200521	intron (NM_001360284, intron 10 of 20)	-16101	Mydgf	protein-coding
chr17	67480134	67481134	Intergenic	-126143	Ptprm	protein-coding
chr18	32442215	32443215	Intergenic	-64431	A830052D11Rik	ncRNA
chr18	40307795	40308795	intron (NM_026135, intron 1 of 1)	49934	Kctd16	protein-coding
chr18	54370725	54371725	Intergenic	-51070	Redrum	ncRNA
chr2	39594306	39595306	Intergenic	-368457	Ppp6c	protein-coding
chr3	85073883	85074883	Intergenic	122172	Fbxw7	protein-coding
chr4	25371658	25372658	Intergenic	-90337	Ufl1	protein-coding
chr4	98547353	98548353	intron (NM_001005787, intron 1 of 15)	1219	Patj	protein-coding
chr6	73119596	73120596	intron (NM_001164669, intron 39 of 76)	101535	Dnah6	protein-coding
chr6	126535182	126536182	promoter-TSS (NM_145983)	-127	Kcna5	protein-coding
chrX	90794896	90795896	Intergenic	-96746	Gm44	ncRNA

Next, I used MEME software to identify a consensus motif from the 20 selected sites. I was able to unravel a 28 bp purine-rich motif present in 12 of 20 regions with a p-value of 1.8e-32 and a FIMO score value of at least 32 (Figure 41E). Using FIMO tool, I predicted all genomic regions of this motif in mouse, human, and zebrafish genome (Figure 41G) using a cut-off score value of 32. Whereas we identified only 18 regions in the zebrafish genome, we identified 8,000 sites in the human and 7,050 in the mouse genome. 1,792 regions are well conserved by synteny. Using GO enrichment analysis, I figured out that syntenic conserved sites are close to neural relevant genes (Figure 41G). During the investigation of FIMO sites, I figured out that RUS occupancy is increased if the motif is repeated within a distance of 55bp. Those repetitive regions are occurring 75x times in the mouse genome and 80x times in the human genome. Twenty-five sites of the repetitive regions are conserved by synteny. Next, I calculated the overall occupancy of ChRIP reads to those regions (Figure 41H) using the Rsubread package. As evaluated by CHIRP, RUS covers 18 sites, approximately 25% of all predicted sites. According to this result, we postulated that the tandem repeat of this motif is required for genomic RUS binding.

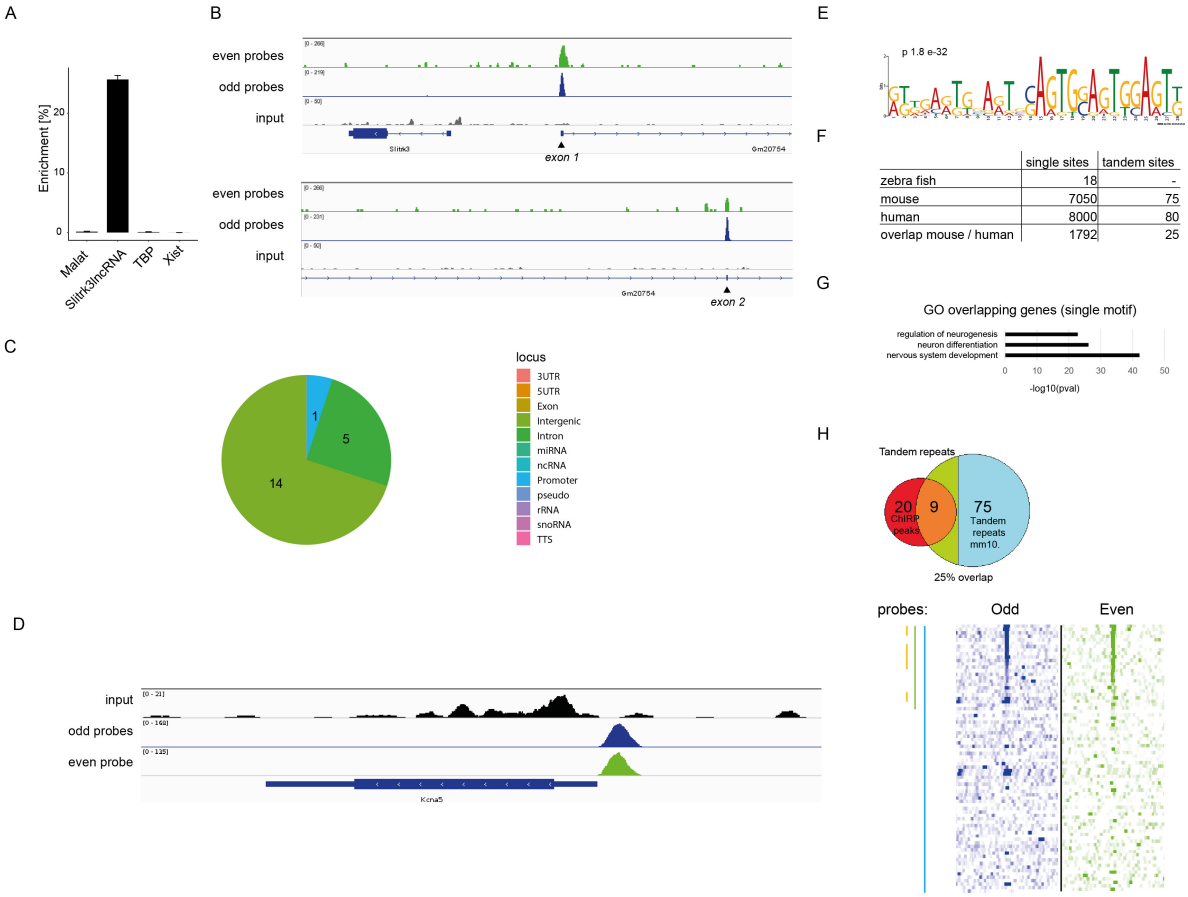


Figure 41: **Determination of RUS bound chromatin regions by ChIRP. A:** enrichment of RUS, TBP, Malat, and Xist by ChIRP as measured by quantitive RT-PCR. Ct values of even and odd samples were normalized to input. **B:** Tracks of ChIRP samples (green: even, blue: odd, grey: input) in the genomic region of RUS. **C:** Genomic distribution of top-20 ranked ChIRP peaks. **D:** Tracks of ChIRP samples (green: even, blue: odd, black: input) in the genomic region of Kcna5. **E:** DNA binding motif discovered for RUS from 20 top-ranked peaks. **F:** Table of the single motif and tandem motif in mouse, human, and zebrafish genome. **G:** GO enrichment analysis of overlapping genes next to the single motif. **H:** calculation of RUS occupancy in the mouse genome. Venn Diagram to show the overlap of tandem motif sites in the genome and covered site. The density map shows the read density in the odd and even samples +/-5 kbp around the tandem repeat motif. Sites were ranked according to mean coverage in the even and odd sample. Blue bar: all 75 tandem repeat sites in the mouse genome. Green bar: covered by RUS. Yellow bars: top-20 ranked peaks.

6.3.6 Analysis of genes neighboring RUS bound regions in ctrl and RUS depleted cells

Next, I compared the expression of genes nearby RUS bound regions in ctrl and SH1 and SH2 RUS KD condition using RNA-Seq results from samples collected on day 7 (Figure 42, 6.2.2). From 16 genes, three genes: Cdh7, Kctd16, Ppc6 were downregulated, and five genes: Arid1b, Dpp9, Gdnf, Kcna5, and Ptprm were upregulated in both SH1 and SH2 KD conditions. Mygdf, Bin1, and Gm17644 were solely upregulated by SH1, maybe due to a distinct function of isoform1 on those genes. These results suggested that RUS has a repressive function on its target genes rather.

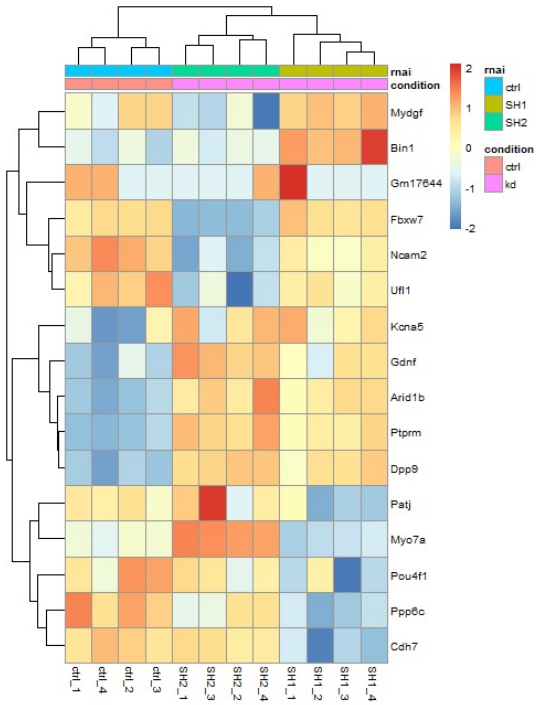


Figure 42: **Analysis of genes neighboring RUS bound regions in ctrl and RUS depleted cells.** Euclidean distance method: grouping of the z-scores of 16 genes located next to the top 20 RUS peaks using RNA-seq data from samples collected on day 7 (paragraph 6.2.2).

6.3.7 Chromatin Immune Precipitation (CHIP) in ctrl and RUS depleted cells: RUS binds to Brd2 occupied repetitive genomic regions

To investigate whether RUS has any recruiting or titrating effects on its interacting proteins, we performed ChIP experiments (chromatin immune precipitation). Therefore, I focused on the potential interplay between Brd2 and Smarca5. Since ChIP-Seq approved antibodies were commercially available for Brd2, we performed ChIP against Brd2.

We used chromatin isolated from formaldehyde crosslinked ctrl KD and RUS SH1 KD cells that we differentiated for two days. The EMBL-Gene Core facility carried out the

preparation and sequencing of ChIP and input sequencing libraries on a NextGen 500. Sequencing reads were aligned against mm10 mouse genome using bowtie2. Peaks were called with MACS14 using the respective input as control.
Although only 500,000-900,000 reads were aligned to mm10 genome, peaks-calling yielded 4501 high confident Brd2 peaks with an -log10 p-value greater than 100 in all four samples. As expected, a significant amount of 1253 peaks are located in the promoter regions of genes (Figure 43A). Next, I calculated the overlap between Brd2 peaks and RUS bound regions. Brd2 occupies approximately 33% (13 of 39, Table 32) of initial determined RUS bound regions. As observed before, these sites are either intergenic and intronic located (Figure 43B) as in the case of Dpp9 (Figure 43C). I did not observe any differences in Brd2 occupancy between ctrl and RUS at this site, arguing that RUS has no impact on Brd2 occupancy. However, I was able to identify using MEME software a 29bp long AATGG repetitive motif that was present in all 13 peaks with a p-value of 1e-122 (Figure 43D) and a score value greater than 42. I used the FIMO tool to extract 42 regions showing this motif with a score higher than 42 in the mouse genome. Comparing those regions with Brd2 peaks yielded an overlap of 31 sites.

Table 32: Genomic location and annotation of the 13 Brd2 bound RUS target sites

Chr	Start	End	Annotation	Distance to TSS	Gene Name	Gene Type
chr1	12644659	12644957	Intergenic	-22755	Gm17644	ncRNA
chr1	44951107	44951405	Intergenic	-206307	4930521E06Rik	ncRNA
chr11	25099952	25100322	Intergenic	-225970	4933427E13Rik	ncRNA
chr14	104336827	104337125	Intergenic	131023	Pou4f1	protein-coding
chr17	4626046	4626699	Intergenic	-8608	Arid1b	protein-coding
chr17	56199893	56200191	intron (NM_001360284, intron 10 of 20)	-16122	Mydgf	protein-coding
chr2	60402650	60402948	Intergenic	-8726	Gm13580	ncRNA
chr3	153123639	153123937	Intergenic	28320	4930482G09Rik	ncRNA
chr6	10186269	10186567	Intergenic	787960	AA545190	ncRNA
chr6	125876793	125877091	intron (NM_153589, intron 13 of 25)	186523	Ano2	protein-coding
chr6	132463611	132463909	Intergenic	-99625	Gm4736	protein-coding
chr7	98059921	98060219	intron (NM_008663, intron 39 of 48)	49178	Myo7a	protein-coding
chr9	19575323	19575621	Intergenic	-8090	Olfr854	protein-coding

I calculated with R subread the occupancy of RUS using the ChIRP odd and even sample. RUS covers 14 of 42 genomic motif sites matching the 39 RUS targets sites initially determined by ChIRP (Figure 43E). Next, I calculated Brd2's occupancy +/- 5kbp around RUS occupied- and unoccupied motif sites in ctrl and KD samples by applying Homer tools and the SeqMiner software (Figure 43F & G). Notable, RUS sites showed an increased occupancy by Brd2, arguing that RUS impacts Brd2 binding to those sites. However, as observed for Dpp9, KD of RUS resulted in no change of Brd2 binding.
Intriguingly, the same 13 sites occupied by Brd2 match the tandem repeats I determined before. These results indicate that the binding of RUS occurs predominantly in repeti-

tive genomic tracts and is primarily determined by the occupancy of Brd2. To identify other proteins that mediate the binding of RUS to the determined motif site *in silico*, I used the TomTom software. As a hit, I obtained the enhancer factor TEAD1. However, this factor is low expressed in the nervous system and does not appear in the proteome data as a potential interactor.

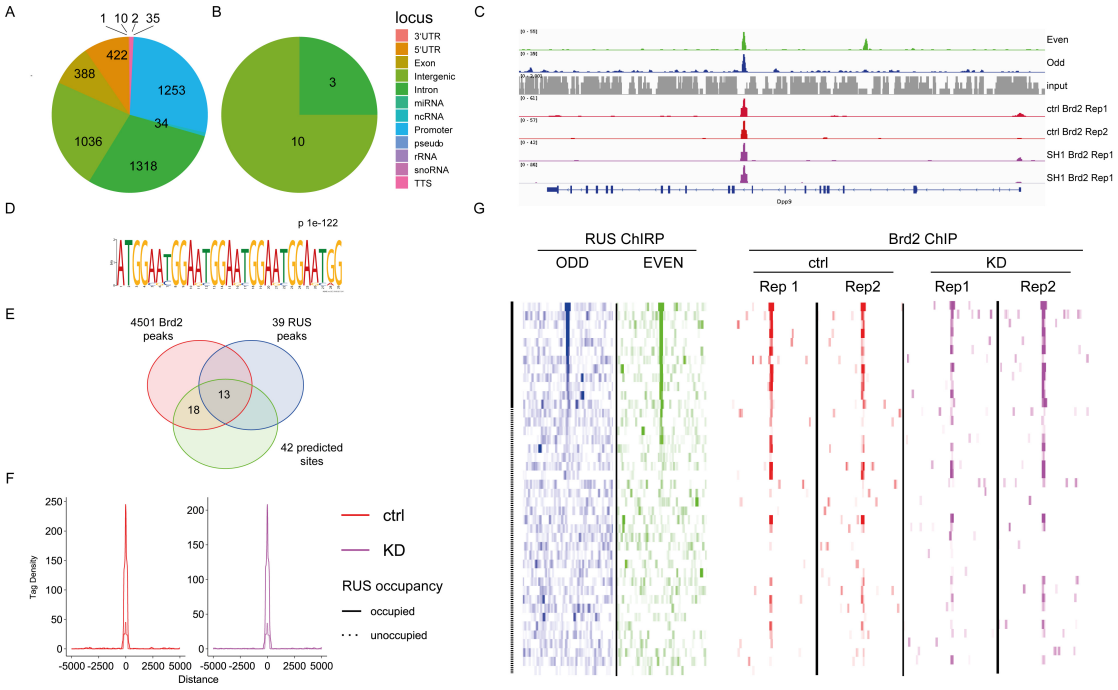


Figure 43: **RUS binds to Brd2 covered ATGGA repeats. Overlap of genomic regions bound by Brd2 and RUS. A:** Genomic distribution of Brd2 peaks analyzed by ChIP. **B:** Genomic distribution of sites occupied by RUS determined by ChIRP and Brd2 determined by ChIP. **C:** Tracks of RUS ChIRP samples in Wt cells (green: even, blue: odd, grey: input) and Brd2 ChIP samples in ctrl (red) and RUS KD (magenta) cells show the coding region of Dpp9. **D:** DNA binding motif present in CHIP RUS (ChIRP) and Brd2 (ChIP) covered sites. **E:** Venn Diagram shows the overlap of ChRIRP RUS peaks, Brd2 ChIP peaks, and genomic sites harboring the DNA binding motif. **F:** Occupancy of Brd2 in ctrl (red) and KD (magenta) cells around RUS covered (solid line) and uncovered (dotted line) motif sites. **G:** Density map shows the read density in the odd (blue) and even (green) RUS ChIRP samples in Wt cells and Brd2 ChIP samples in ctrl (red) and Slitrk KD cells (magenta) +/- 5kbp around tandem repeat motif. Sites were ranked according to mean coverage in the even and odd sample.

6.3.8 Unspliced RUS may bind to repetitive genomic regions by the formation of RNA-DNA triplexes or RNA-RNA duplexes

Although chromatin-associated proteins frequently mediate lncRNAs' chromatin binding, lncRNAs also bind directly to chromatin by RNA-DNA triplex formation [107]. Tandem repeats are prone to form RNA-DNA triplexes [174]. Besides, tandem repeats

are transcriptionally active and are transcribed into enhancer RNAs [175]. These facts raise the two hypotheses: (1) RUS forms an RNA-DNA triplex with the tandem repeats, or (2) RUS base pairs the potential eRNA expressed from the repeat region. In cases of intronic repeat regions, the premature mRNA, such as Dpp9 by itself, may act as an RNA substrate for RUS. ATGGA represents a simple DNA repeat and is prone to form RNA-DNA triplexes [174, 176]. I used the Triplexator software using the peaks site sequences as triplex targeting sites (TTS) and the spliced RUS transcript sequence as the triplex-forming oligonucleotide (TFO). This task yielded only 2 of 13 sites containing a TFOs. Besides, potential triplexes do not match the 29 bp long GAATG repetitive motif shown in (Figure 43D). That indicates that DNA-RNA triplex formation has no impact on chromatin binding. However, I repeated the Triplexator search with the 29 bp repetitive motif as a TTS and the non-spliced premature RUS transcript as a TFO. This approach yielded a 29 nt long parallel TFO in the premature transcript lying in the intron between exon 1 and 2 (Figure 44A & B). The human unspliced-transcript has two antiparallel TFOs located in the introns between exon 1 & 2 and between exon 3 & 4. The same TFO site in the murine unspliced transcript may form a 27 nt long double-stranded RNA product with the transcribed repetitive region displaying a mismatch every fifth base pair (Figure 44B). In humans, consequently, the reverse complementary strand must be transcribed from the repetitive regions.

To test for the predicted triplex's existence, we chemically labeled 5' amino-modified pyrimidine TTS-DNA strand with NHS activated TAMRA and annealed it with the unmodified purine TTS-DNA strand. Besides, we chemically labeled 5' amino-modified pyrimidine parallel TFO-RNA with NHS-activated FITC. Next, we performed EMSA assays with 4 pmol TAMRA-labeled TTS-DNA and increased the amount of TFO-RNA, ranging from 20 pmol to 320 pmol. Since parallel triplexes required the cytosine protonation to form Hoogsten base pairs, we performed the EMSA in 20 mM $MgCl_2$ and Tris-acetate pH 5. However, the low pH- condition resulted in the breakdown of the DNA duplex. Then we performed the EMSA assay in 20 mM $MgCl_2$ and Tris-acetate pH 7.4 (Figure 44C). This condition preserved the DNA duplex structure but resulted in the formation of a weak RNA-DNA triplex. To study the predicted RNA duplex, we annealed the FITC labeled RNA with the GGUAA repetitive RNA. The annealing reaction yielded a stable RNA duplex structure (Figure 44C). These results indicate that substrate sensing by RUS may happen co-transcriptionally.

Since Brd2 occupies only 13 of 41 ChIRP peaks, I suppose that RUS acts on two or more distinct regulatory pathways comprising one Brd2-dependent. However, in several cases, the next Brd2 peak was close to the ChIRP peak as the case for Kcna5.
I used the R BioMart packages to map all 13 Brd2 and RUS bound sites to genomic regulatory features. All regions coincide with promoter flanking and enhancer.

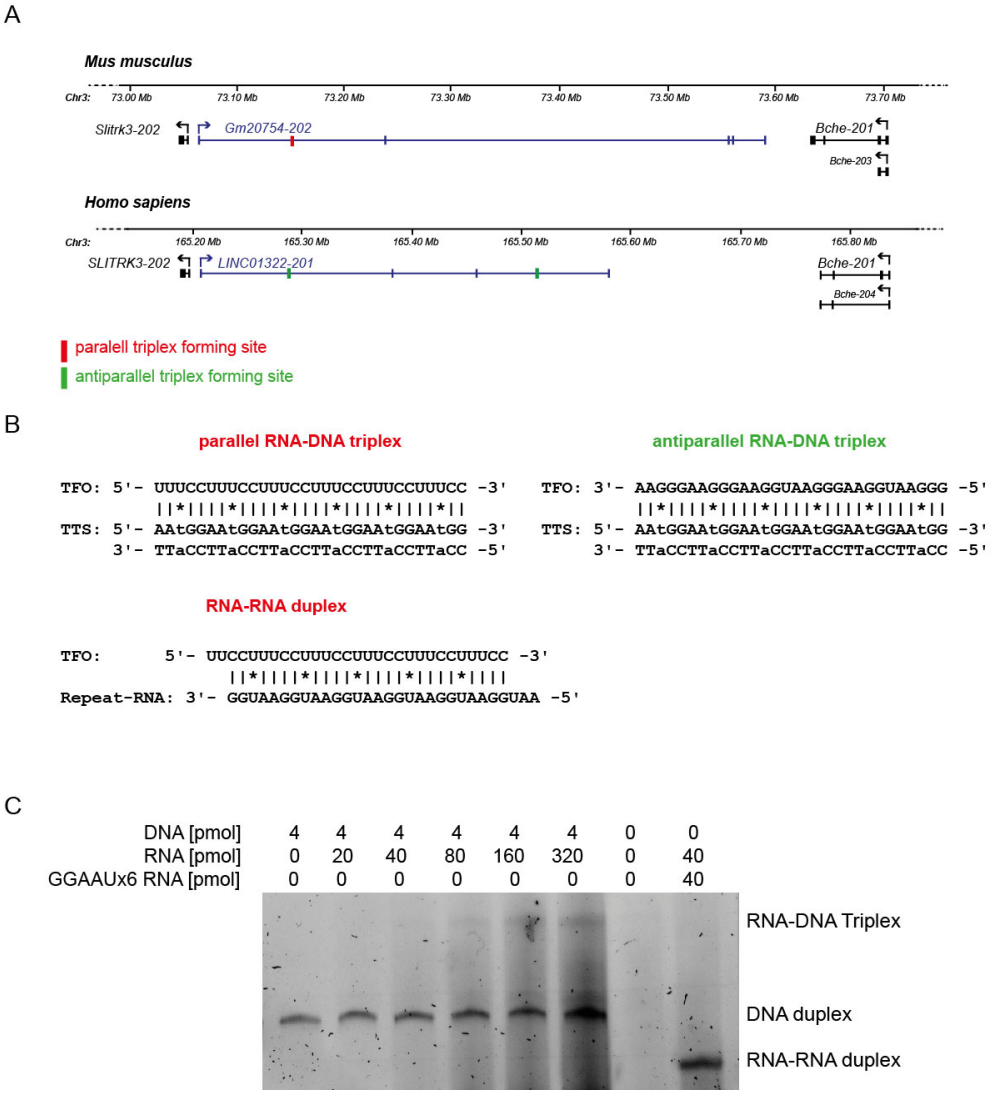


Figure 44: **Co-transcriptional substrate sensing by RUS. A:** Conserved Triplex forming sites in the unspliced mouse and human RUS transcript. RUS locus in mouse and human. (red: parallel triplex forming site, green: antiparallel triplex-forming site). **B:** top: Sequence of parallel and antiparallel triplexes around the repeat region in mice and humans, respectively. Bottom: potential RNA-RNA duplex between the triplex-forming site with transcripts produced from repeat regions in mice **C:** EMSA assay of TAMRA stained TTS-DNA and FITC stained TFO-RNA. Annealing reactions were heated to 95°C for 5 min and cooled down slowly. Reactions were performed in 20 mM Tris-acetate pH 7.4, 20 mM $MgCl_2$ and 150 mM KCl for 2 h at RT.

6.3.9 RUS and Brd2 bind together on hippocampal specific CTCF sites

Smarca5 positively regulates and maintains the occupancy of CTCF on chromatin [43, 76]. Thus, Smarca5 and RUS might act on the same genomic sites predominantly enriched for CTCF. Public available ChIP-Seq Data sets performed in many different murine cell-types well described the occupancy of CTCF.

To test whether CTCF binds RUS targets sites, I analyzed the publicly available data from CTCF ChIP experiments performed in as ESC, NSC, cortical neurons (CX), and hippocampal neurons (Hip) (Figure 45) [177–180]. Thereby I focused on the 13 RUS target sites bound by Brd2.
RUS target sites, as examples around the protein-coding regions of Arid1b (AT-Rich Interaction Domain 1B, Figure 45A), Dpp9 (Dipeptidyl peptidase 9, Figure 45B), and Pou4f1 (Pou domain class 4 transcription factor 1, Figure 45C) displayed strong CTCF peaks in hippocampal neurons. As examples around Arid1b and Pou4f1, several RUS target sites displayed weak CTCF peaks in ESC. No CTCF peaks were observed in neural stem cells and cortical neurons. These results suggested that RUS acts primarily on hippocampal specific CTCF sites.
To evaluate the hippocampal specificity, I calculated the overall occupancy of CTCF +/- 5kbp around the 13 Brd2 bound RUS target sites in ESC, NSC, cortical, and hippocampal neurons using Homer tools. As background control, I called all CTCF peaks in the four samples separately using MACS, combined them, and calculated the CTCF occupancy +/- 5kbp them in the four samples (Figure 45D).
In line with previous observations (Figure 45A-C), hippocampal neurons showed the strongest and ESCs a weak tag density enrichment around the 13 Brd2 bound RUS target sites. This overrepresentation of hippocampal tag density was not observed in the background control. As expected, all samples showed an enriched tag density around all called CTCF peaks, whereby cortical neurons showed the strongest enrichment.
To evaluate the overrepresentation of CTCF peaks around Brd2 bound RUS target sites in hippocampal neurons statistically, I compared the distribution of peak intensities between all CTCF and the 13 Brd2 bound RUS targets sites for each of the four samples (Figure 45E). Therefore, I calculated the normalized tag reads for all CTCF and Brd2 bound RUS target sites using homer tools to calculate the p-value between the two peak-sets using a two-tailed student's t-test. Only in hippocampal neurons, the mean CTCF intensity around the 13 Brd2 bound RUS target sites overtopped the mean CTCF intensities of all CTCF sites significantly with a p-value of 9.75e-4.
These data statistical confirmed that RUS and Brd2 are acting on hippocampal specific CTCF sites, a region where RUS is highly expressed (Figure 17F). Since ESC also displayed CTCF peaks but not NSC on those sites, let suggest that RUS might be involved in regaining CTCF occupancy at its target sites during the transition from NSCs to hippocampal neurons.

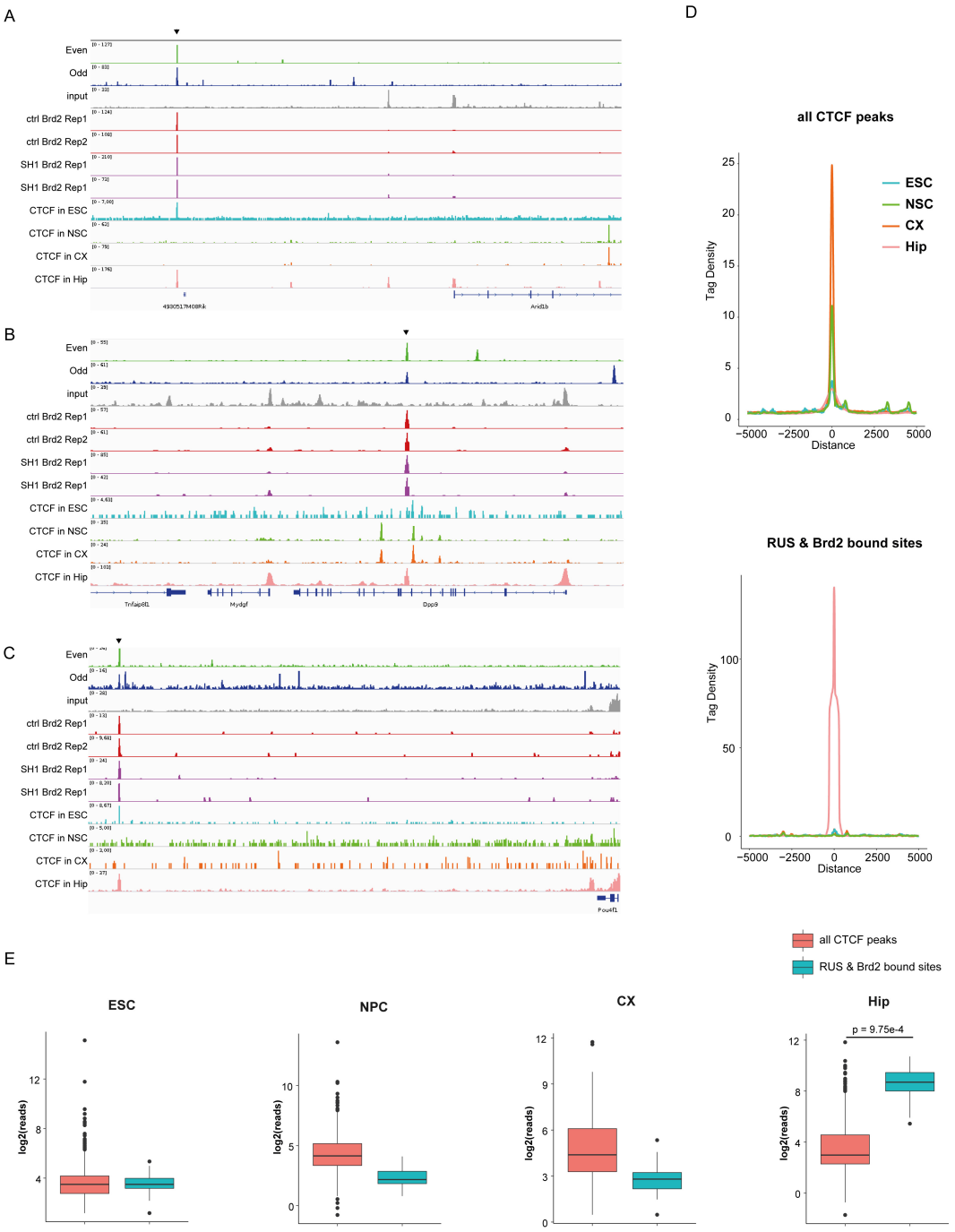


Figure 45: **RUS and Brd2 acts on hippocampal specific CTCF sites. A-C:** Tracks of RUS ChIRP samples in Wt cells (Even, Odd, and input), Brd2 ChIP samples in ctrl and RUS KD cells, and CTCF ChIP samples in ESC, NSC, CX (cortical neurons), and Hip (hippocampal neurons) showing the regions around Arid1b **(A)**, Dpp9 **(B)**, and Pou4f1 **(C)**(co-occupied sites indicated by arrows). **D:** Occupancy of CTCF in ESC (Robin's Egg blue), NSC (Harlequin's green), CX (orange), and Hip (Light salmon pink) cells +/- 5kbp around all called CTCF (left) and the 13 Brd2 bound RUS target sites (right). **E:** Box plots showing the distributions of normalized tag reads in all CTCF and the 13 Brd2 bound RUS target sites in ESC, NSC, CX, and Hip (two-tailed t-test).

6.3.10 RUS regulates CTCF occupancy by Smarca5's activity modulation

Next, I tested whether RUS regulates the occupancy of CTCF at its target sites by modulating the occupancy or activity of Smarca5. Therefore, we performed ChIP experiments against CTCF (Figure 46A-D) combined with quantitative RT-PCR in ctrl, RUS, and Smarca5 KD NSC to measure whether RUS KD affects the occupancy of CTCF and its target sites and whether this effect depends on the activity of Smarca5. Besides, we performed ChIP experiments against Smarca5 to depict whether a reduced Smarca5's occupancy can explain the observed effect (Figure 46E-H). Thereby, I focused on the RUS target sites around the genomic regions of Arid1 (Figure 32A & E), Dpp9 (Figure 32B & F), and Pou4f1 (Figure 32C & G). As a control region, we used the boundary element called rs14, located downstream of the transcriptional locus of Xist. Rs14 is bound by CTCF and required for Xist spreading [181]. ChIP experiments against CTCF were performed in biological duplicates, including 1-2 technical replicates. Chip experiments against Smarca5 were performed in biological duplicates. Ct values in ChIP samples were normalized to the respective input sample's Ct value considering the ChIP and input sample's dilution coefficient. Normalized ChIP values were compared with the ctrl KD's normalized value.

As expected, Smarca5 KD diminished the occupancy of CTCF around RUS's target sites around Arid1b (by 83%), Dpp9 (by 76%), Pou4f1 (by 62%), and at the control region rs14 (by 65%) globally (Figure 46A-D). Diminished CTCF occupancy after Smarca5 depletion is the consequence of reduced Smarca5 levels at those sites (Arid1b by 35%, Dpp9 by 70%, Pou4f1 by 83%, and rs14 by 55%, Figure 46E-H).

Intriguingly, RUS's KD diminished the CTCF's occupancy selectively at its target sites around Arid1b (by 62%), Dpp9 (by 72%), Pou4f1 (by 34%), but not at the control region rs14 (Figure 46A-D). However, RUS's KD did not affect Smarca5 occupancy (Figure 46E-H). Thus, the reduced CTCF occupancy can not be explained by reduced Smarca5's levels at RUS's target sites.

These results demonstrated that RUS does not affect the occupancy of Smarca5 but rather activates Smarca5's chromatin-remodeling activity to enable CTCF binding at hippocampal specific CTCF sites.

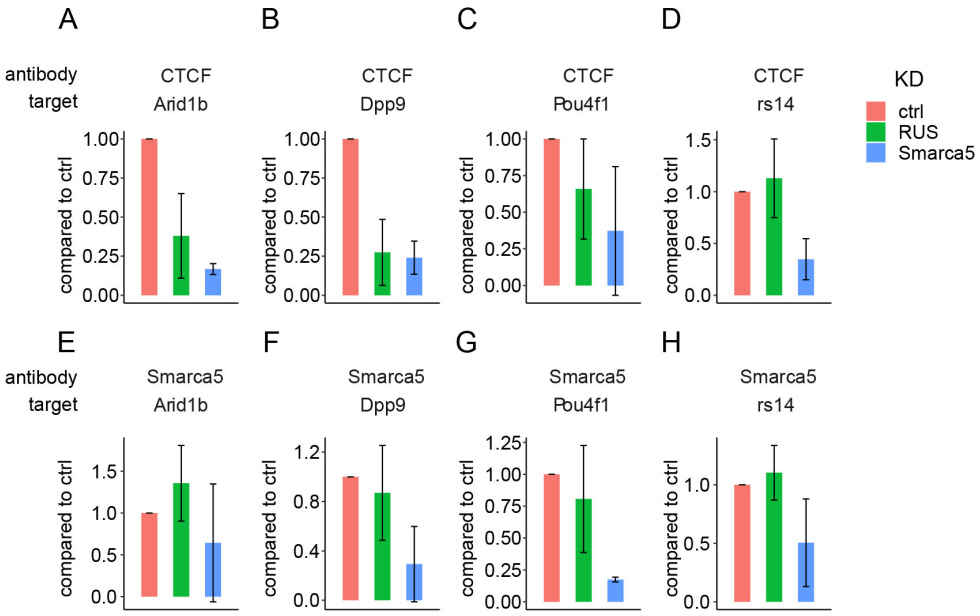


Figure 46: **RUS regulates CTCF occupancy by modulating Smarca5's activity.** ChIP experiments against CTCF (**A-D**, performed in biological duplicates including 1- 2 technical replicates) and Smarca5 (**F-H**, performed in biological duplicates) in ctrl (red), RUS KD (green), and Smarca5 KD (blue) cells. ChIP experiments were analyzed by quantitative RT-PCR using primers targeting the RUS target regions around the coding regions of Arid1b **(A+E)**, Dpp9 **(B+F)**, and Pou4f1 **(C+G)** and the control region rs14 **(D+H)**. ChIPs' Ct values were normalized to input and compared with ctrl KD's normalized values.

7 Discussion

This doctoral thesis describes the identification of an unknown and still undescribed lncRNA by a KD-phenotype screening approach. Potential candidates were preselected from publicly available transcriptome datasets. We called the lncRNA RUS (RNA upstream to Slitrk3) because it is transcribed near the transcription locus of the neurodevelopmental transmembrane protein Slitrk3. I identified that RUS's KD inhibits the formation of neurons in our *ex vivo* model. We identified that RUS acts on Brd2 occupied GAATG repetitive genomic sites and might activate the chromatin remodeler Smarca5.

7.1 RUS in neurogenesis

I recapitulated the formation in vitro by isolating neural stem cells (NSC) from embryonic cortices at E13-E16 that we cultured in the presence of basic FGF and differentiated them into neurons in the absence of basic FGF. Neurons typically appeared 2 days after differentiation in the absence of basic FGF as visualized by the neuronal marker βtubulin-III and Mapt. Using quantitative RT-PCR, I figured out that neuroectodermal and neural-stem-cell markers decreased, and the neuronal-markers increased over differentiating time courses. This assay typically yielded 20-40% neurons. Newly formed neurons typically showed axonal and dendritic morphologies. Due to culture conditions, cells can be kept for a maximum of 4-6 days in the absence of basic FGF. Hence, neurons were not forming synapses and other mature hallmarks of neuron formation. Thus, the assay was adequate to recapitulate the formation but not the maturation of neurons. In contrast to other differentiation models using a defined mixture of growth factors to produce the defined neuronal cell type, the presented model system produces a mixture of different neuronal cell types.
RUS expression is highly restricted to neural tissues and highest in the hippocampus. In accordance, RUS's expression increased over the differentiating time course in our cellular model. During differentiation, RUS underwent alternative splicing. The first occurring isoform contains exon 1,2,3, and 5 but lacks exon 4. Isoform 2 includes exon 4 but lacks exon 5, and its expression is delayed. Isoform 3 contains all exons and is similar to isoform 2 expressed. The two used KD shRNAs targeted different isoforms of RUS. Both KD resulted in a similar phenotype, arguing that both isoforms have the same function. ShRNA 1 targets exon 5 of the most and earlier abundant isoform 1 and less abundant isoform 3. ShRNA 2 targets exon 5 of the lower abundant isoforms 2 and 3. In accordance, shRNA 1 resulted in a stronger reduction of RUS and a stronger phenotype than shRNA2.
The KD of RUS in our *ex vivo* model resulted in a reduced number of proliferative cells, a reduced number of βtubulin-III, Mapt, and Map2 positive cells. Accordingly, RUS's KD in cortical neuron culture resulted in reduced Mapt, Sox2, and nestin-positive cells. Also, I observed more apoptotic cells in the RUS KD conditions.
Since lncRNAs may interfere in the expression of the direct neighbor gene, I tested

whether RUS affects the expression of Slitrk3. Although I observed a decreased expression of Slitrk3 by RUS KD in NSC, I did not observe this effect in cortical neurons. Additionally, Slitrk3's KD does not copy RUS's KD phenotype. In general, neuron-specific lncRNAs locate genomically next to neurodevelopmental genes and are coexpressed. lncRNAs are supposed to increase the complexity of gene-regulation rather than regulate the expression of the neighboring gene. Thus, an increased number of lncRNAs correlates positively with brain size in an exponential way [1].
The presented RNA-seq experiments that we performed together with the EMBL-Gene Core facility confirmed the KD-phenotype observed by staining against proliferative and apoptotic markers. RUS's KD downregulated genes associated with cell proliferation as the cell cycle genes (e.g., Cdc7, Cdc25, and E2F7) and genes associated with RNA metabolism as splicing factors (e.g., Srsf11 13), members of the exome complex (e.g., Exosc7 and -11) and ribosomal genes (e.g., Rsp7 and Rsp5). In turn, RUS's KD upregulated genes involved in the apoptotic pathway (e.g., Casp6 and -9). Decreased expression of ribosomal genes explained the observation that the neuronal markers β-tubulin III and Mapt were not effectively reduced on the RNA but protein level. Intriguingly genes blocking post-mitotic neuronal differentiation (e.g., Notch1 and Rest) were first downregulated than upregulated.

However, we have to consider that the isolated NSC used for our *ex vivo* model comprising different neural precursors and differentiation of those cells by basic FGF withdrawing, producing a mixed population of neurons, astrocytes, and other glial cells. Thus, not all cell types might express RUS. Therefore, the affected genes do not necessarily represent RUS's target genes but display RUS KD arrested cells in their differentiation stage and undergo apoptosis. Moreover, RUS's target genes might be less affected, and transcriptional changes on these genes by RUS's KD are not well resolved in the transcriptome data. In general, secondary and tertiary effects are common in biological systems and make it difficult to assign target genes of a particular gene-regulatory factor from the genes affected by the respective KD or KO.
Moreover, the expression of neurogenic lncRNAs is often restricted to one specific neuronal cell type [1]. Thus it would be worthwhile to unravel the neuronal cell type expressing RUS by *in vivo* experiments and the effect of RUS's KD phenotype in those cells. Latter may better characterize the function of RUS on its target genes.

7.2 Molecular mechanism

7.2.1 Identification of RUS's interacting proteins

For the subsequent studies, I focused on the more abundant RUS isoform 1. The 5' region of RUS is highly conserved between mouse and human. The presented rescue experiments show that the overexpression of full length, the $\triangle$3' but not the $\triangle$5' RUS construct, rescued the observed RUS' neuronal KD phenotype. Thus, we considered

the conserved 5' domain to be functionally relevant and act as a binding domain for gene-regulatory proteins. RUS is localized in the nucleus and acts on the chromatin. Therefore, we purified interacting proteins from the nuclear extract using the MS2 phage aptamer as an RNA-affinity tag. Instead of transcribing the RNA *in vitro*, I overexpressed the tagged RNA in stable neuron-related Neuro2A cells to circumvent folding artifacts. Together with our in-house proteomic research group of Prof. Dr. Stefan Lichtentahler, we quantified and sequenced the RNAse A eluted proteins by LC-MS. Although several studies overexpress MS2 tagged RNAs in mammalian cells to study RNA-interacting proteins, our methodology is distinct and not described elsewhere. Overexpression of the MS2 coat protein may produce site artifacts by mislocalization of the MS2 tagged RNA. Compared to other studies, we preincubated the nuclear lysate with the MS2 binding protein instead of overexpressing it [129]. To ensure a similar abundance of overexpressed constructs, we generated stable MS2-tagged RNA overexpressing cell lines.

To unravel relevant interacting proteins, we focused on the conserved 5' region. Therefore, we measured the interactome of full-length RUS, the △5' construct, and the 5' conserved domain itself. Statistically, relevant proteins should be enriched by comparing the binding proteins of the full-length and △5' construct and comparing the binding protein of the 5' conserved domain and △5' construct. So far, only NAT10, Hist1h4, Dld1, and Smarac5 were significantly enriched in both comparisons. NAT10 is involved in rRNA processing [166]. Dld is involved in the succinylation of lysine 79 in histone 3 [182]. Thus, only the ISWI chromatin remodeler Smarca5 represented an interesting protein. Other interesting proteins such as Brd2, Brd4, and Lbr were only enriched by comparing the interactome of the full-length and △5'. Lbr was completely absent in the △5' interactome. Thus, potential long-range RNA structures are required for the binding of Brd2, Brd4, and Lbr. Overexpression of △3' is still rescuing the KD phenotype. This mutant still includes exon 2 and 3 that might be relevant for the binding of Lbr, Brd2, and Brd4. Several studies already described the interaction of Smarca5 and Lbr to lncRNAs. Besides, it is known that lncRNAs modulate Smarca5's remodeling activity by its RNA helicase [2, 68]). So far, the interaction of Bet proteins with RNAs is described only for enhancer RNAs (eRNAs) [83]. eRNAs are transcribed from enhancer regions and to recruit Brd4 and the histone acetylates CBP-p300 locally to their transcription site [83, 84]. However, the interaction of Brd2 and Brd4 with *trans* acting lncRNAs is not described else. Thus, we supposed the interaction of lncRNAs with Brd2 and Brd4 represents a novel and undescribed feature for lncRNA - especially for those regulating enhancer-promoter interactions. RUS exhibits at the 3' end long tracts enriched for uridines and cytosines. Compared to the 5'domain, the △5' construct enriched for proteins binding to poly uridines. This result ensured the robustness of our methodology.

All interactions were confirmed reciprocally by RIP experiments and by western blot analysis. We repeated the affinity purifications for western blot analysis and eluted the proteins using amylose as an additional control. Using western blot analysis, I fig-

ured out that RUS's interaction with Brd2 and Brd4 depends on their extra terminal domain. It is known that this domain is required for protein-protein interactions at promoter enhancer loops but not for the interaction with RNAs. So far, Brd2, Brd4, and Smarca5 are not described to interact with each other. Thus, it remains elusive whether RUS sequesters them in the same RNP or whether the interaction to those proteins is dynamic.

7.2.2 RUS binding on chromatin

To identify RUS binding sites, we performed together with the EMBL Gene Core facility, ChIRP analysis against RUS in Wt cells and ChIP analysis against Brd2 in ctrl and RUS KD cells. Whereby Brd2 binds more than 4000 times in the genome, RUS binds only 39 times. We identified a repetitive motif with a p-value of 1.8 e-32 for all RUS-bound regions using motif enrichment analysis.
By comparing the regions bound by RUS and Brd2, we identified that Brd2 and RUS bound together on 13 genomic sites. Thus, more than 30% of RUS sites are bound by Brd2. Using motif enrichment analysis for Brd2 and RUS bound regions, I determined a high confident 29 nt long motif containing 6 x GAATG repeats with a p-value of 1e-132. By predicting the repetitive GAATG motif in the mouse genome, I figured out that Brd2 bound 31 of 41 genomic motif sites. Hence, 42% from those Brd2 occupied sites are targeted by RUS. Only one repetitive site in the genome was bound by RUS and not by Brd2. Thus, I suggested that the repetitive regions and the binding of Brd2 to those are determinants for RUS binding. Moreover, the observed improvement of the p-value during motif enrichment analysis suggests that RUS binds beyond the repetitive regions on Brd2 unbound sites. Thus, RUS acts on two or more distinct regulatory pathways comprising one Brd2-dependent on which I focused.

I anticipated that RUS recruited Brd2 to its target site. However, the KD of RUS did not affect the binding of Brd2 to those sites. Thus, Brd2 recruits RUS to its target sites and not vice versa. Still elusive remains the specificity towards the GAATG tandem repeat. Frequently, proteins are involved in the binding of lncRNAs to their respective target sites as hnRNP-U. Although hnRNP-U appears in our LC-MS data, it was not enriched by comparing the interactome of the full-length and $\triangle$5' construct. Besides, hnRNP-U was not enriched by comparing the interactomes of the $\triangle$5' construct and the 5' domain. Thus, I cannot conclude that hnRNP-U is binding to RUS. Motif enrichment analysis determined the enhancer factor TEAD1 as potential RUS' genomic binding-factor. However, TEAD1 was absent in our LC-MS data and can be excluded. We know that tandem repeats are transcriptionally active from the literature and are transcribed into enhancer RNAs [175]. Besides, tandem repeats are prone to form RNA-DNA triplexes [174]. These facts raise the two hypotheses: (1) RUS forms an RNA-DNA triplex with the tandem repeats, or (2) RUS base pairs the potential eRNA expressed from the repeat region. In cases of intronic repeat regions, the premature mRNA, such

as Dpp9 by itself, may act as an RNA substrate for RUS.
To test hypothesis 1, I predicted potential RNA-DNA triplexes between RUS and the repeat domain *in silico*. Using the Triplexator software yielded no sites forming an RNA-DNA triplex with the repeat domain in the matured RUS transcript. Surprisingly, the premature and unspliced murine RUS transcript contains in intron 1 a pyrimidine rich parallel forming triplex site. Controversially, the human RUS contains two antiparallel potential purine-rich triplex-forming sites. We studied the triplex formation by performing EMSA with fluorescent-labeled oligonucleotides. Parallel triplex formation requires the acidification of cytosine. However, we observed that acidic pH decreased the stability of the DNA duplex. Triplex formation under ambient pH-value resulted in a weak triplex.
To test hypothesis 2, we annealed the same fluorescent-labeled RNA strand with the potential expressed GAAUG repeat. Performing EMSA confirmed that both RNAs are forming a stable RNA-RNA duplex. However, the human premature RUS contains an antiparallel triplex-forming site preferential anneal to the CAUUC repeat, the reverse complementary repeat of GAAUG. In contrast to mice, humans must express the reverse complementary repeat consequently.
Both hypotheses would explain that lncRNAs might sense its targets site during transcription. After splicing, the chromatin sensing site is removed from the RUS. Consequently, RUS stays bound by interacting with Brd2 bound at chromatin. The binding of a lncRNA on chromatin is frequently explained by interaction with chromatin-associated proteins, although the protein binds to more sites than the lncRNA does. Like RNA-DNA triplex formation of matured lncRNAs, co-transcriptional substrate sensing, as I suggested for RUS, could explain the specificity towards a DNA consensus motif that is not defined for a gene-regulatory protein. Although further experiments like the mutation of the potential sensing site must be performed for full approval, our observations yielded the first biochemical evidence for co-transcriptional substrate sensing. Moreover, the removal of introns during splicing could bring the target sites close to the RUS's transcription site. In this fashion, lncRNA might influence the nuclear 3-dimensional organization.
Nevertheless, it is suggested that repeats are spatial clustering in the nucleus [92, 183]. Accordingly, RUS's target regions may be in proximity to its transcription locus. Thus RUS would bind to chromatin according to the well-described proximity and affinity guided mechanisms [109].
Vice versa, the transcribed repeat RNAs may attach to the transcriptional locus of RUS by forming RNA-DNA triplexes. However, eRNAs are short and un-polyadenylated RNA-Pol II transcripts and supposed to be short-lived RNAs. Thus eRNAs act primarily in close vicinity to its transcription site [83, 84].
Moreover, studying eRNA triplexes' formation in shaping the three-dimensional nuclear structure is only notional, was out of this doctoral thesis's scope, and not performed.

Both scenarios: co-transcriptional substrate sensing or proximity and affinity-guided

chromatin binding, suggest that the rescue experiments' results may be controversial. How can the overexpression from the non-endogenous locus of an unspliced RUS missing intronic triplex-forming site rescue the neuronal KD phenotype? As a potential explanation, we have to consider that RUS was tremendously overexpressed. The increased level and Brd2 binding capacity potentially enabled the exogenous expressed RUS to bind to the endogenous RUS target sites.

7.2.3 RUS's action on chromatin

I selected from the 39 RUS bound sites the top 20 highest confident peaks. From those, RUS's KD downregulates the three genes: Pou4f1, Cdh7, and Ppc6c and upregulates the five genes significantly: Arid1b, Dpp9, Gdnf, Kcna5, and Ptprm. Genes as Bin1 and Ncam1 were reduced only in one KD condition. Intriguingly, most of the genes located next to RUS binding sites exert important neuronal functions.
Kcna5 is a potassium voltage-gated channel and has known important functions in neurons [184]. The neural adhesion molecules Ncam1, the cadherin Cdh7, and the tyrosine phosphatase-receptor Ptprm1 are required for cell-cell adhesion and signaling [185–187]. Pp6c is a protein phosphatase regulating cell cycle progression [188]. The glia-derived neurotrophic factor Gdnf1 regulates the expansion of neural stem cells and neurons' survival [189]. Arid1b is a subunit of BAF remodeling complex subunit and exerts key-neurodevelopmental gene-regulatory functions [190]. Pou4f1 is neuron-specific transcription factor and protects sensory neurons from apoptosis [191].
During ongoing analyses, I focused on the 13 sites bound by Brd2 and RUS. None of them are located in promoter regions. However, three sites are located next to genes: Pou4f1, Dpp9, and Arid1b; those expressions' were affected by RUS's KD. Overall, RUS depletion resulted in a reduced expression of genes located next to RUS's target sites. Thus, I anticipated that RUS has a repressive function on gene expression rather.
Several chromatin remodelers are repressed by ncRNA. pRNAs repress Smarca5 in the NoRC complex to silence rRNA genes, and the lncRNA Evf2 represses Brg1 on the conserved enhancer in the Dlx5/6 locus [2, 3]. Both remodellers contain a DNA and RNA helicase domain. That suggests that RNAs act as a competing substrate for chromatin remodeling. Chromatin remodelers are not acting sequence-specific. Thus, chromatin remodeling repression by lncRNAs would present an adequate mechanism for site-specific chromatin remodeling regulation.
Brd2 cooperates with CTCF for enhancer-promoter looping and maintaining insulation boundaries [4, 80]. Besides, Smarca5 regulates CTCF deposition and nucleosome spacing adjacent to CTCF sites [43, 76]. To prove whether RUS also represses the activity of Smarca5, I focused on the interplay between Smarca5 and CTCF. If RUS represses Smarca5 activity, RUS targets sites should show reduced CTCF occupancy in neuronal tissues. Using publicly available ChIP data against CTCF in ESC, NSC, cortical -, and hippocampal-neurons showed, however, that RUS bound on CTCF sites that are

distinct for the hippocampus where RUS is at the highest expressed. This observation indicated that RUS rather activates Smarca5's chromatin remodeling activity. Thus, we performed ChIP experiments against CTCF and Smarca5 combined with quantitative RT-PCR using primers against the 3 RUS target sites next to the genes: Arid1b, Dpp9, and Pouf4f1 and rs14 as control CTCF occupied region in ctrl, RUS, and Smarca5 KD cells. As expected, Smarca5 depletion reduced Smarca5 and consequently CTCF occupancy at all tested regions. RUS's depletion selectively reduced the CTCF occupancy at RUS target sites. The level of Smarca5 remained unaffected by RUS's KD. These observations confirm that RUS activates Smarca5's chromatin remodeling activity without affecting Smarca5's occupancy.

Hence, I hypothesize that RUS binds to GAATG repetitive genomic regions bound by Brd2 that are also bound but not remodeled by Smarca5. Thus, the CTCF binding site remains wrapped around the nucleosome and inaccessible for CTCF. Either RUS may bind proximity and affinity guided to its target sites potentially located in vicinity to RUS's transcription locus or may bind co-transcriptionally via the intronic triplex forming site (TFO). Thereby, the strong RNA-RNA duplex formation of the RNA produced from the repetitive region and the intronic TFO seemed to be most relevant for co-transcriptional chromatin binding. The binding of RUS to its target regions activates Smarca5's chromatin remodeling. Smarca5 slides nucleosomes around the target site to liberate the CTCF binding site and facilitate CTCF binding. Elusive remains, why Smarca5 binds but not remodels to RUS target sites. However, the potential RNA expressed from the repetitive regions may inhibit Smarca5 activity in *cis*. After binding of RUS, Smarca5's activity is induced in *trans*.

Although further biochemical validations are required, our observation showed that RUS represents a lncRNA that activates and does not repress Smarca5's remodeling activity to deposit CTCF site-specific. In this fashion, chromatin remodeling activation by lncRNAs represents a novel unknown and versatile mechanism for spatiotemporal chromatin remodeling and chromatin shaping.

Consequently, the genes as Arid1, Dpp9, and Pou4f1 adjacent to the target sites are supposed to be reduced or increased during hippocampal differentiation. Since RUS interacts with Brd2 and Brd4, I reasoned that RUS is involved in enhancer-promoter looping, which should be manifested by a decrease of gene expression of the genes located next to RUS's target sites after RUS depletion. That was only the case for Pou4f1. During neuronal differentiation, tumor suppressor p53's expression is elevated and determines the neuronal fate and neurons' susceptibility to apoptosis. Pou4f1 antagonizes p53 mediated expression of the pro-apoptotic factor Noxa and Bax and protects sensory neurons against apoptosis [192]. Nevertheless, Brd2 and CTCF also act together as chromatin insulator [4]. In this fashion, RUS reduces the expression next to RUS's target sites as observed for Arid1b and Dpp9 that were significantly upregulated after RUS's KD. Conversely, the expression Arid1b regulating cell cycle progression in NSC is still maintained in neurons [190]. Moreover, Arid1b represses Wnt signaling required for neuronal differentiation in intermediate progenitors [193, 194]. So far, the literature did

not show whether Arid1 is absent in a particular neuronal cell type responding to Wnt signaling. That would indicate RUS acts in a particular neuronal cell type responding to Wnt signaling to repress Arid1b. Furthermore, the role of Dpp9 in neural stem cells is not described. Dpp9 represses EGF-mediated PI3K/Akt signaling, regulates migration, cell -adhesion, and induces apoptosis by activating Caspase 9 [195]. Thus, repressed Wnt signaling by increased Arid1b expression, the increased Dpp9 expression, and the decreased Pou4f1 expression by RUS's KD would explain that RUS KD cells are arrested in differentiation and undergo apoptosis.
Thus, RUS regulates the cell-type-specific chromatin shaping and organization by regulating CTCF deposition instead of directly affecting gene-transcription. Nevertheless, a misshaped and disorganized chromatin structure by altered CTCF occupancy produced by RUS's KD may also result in the arrest of differentiation and induces apoptosis in differentiating neurons.
Elusive remains the role of Lbr that specifically binds on the 5'conserved domain of RUS during RUS mediated chromatin reorganization. Either Lbr could be required to tether RUS target sites to the nuclear envelope as observed for Xist, or RUS could use Lbr to bind to its target sites [68].

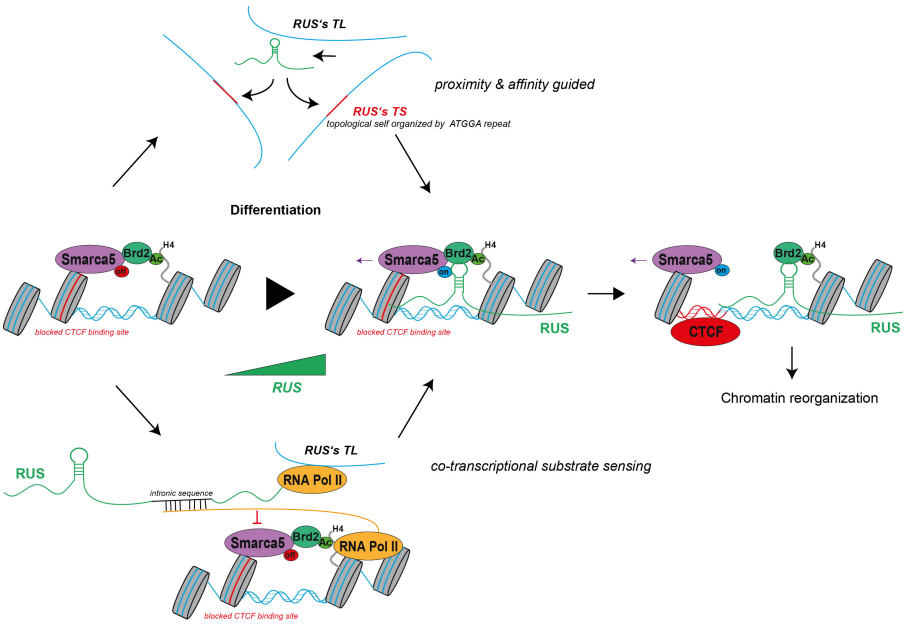


Figure 47: **A proposed model of acting:** RUS binds to Brd2 occupied GAATG repetitive regions and activates the remodeling activity of Smarca5 to induces CTCF binding and chromatin reorganization. Thereby RUS might sense its chromatin substrates either by proximity and affinity-guided mechanisms (top) or co-transcriptionally (bottom). TS: target sites, TL: transcription locus

7.3 Outlook

Ultimately, the presented doctoral study provides us novel insights into lncRNA biology and helps us understand the molecular function of lncRNAs. Although the presented results shed light on many points of the molecular mechanism of RUS, more questions come up and have to be addressed.
From our observation, I conclude that RUS is required during embryonal development for proper neural stem cell function and neuronal differentiation. Still elusive remains whether RUS acts in all neuronal cell types or only in one distinct. Thus, FISH experiments in whole mouse embryos and adult mouse brains could be performed to determine the developmental stage, the regiospecific, and the neuronal subtype-specific expression of RUS. Consequently, defined differentiation protocols of stem cells into RUS expressing neuronal linage can be applied to study RUS's KD's effect on the cellular level in more detail. Furthermore, novel CRISPR-Cas9 editing tools can be applied to generated RUS knockout (KO) mouse lines to decipher the function of RUS during central nervous system development [196]. The questions of whether RUS KO is embryonic lethal, result in an anatomical change, or depletion of the RUS's expressing cell type can be addressed.
To understand the molecular mechanism of RUS, the activating effect of RUS on Smarca5, I suggested, has to be validated. Therefore, biochemical and functional experiments can be performed. One approach represents nucleosome mobility assay with purified components [3, 197]. Therefore, nucleosomes are reconstituted and incubated with Smarca5 in the presence of ATP. Smarca5 remodels the DNA wrapped around the nucleosome resulting in a shift on a native polyacrylamide gel. Activation of Smarca5 by *in vitro* transcribed RUS should increase the shift.
Smarca5 liberates CTCF binding sites by sliding nucleosomes, also known as nucleosome phasing, which can be probed by DNAse I sensitive assays [198]. In general, nucleosomes protect DNA against cleavage by DNAse I. The flexible linker DNA between nucleosome is prone to get degraded by DNAse I. Thus CTCF bound sites are prone to get cleaved. The same primers used to evaluate ChiP experiments targeting RUS's target sites and rs14 can be used to analyze by quantitative RT-PCR whether the target sites are cleaved by DNAse I or protected against DNAse I. Similar to performed ChIP experiments, DNAse sensitive assays have to be performed in ctrl, RUS, and Smarca5 KD cells to investigate Smarca5 and RUS's effect on DNAse sensitivity. I would expect that in Wt and ctrl KD cells, the sites are unprotected against DNAse-I. Smarca5 KD would be manifested by increased global nucleosome occupancy and decreased DNAse sensitivity. Knocking down RUS activating Smarca5 would show the same effect selectively on its target sites but not on the control region.
To understand the chromatin binding of RUS, the predicted sensing domain in the premature RUS transcript has to removed using genome editing tools as CRISPR-Cas 9 in isolated neural stem cells [196]. Performing ChIRP against RUS analyzed by quantitative RT-PCR would reveal whether the identified target sites' purification is diminished in edited cells. If so, additional KD experiments of potential eRNAs expressed from

repeat regions can be performed to test whether RUS still binds to the respective repeat region by using ChIRP against RUS analyzed by quantitative RT-PCR. In this fashion, we can distinguish whether RUS senses its substrate co-transcriptional by forming RNA-DNA triplexes with the GAATG repeat regions or base pair with the RNAs potential transcribed from those regions.
In addition to editing the sensing domain, the conserved 5' domain can be removed. Novel visualization allows for the staining of RNA and particular DNA regions in living cells using fluorescent-labeled Cas 13 and Cas 9 proteins, respectively [199]. Together with fluorescent-tagged Brd2, Brd4, Smarca5, and Lbr, the interplay of RUS with those factors at its target sites can be visualized. Especially, the interplay of RUS and Lbr could unravel whether Lbr tethers RUS's target sites to the nuclear envelope. That can be done in Wt and edited cells to dissect the edited RNA-domains' molecular function.

8 Bibliography

References

[1] B. S. Clark and S. Blackshaw, "Understanding the Role of lncRNAs in Nervous System Development," *Adv Exp Med Biol*, vol. 1008, pp. 253–282, 2017.

[2] L. Manelyte, R. Strohner, T. Gross, and G. Längst, "Chromatin Targeting Signals, Nucleosome Positioning Mechanism and Non-Coding RNA-Mediated Regulation of the Chromatin Remodeling Complex NoRC," *PLoS Genetics*, vol. 10, no. 3, 2014.

[3] I. Cajigas, D. E. Leib, J. Cochrane, H. Luo, K. Swyter, S. Chen, B. S. Clark, J. Thompson, J. R. Yates, R. E. Kingston, and J. D. Kohtz, "Evf2 lncRNA/BRG1/DLX1 interactions reveal RNA-dependent chromatin remodeling inhibition.," *Development (Cambridge, England)*, vol. 170, pp. dev.126318–, 2015.

[4] S. C. Hsu, T. G. Gilgenast, C. R. Bartman, C. R. Edwards, A. J. Stonestrom, P. Huang, D. J. Emerson, P. Evans, M. T. Werner, C. A. Keller, B. Giardine, R. C. Hardison, A. Raj, J. E. Phillips-Cremins, and G. A. Blobel, "The BET Protein BRD2 Cooperates with CTCF to Enforce Transcriptional and Architectural Boundaries," *Molecular Cell*, vol. 66, no. 1, pp. 102–116.e7, 2017.

[5] B. J. GORDON, D. PETER, J. EDDIE, O. K. JODY E. JOHNSON, K. DEAN, P. BRANDON, A. W. JAMES, W. MARK, and Y. KELLY A., *Anatomy and Physiology*. OpenStax, 2013.

[6] A. P. Barnes and F. Polleux, "Establishment of axon-dendrite polarity in developing neurons," *Annual Review of Neuroscience*, vol. 32, pp. 347–381, 2009.

[7] A. M. Sousa, K. A. Meyer, G. Santpere, F. O. Gulden, and N. Sestan, "Evolution of the human nervous system function, structure, and development," *Cell*, vol. 170, pp. 226–247, July 2017.

[8] T. Sadler, "Embryology of neural tube development," *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, vol. 135C, no. 1, pp. 2–8, 2005.

[9] M. Götz and W. B. Huttner, "The cell biology of neurogenesis," *Nature Reviews Molecular Cell Biology*, vol. 6, pp. 777–788, Oct. 2005.

[10] L. C. Greig, M. B. Woodworth, M. J. Galazo, H. Padmanabhan, and J. D. Macklis, "Molecular logic of neocortical projection neuron specification, development and diversity," *Nature Reviews Neuroscience*, vol. 14, pp. 755–769, Oct. 2013.

[11] E. Hartfuss, R. Galli, N. Heins, and M. Go, "Characterization of CNS Precursor Subtypes and Radial Glia," vol. 30, pp. 15–30, 2001.

[12] P. Malatesta, M. A. Hack, E. Hartfuss, H. Kettenmann, W. Klinkert, F. Kirchhoff, and M. Götz, "Neuronal or glial progeny: Regional differences in radial glia fate," *Neuron*, vol. 37, no. 5, pp. 751–764, 2003.

[13] D. S. Rice and T. Curran, "Role of the reelin signaling pathway in central nervous system development," *Annual Review of Neuroscience*, vol. 24, pp. 1005–1039, Mar. 2001.

[14] M. Nieto, E. S. Monuki, H. Tang, J. Imitola, N. Haubst, S. J. Khoury, J. Cunningham, M. Gotz, and C. A. Walsh, "Expression of Cux-1 and Cux-2 in the subventricular zone and upper layers II-IV of the cerebral cortex," *Journal of Comparative Neurology*, vol. 479, no. 2, pp. 168–180, 2004.

[15] R. W. Oppenheim, "The neurotrophic theory and naturally occurring motoneuron death," *Trends in Neurosciences*, vol. 12, pp. 252–255, Jan. 1989.

[16] K. Takahashi and S. Yamanaka, "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors," *Cell*, vol. 126, pp. 663–676, aug 2006.

[17] E. G. Bernstine, M. L. Hooper, S. Grandchamp, and B. Ephrussi, "Alkaline phosphatase activity in mouse teratoma," *Proceedings of the National Academy of Sciences*, vol. 70, pp. 3899–3903, Dec. 1973.

[18] A. Leahy, J.-W. Xiong, F. Kuhnert, and H. Stuhlmann, "Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation," *Journal of Experimental Zoology*, vol. 284, pp. 67–81, June 1999.

[19] J.-H. Chuang, " Neural differentiation from embryonic stem cells in vitro : An overview of the signaling pathways ," *World Journal of Stem Cells*, vol. 7, no. 2, p. 437, 2015.

[20] E. Arenas, M. Denham, and J. C. Villaescusa, "How to make a midbrain dopaminergic neuron," *Development (Cambridge)*, vol. 142, no. 11, pp. 1918–1936, 2015.

[21] B. a. Reynolds and S. Weiss, "Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system.," *Science*, vol. 255, no. 5052, pp. 1707–10, 1992.

[22] G. A. Banker and W. M. Cowan, "Rat hippocampal neurons in dispersed cell culture," *Brain Research*, vol. 126, no. 3, pp. 397–425, 1977.

[23] J. Watson, N. Hopkins, J. Roberts, J. Steitz, and A. Weiner, *Molecular Biology of the Gene*. No. Bd. 1 in Molecular Biology of the Gene, Benjamin/Cummings Publishing Company, 1988.

[24] J. Soutourina, "Transcription regulation by the mediator complex," *Nature Reviews Molecular Cell Biology*, vol. 19, pp. 262–274, Dec. 2017.

[25] S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, and M. T. Weirauch, "The human transcription factors," *Cell*, vol. 172, pp. 650–665, Feb. 2018.

[26] W. W. Wasserman and A. Sandelin, "Applied bioinformatics for the identification of regulatory elements," *Nature Reviews Genetics*, vol. 5, pp. 276–287, Apr. 2004.

[27] H. J. Baek, Y. K. Kang, and R. G. Roeder, "Human mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly," *Journal of Biological Chemistry*, vol. 281, pp. 15172–15181, June 2006.

[28] F. Eyboulet, S. Wydau-Dematteis, T. Eychenne, O. Alibert, H. Neil, C. Boschiero, M.-C. Nevers, H. Volland, D. Cornu, V. Redeker, M. Werner, and J. Soutourina, "Mediator independently orchestrates multiple steps of preinitiation complex assemblyin vivo," *Nucleic Acids Research*, vol. 43, pp. 9214–9231, Aug. 2015.

[29] A. C. Schier and D. J. Taatjes, "Structure and mechanism of the RNA polymerase II transcription machinery," *Genes & Development*, vol. 34, pp. 465–488, Apr. 2020.

[30] D. S. Luse, "The RNA polymerase II preinitiation complex," *Transcription*, vol. 5, p. e27050, Nov. 2013.

[31] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 Å resolution," *Nature*, vol. 389, no. 6648, pp. 251–260, 1997.

[32] R. D. Kornberg, "Chromatin structure: A repeating unit of histones and DNA," *Science*, vol. 184, pp. 868–871, May 1974.

[33] C. L. van Emmerik and H. van Ingen, "Unspinning chromatin: Revealing the dynamic nucleosome landscape by NMR," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 110, no. April, pp. 1–19, 2019.

[34] J. A. West, C. P. Davis, H. Sunwoo, M. D. Simon, R. I. Sadreyev, P. I. Wang, M. Y. Tolstorukov, and R. E. Kingston, "The Long Noncoding RNAs NEAT1 and MALAT1 Bind Active Chromatin Sites," *Molecular Cell*, vol. 55, no. 5, pp. 791–802, 2014.

[35] T. C. James and S. C. Elgin, "Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene.," *Molecular and Cellular Biology*, vol. 6, no. 11, pp. 3862–3872, 1986.

[36] D. Canzio, E. Y. Chang, S. Shankar, K. M. Kuchenbecker, M. D. Simon, H. D. Madhani, G. J. Narlikar, and B. Al-Sady, "Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly," *Molecular Cell*, vol. 41, no. 1, pp. 67–81, 2011.

[37] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, "Formation of Chromosomal Domains by Loop Extrusion," *Cell Reports*, vol. 15, no. 9, pp. 2038–2049, 2016.

[38] H. Zheng and W. Xie, "The role of 3D genome organization in development and cell differentiation," *Nature Reviews Molecular Cell Biology*, vol. 20, no. 9, pp. 535–550, 2019.

[39] I. F. Davidson, B. Bauer, D. Goetz, W. Tang, G. Wutz, and J.-M. Peters, "DNA loop extrusion by human cohesin," *Science*, vol. 366, pp. 1338–1345, Nov. 2019.

[40] C.-T. Ong and V. G. Corces, "CTCF: an architectural protein bridging genome topology and function," *Nature Reviews Genetics*, vol. 15, pp. 234–246, Mar. 2014.

[41] E. de Wit, B. A. M. Bouwman, Y. Zhu, P. Klous, E. Splinter, M. J. A. M. Verstegen, P. H. L. Krijger, N. Festuccia, E. P. Nora, M. Welling, E. Heard, N. Geijsen, R. A. Poot, I. Chambers, and W. de Laat, "The pluripotent genome in three dimensions is shaped around pluripotency factors," *Nature*, vol. 501, pp. 227–231, July 2013.

[42] S. Schoenfelder, M. Furlan-Magaril, B. Mifsud, F. Tavares-Cadete, R. Sugar, B. M. Javierre, T. Nagano, Y. Katsman, M. Sakthidevi, S. W. Wingett, E. Dimitrova, A. Dimond, L. B. Edelman, S. Elderkin, K. Tabbada, E. Darbo, S. Andrews, B. Herman, A. Higgs, E. LeProust, C. S. Osborne, J. A. Mitchell, N. M. Luscombe, and P. Fraser, "The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements," *Genome Research*, vol. 25, no. 4, pp. 582–597, 2015.

[43] N. Wiechens, V. Singh, T. Gkikopoulos, P. Schofield, S. Rocha, and T. Owen-Hughes, "The chromatin remodelling enzymes snf2h and snf2l position nucleosomes adjacent to ctcf and other transcription factors," *PLOS Genetics*, vol. 12, pp. 1–25, 03 2016.

[44] Y. Atlasi and H. G. Stunnenberg, "The interplay of epigenetic marks during stem cell differentiation and development," *Nature Reviews Genetics*, vol. 18, no. 11, pp. 643–658, 2017.

[45] V. N. Budhavarapu, M. Chavez, and J. K. Tyler, "How is epigenetic information maintained through DNA replication?," *Epigenetics & Chromatin*, vol. 6, no. 1, p. 32, 2013.

[46] M. V. C. Greenberg and D. Bourc'his, "The diverse roles of DNA methylation in mammalian development and disease," *Nature Reviews Molecular Cell Biology*, vol. 20, pp. 590–607, Aug. 2019.

[47] C. A. Musselman, M.-E. Lalonde, J. Côté, and T. G. Kutateladze, "Perceiving the epigenetic landscape through histone readers," *Nature Structural & Molecular Biology*, vol. 19, pp. 1218–1227, Dec. 2012.

[48] H. Santos-Rosa, R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor, and T. Kouzarides, "Active genes are tri-methylated at K4 of histone H3," *Nature*, vol. 419, no. 6905, pp. 407–411, 2002.

[49] C. C. Sze, P. A. Ozark, K. Cao, M. Ugarenko, S. Das, L. Wang, S. A. Marshall, E. J. Rendleman, C. A. Ryan, D. Zha, D. Douillet, F. X. Chen, and A. Shilatifard, "Coordinated regulation of cellular identity–associated H3K4me3 breadth by the COMPASS family," *Science Advances*, vol. 6, no. 26, pp. 1–12, 2020.

[50] J. F. Flanagan, L. Z. Mi, M. Chruszcz, M. Cymborowski, K. L. Clines, Y. Kim, W. Minor, F. Rastinejad, and S. Khorasanizadeh, "Double chromodomains cooperate to recognize the methylated histone H3 tail," *Nature*, vol. 438, no. 7071, pp. 1181–1185, 2005.

[51] Y. Shi, F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero, and Y. Shi, "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1," *Cell*, vol. 119, no. 7, pp. 941–953, 2004.

[52] H. Shen, W. Xu, and F. Lan, "Histone lysine demethylases in mammalian embryonic development," *Experimental and Molecular Medicine*, vol. 49, no. 4, pp. 1–7, 2017.

[53] Y. Liu, S. Qin, T. Y. Chen, M. Lei, S. S. Dhar, J. C. Ho, A. Dong, P. Loppnau, Y. Li, M. G. Lee, and J. Min, "Structural insights into trans-histone regulation of H3K4 methylation by unique histone H4 binding of MLL3/4," *Nature Communications*, vol. 10, no. 1, 2019.

[54] J. Cheng, R. Blum, C. Bowman, D. Hu, A. Shilatifard, S. Shen, and B. D. Dynlacht, "A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers," *Molecular Cell*, vol. 53, no. 6, pp. 979–992, 2014.

[55] R. Raisner, S. Kharbanda, L. Jin, E. Jeng, E. Chan, M. Merchant, P. M. Haverty, R. Bainer, T. Cheung, D. Arnott, E. M. Flynn, F. A. Romero, S. Magnuson, and K. E. Gascoigne, "Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation," *Cell Reports*, vol. 24, no. 7, pp. 1722–1729, 2018.

[56] Y. Li, J. Syed, and H. Sugiyama, "RNA-DNA Triplex Formation by Long Noncoding RNAs," *Cell Chemical Biology*, pp. 1–9, 2016.

[57] S. Cuddapah, R. Jothi, D. E. Schones, T. Y. Roh, K. Cui, and K. Zhao, "Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains," *Genome Research*, vol. 19, no. 1, pp. 24–32, 2009.

[58] H. Marks, J. C. Chow, S. Denissov, K. J. Françoijs, N. Brockdorff, E. Heard, and H. G. Stunnenberg, "High-resolution analysis of epigenetic changes associated with X inactivation," *Genome Research*, vol. 19, no. 8, pp. 1361–1373, 2009.

[59] K. Kurimoto, Y. Yabuta, K. Hayashi, H. Ohta, H. Kiyonari, T. Mitani, Y. Moritoki, K. Kohri, H. Kimura, T. Yamamoto, Y. Katou, K. Shirahige, and M. Saitou, "Quantitative dynamics of chromatin remodeling during germ cell specification from mouse embryonic stem cells," *Cell Stem Cell*, vol. 16, no. 5, pp. 517–532, 2015.

[60] K. H. Hansen, A. P. Bracken, D. Pasini, N. Dietrich, S. S. Gehani, A. Monrad, J. Rappsilber, M. Lerdrup, and K. Helin, "A model for transmission of the h3k27me3 epigenetic mark," *Nature Cell Biology*, vol. 10, pp. 1291–1300, Oct. 2008.

[61] H. Wu, V. Coskun, J. Tao, W. Xie, W. Ge, K. Yoshikawa, E. Li, Y. Zhang, and Y. E. Sun, "Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes," *Science*, vol. 329, no. 5990, pp. 444–447, 2010.

[62] M. Tachibana, K. Sugimoto, T. Fukushima, and Y. Shinkai, "SET Domain-containing Protein, G9a, is a Novel Lysine-preferring Mammalian Histone Methyltransferase with Hyperactivity and Specific Selectivity to Lysines 9 and 27 of Histone H3," *Journal of Biological Chemistry*, vol. 276, no. 27, pp. 25309–25317, 2001.

[63] N. Fujita, S. Watanabe, T. Ichimura, S. Tsuruzoe, Y. Shinkai, M. Tachibana, T. Chiba, and M. Nakao, "Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression," *Journal of Biological Chemistry*, vol. 278, no. 26, pp. 24132–24138, 2003.

[64] X. Liu, Q. Gao, P. Li, Q. Zhao, J. Zhang, J. Li, H. Koseki, and J. Wong, "UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9," *Nature Communications*, vol. 4, 2013.

[65] T. Yadav and I. Whitehouse, "Replication-Coupled Nucleosome Assembly and Positioning by ATP-Dependent Chromatin-Remodeling Enzymes," *Cell Reports*, vol. 15, no. 4, pp. 715–723, 2016.

[66] A. Sabantsev, R. F. Levendosky, X. Zhuang, G. D. Bowman, and S. Deindl, "Direct observation of coordinated DNA movements on the nucleosome during chromatin remodelling," *Nature Communications*, vol. 10, no. 1, pp. 1–12, 2019.

[67] Y. Zhang, H. H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird, and D. Reinberg, "Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation," *Genes and Development*, vol. 13, no. 15, pp. 1924–1935, 1999.

[68] J. T. L. Anand Minajigi, John E. Froberg, Chunyao Wei, Hongjae Sunwoo, Barry Kesner, David Colognori, Derek Lessing, Bernhard Payer, Myriam Boukhali, Wilhelm Haas, "A comprehensive Xist interactome reveals cohesin repulsion and an RNA- directed chromosome conformation Anand," *Science*, vol. 349, no. 6245, pp. 282–294, 2015.

[69] J. Lessard, J. I. Wu, J. A. Ranish, M. Wan, M. M. Winslow, B. T. Staahl, H. Wu, R. Aebersold, I. A. Graef, and G. R. Crabtree, "An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development," *Neuron*, vol. 55, no. 2, pp. 201–215, 2007.

[70] X. Zhan, X. Shi, Z. Zhang, Y. Chen, and J. I. Wu, "Dual role of Brg chromatin remodeling factor in Sonic hedgehog signaling during neural development," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 31, pp. 12758–12763, 2011.

[71] A. Alfert, N. Moreno, and K. Kerl, "The BAF complex in development and disease," *Epigenetics & Chromatin*, vol. 12, Mar. 2019.

[72] S. He, Z. Wu, Y. Tian, Z. Yu, J. Yu, X. Wang, J. Li, B. Liu, and Y. Xu, "Structure of nucleosome-bound human BAF complex," *Science*, vol. 367, pp. 875–881, Jan. 2020.

[73] G. P. Dann, G. P. Liszczak, J. D. Bagert, M. M. Müller, U. T. Nguyen, F. Wojcik, Z. Z. Brown, J. Bos, T. Panchenko, R. Pihl, S. B. Pollock, K. L. Diehl, C. D. Allis, and T. W. Muir, "ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference," *Nature*, vol. 548, no. 7669, pp. 607–611, 2017.

[74] R. Santoro and I. Grummt, "Epigenetic Mechanism of rRNA Gene Silencing: Temporal Order of NoRC-Mediated Histone Modification, Chromatin Remodeling, and DNA Methylation," *Molecular and Cellular Biology*, vol. 25, no. 7, pp. 2539–2546, 2005.

[75] Ö. Z. Aydin, W. Vermeulen, and H. Lans, "ISWI chromatin remodeling complexes in the DNA damage response," *Cell Cycle*, vol. 13, no. 19, pp. 3016–3025, 2014.

[76] D. Barisic, M. B. Stadler, M. Iurlaro, and D. Schübeler, "Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors," *Nature*, vol. 569, no. 7754, pp. 136–140, 2019.

[77] Y. Cai, J. Jin, T. Yao, A. J. Gottschalk, S. K. Swanson, S. Wu, Y. Shi, M. P. Washburn, L. Florens, R. C. Conaway, and J. W. Conaway, "YY1 functions with INO80 to activate transcription," *Nature Structural and Molecular Biology*, vol. 14, no. 9, pp. 872–874, 2007.

[78] E.-J. Park, S.-K. Hur, and J. Kwon, "Human INO80 chromatin-remodelling complex contributes to DNA double-strand break repair via the expression of rad54b and XRCC3 genes," *Biochemical Journal*, vol. 431, pp. 179–187, Sept. 2010.

[79] R. Kumar, D.-Q. Li, S. Müller, and S. Knapp, "Epigenomic regulation of oncogenesis by chromatin remodeling," *Oncogene*, vol. 35, pp. 4423–4436, Jan. 2016.

[80] K. L. Cheung, F. Zhang, A. Jaganathan, R. Sharma, Q. Zhang, T. Konuma, T. Shen, J. Y. Lee, C. Ren, C. H. Chen, G. Lu, M. R. Olson, W. Zhang, M. H. Kaplan, D. R. Littman, M. J. Walsh, H. Xiong, L. Zeng, and M. M. Zhou, "Distinct Roles of Brd2 and Brd4 in Potentiating the Transcriptional Program for Th17 Cell Differentiation," *Molecular Cell*, vol. 65, no. 6, pp. 1068–1080.e5, 2017.

[81] R. Andersson and A. Sandelin, "Determinants of enhancer and promoter activities of regulatory elements," *Nature Reviews Genetics*, vol. 21, pp. 71–87, Oct. 2019.

[82] S. Nagarajan, T. Hossan, M. Alawi, Z. Najafova, D. Indenbirken, U. Bedi, H. Taipaleenmäki, I. Ben-Batalla, M. Scheller, S. Loges, S. Knapp, E. Hesse, C.-M. Chiang, A. Grundhoff, and S. A. Johnsen, "Bromodomain protein BRD4 is required for estrogen receptor-dependent enhancer activation and gene transcription," *Cell Reports*, vol. 8, pp. 460–469, July 2014.

[83] H. Rahnamoun, J. Lee, Z. Sun, H. Lu, K. M. Ramsey, E. A. Komives, and S. M. Lauberth, "RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation," *Nature Structural and Molecular Biology*, vol. 25, no. 8, pp. 687–697, 2018.

[84] D. A. Bose, G. Donahue, D. Reinberg, R. Shiekhattar, R. Bonasio, and S. L. Berger, "RNA Binding to CBP Stimulates Histone Acetylation and Transcription," *Cell*, vol. 168, no. 1-2, pp. 135–149.e22, 2017.

[85] F. Lai, U. A. Orom, M. Cesaroni, M. Beringer, D. J. Taatjes, G. A. Blobel, and R. Shiekhattar, "Activating RNAs associate with mediator to enhance chromatin architecture and transcription," *Nature*, vol. 494, pp. 497–501, Feb. 2013.

[86] J. A. Shapiro and R. von Sternberg, "Why repetitive DNA is essential to genome function," *Biological Reviews*, vol. 80, pp. 227–250, May 2005.

[87] G.-F. Richard, A. Kerrest, and B. Dujon, "Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes," *Microbiology and Molecular Biology Reviews*, vol. 72, no. 4, pp. 686–727, 2008.

[88] S. Leclercq, E. Rivals, and P. Jarne, "DNA slippage occurs at microsatellite loci without minimal threshold length in humans: A comparative genomic approach," *Genome Biology and Evolution*, vol. 2, pp. 325–335, May 2010.

[89] C. W. Greider and E. H. Blackburn, "Identification of a specific telomere terminal transferase activity in tetrahymena extracts," *Cell*, vol. 43, no. 2 PART 1, pp. 405–413, 1985.

[90] H. L. Levin and J. V. Moran, "Dynamic interactions between transposable elements and their hosts," *Nature Reviews Genetics*, vol. 12, no. 9, pp. 615–627, 2011.

[91] J. J. YUNIS, L. ROLDAN, W. G. YASMINEH, and J. C. LEE, "Staining of satellite DNA in metaphase chromosomes," *Nature*, vol. 231, pp. 532–533, June 1971.

[92] A. Cournac, R. Koszul, and J. Mozziconacci, "The 3D folding of metazoan genomes correlates with the association of similar repetitive elements," *Nucleic Acids Research*, vol. 44, no. 1, pp. 245–255, 2016.

[93] T. H. Eickbush and D. G. Eickbush, "Finely orchestrated movements: Evolution of the ribosomal RNA genes," *Genetics*, vol. 175, no. 2, pp. 477–485, 2007.

[94] G. E. Palade, "A SMALL PARTICULATE COMPONENT OF THE CYTOPLASM," *The Journal of Biophysical and Biochemical Cytology*, vol. 1, pp. 59–68, Jan. 1955.

[95] M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate, and H. F. Noller, "Crystal structure of the ribosome at 5.5 Å resolution," *Science*, vol. 292, no. 5518, pp. 883–896, 2001.

[96] H. Domdey, B. Apostol, R. J. Lin, A. Newman, E. Brody, and J. Abelson, "Lariat structures are in vivo intermediates in yeast pre-mRNA splicing," *Cell*, vol. 39, no. 3 PART 2, pp. 611–621, 1984.

[97] C. L. Will and R. Lührmann, "Spliceosome structure and function," *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 7, pp. 1–2, 2011.

[98] R. C. Lee, R. L. Feinbaum, and V. Ambros, "The c. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14," *Cell*, vol. 75, pp. 843–854, Dec. 1993.

[99] S. M. Hammond, E. Bernstein, D. Beach, and G. J. Hannon, "An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells," *Nature*, vol. 404, no. 6775, pp. 293–296, 2000.

[100] L. D. Sacco, A. Baldassarre, and A. Masotti, "Bioinformatics tools and novel challenges in long non-coding RNAs (lncRNAs) functional analysis," *International Journal of Molecular Sciences*, vol. 13, pp. 97–114, Dec. 2011.

[101] S. J. Liu, T. J. Nowakowski, A. A. Pollen, J. H. Lui, M. A. Horlbeck, F. J. Attenello, D. He, J. S. Weissman, A. R. Kriegstein, A. A. Diaz, and D. A. Lim, "Single-cell analysis of long non-coding RNAs in the developing human neocortex," *Genome Biology*, vol. 17, no. 1, pp. 1–17, 2016.

[102] C. Gong and L. E. Maquat, "LncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 39 UTRs via Alu eleme," *Nature*, vol. 470, no. 7333, pp. 284–290, 2011.

[103] S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L. H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C. Kocks, F. Le Noble, and N. Rajewsky, "Circular RNAs are a large class of animal RNAs with regulatory potency," *Nature*, vol. 495, no. 7441, pp. 333–338, 2013.

[104] K. Gumireddy, A. Li, J. Yan, T. Setoyama, G. J. Johannes, U. a. Orom, J. Tchou, Q. Liu, L. Zhang, D. W. Speicher, G. a. Calin, and Q. Huang, "Identification of a long non-coding RNA-associated RNP complex regulating metastasis at the translational step.," *The EMBO journal*, vol. 32, no. 20, pp. 2672–84, 2013.

[105] M. A. Faghihi, F. Modarresi, A. M. Khalil, D. E. Wood, B. G. Sahagan, T. E. Morgan, C. E. Finch, G. St, L. Iii, P. J. Kenny, and C. Wahlestedt, "Expression of a noncoding RNA is elevated in Alzheimer ' s disease and drives rapid feed-forward regulation of b -secretase," vol. 14, no. 7, pp. 723–730, 2008.

[106] F. a. Buske, D. C. Bauer, J. S. Mattick, and T. L. Bailey, "Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data," *Genome Research*, vol. 22, no. 7, pp. 1372–1381, 2012.

[107] T. Mondal, S. Subhash, R. Vaid, S. Enroth, S. Uday, B. Reinius, S. Mitra, A. Mohammed, A. R. James, E. Hoberg, A. Moustakas, U. Gyllensten, S. J. M. Jones, C. M. Gustafsson, A. H. Sims, F. Westerlund, E. Gorab, and C. Kanduri, "MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures.," *Nature communications*, vol. 6, p. 7743, 2015.

[108] F. A. Buske, D. C. Bauer, J. S. Mattick, and T. L. Bailey, "Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data," *Genome Research*, vol. 22, no. 7, pp. 1372–1381, 2012.

[109] J. M. Engreitz, N. Ollikainen, and M. Guttman, "Long non-coding RNAs: Spatial amplifiers that control nuclear structure and gene expression," *Nature Reviews Molecular Cell Biology*, vol. 17, no. 12, pp. 756–770, 2016.

[110] C. Wang, Y. Duan, G. Duan, Q. Wang, K. Zhang, X. Deng, B. Qian, J. Gu, Z. Ma, S. Zhang, L. Guo, C. Liu, and Y. Fang, "Stress Induces Dynamic, Cytotoxicity-Antagonizing TDP-43 Nuclear Bodies via Paraspeckle LncRNA NEAT1-Mediated Liquid-Liquid Phase Separation," *Molecular Cell*, vol. 79, no. 3, pp. 443–458.e7, 2020.

[111] J. E. Wilusz, C. K. Jnbaptiste, L. Y. Lu, W. F. Marzluff, C.-d. Kuhn, L. Joshua-tor, and P. a. Sharp, "A triple helix stabilizes the 3 9 ends of long noncoding RNAs that lack poly (A) tails," *Genes \& Development*, vol. 26, no. 21, pp. 2392–2407, 2012.

[112] H. Zhang, C. Xue, Y. Wang, J. Shi, X. Zhang, W. Li, S. Nunez, A. S. Foulkes, J. Lin, C. C. Hinkle, W. Yang, E. E. Morrisey, D. J. Rader, M. Li, and M. P. Reilly, "Deep RNA sequencing uncovers a repertoire of human macrophage long intergenic noncoding RNAs modulated by macrophage activation and associated with cardiometabolic diseases," *Journal of the American Heart Association*, vol. 6, no. 11, 2017.

[113] J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. a. Brugmann, H. Goodnough, J. a. Helms, P. J. Farnham, and H. Y. Chang, "Functional Demarcation of Active and Silent Chromatin Domains in Human Loci by Non-Coding RNAs," *Cell*, vol. 129, no. 7, pp. 1311–1323, 2007.

[114] M. Akam, "Hox genes and the evolution of diverse body plans.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 349, no. 1329, pp. 313–319, 1995.

[115] R. a. Gupta, N. Shah, K. C. Wang, J. Kim, H. M. Horlings, D. J. Wong, M.-C. Tsai, T. Hung, P. Argani, J. L. Rinn, Y. Wang, P. Brzoska, B. Kong, R. Li, R. B. West, M. J. van de Vijver, S. Sukumar, and H. Y. Chang, "Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis.," *Nature*, vol. 464, no. 7291, pp. 1071–1076, 2010.

[116] M.-c. Tsai, O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal, and H. Y. Chang, "Modification Complexes," *Science*, vol. 329, no. August, pp. 689–693, 2010.

[117] C. Chu, K. Qu, F. L. Zhong, S. E. Artandi, and H. Y. Chang, "Genomic Maps of Long Noncod-

ing RNA Occupancy Reveal Principles of RNA-Chromatin Interactions," *Molecular Cell*, vol. 44, no. 4, pp. 667–678, 2011.

[118] C. J. Brown, B. D. Hendrich, J. L. Rupert, Y. Xing, J. Lawrence, and F. Willard, "The Human X / ST Gene : Analysis of a 17 kb Inactive X-Specific RNA That Contains Conserved Repeats and Is Highly Localized within the Nucleus I '," vol. 71, 1992.

[119] A. Wutz and R. Jaenisch, "A shift from reversible to irreversible x inactivation is triggered during ES cell differentiation," *Molecular Cell*, vol. 5, pp. 695–705, Apr. 2000.

[120] J. M. Engreitz, A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E. S. Lander, K. Plath, and M. Guttman, "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome," *Science*, vol. 341, no. 6147, pp. 1–9, 2013.

[121] E. M. Darrow, M. H. Huntley, O. Dudchenko, E. K. Stamenova, N. C. Durand, Z. Sun, S. C. Huang, A. L. Sanborn, I. Machol, M. Shamim, A. P. Seberg, E. S. Lander, B. P. Chadwick, and E. L. Aiden, "Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 31, pp. E4504–E4512, 2016.

[122] E. Hacisuleyman, L. A. Goff, C. Trapnell, A. Williams, J. Henao-Mejia, L. Sun, P. McClanahan, D. G. Hendrickson, M. Sauvageau, D. R. Kelley, M. Morse, J. Engreitz, E. S. Lander, M. Guttman, H. F. Lodish, R. Flavell, A. Raj, and J. L. Rinn, "Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre," *Nat Struct Mol Biol*, vol. 21, no. 2, pp. 198–206, 2014.

[123] M. Munschauer, C. T. Nguyen, K. Sirokman, C. R. Hartigan, L. Hogstrom, J. M. Engreitz, J. C. Ulirsch, C. P. Fulco, V. Subramanian, J. Chen, M. Schenone, M. Guttman, S. A. Carr, and E. S. Lander, "The NORAD lncRNA assembles a topoisomerase complex critical for genome stability," *Nature*, vol. 561, no. 7721, pp. 132–136, 2018.

[124] J. Feng, W. Funk, S. Wang, S. Weinrich, A. Avilion, C. Chiu, R. Adams, E. Chang, R. Allsopp, J. Yu, and e. al., "The RNA component of human telomerase," *Science*, vol. 269, pp. 1236–1241, Sept. 1995.

[125] B. E. Jády, P. Richard, E. Bertrand, and T. Kiss, "Cell cycle-dependent recruitment of telomerase RNA and cajal bodies to human telomeres," *Molecular Biology of the Cell*, vol. 17, pp. 944–954, Feb. 2006.

[126] Q. Zhang, N. K. Kim, and J. Feigon, "Architecture of human telomerase RNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 51, pp. 20325–20332, 2011.

[127] J. Moffat, D. a. Grueneberg, X. Yang, S. Y. Kim, A. M. Kloepfer, G. Hinkle, B. Piqani, T. M. Eisenhaure, B. Luo, J. K. Grenier, A. E. Carpenter, S. Y. Foo, S. a. Stewart, B. R. Stockwell, N. Hacohen, W. C. Hahn, E. S. Lander, D. M. Sabatini, and D. E. Root, "A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen," *Cell*, vol. 124, no. 6, pp. 1283–1298, 2006.

[128] M. S. Jurica, L. J. Licklider, S. R. Gygi, N. Grigorieff, and M. J. Moore, "Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis.," *RNA (New York, N.Y.)*, vol. 8, no. 4, pp. 426–439, 2002.

[129] B. P. Tsai, X. Wang, L. Huang, and M. L. Waterman, "Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach.," *Molecular & cellular proteomics : MCP*, vol. 10, p. M110.007385, apr 2011.

[130] M. D. Simon, C. I. Wang, P. V. Kharchenko, J. a. West, B. a. Chapman, a. a. Alekseyenko, M. L. Borowsky, M. I. Kuroda, and R. E. Kingston, "The genomic binding sites of a noncoding RNA," *Proceedings of the National Academy of Sciences*, vol. 108, no. 51, pp. 20497–20502, 2011.

[131] Y. Yang, L. Wen, and H. Zhu, "Unveiling the hidden function of long non-coding RNA by identifying its major partner-protein," *Cell and Bioscience*, vol. 5, no. 1, pp. 1–10, 2015.

[132] J. M. Engreitz, A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E. S. Lander, K. Plath, and M. Guttman, "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome," *Science*, vol. 341, no. 6147, pp. 1–9, 2013.

[133] J. M. Engreitz, K. Sirokman, P. McDonel, A. A. Shishkin, C. Surka, P. Russell, S. R. Grossman, A. Y. Chow, M. Guttman, and E. S. Lander, "RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and

chromatin sites," *Cell*, vol. 159, no. 1, pp. 188–199, 2014.

[134] N. Rani, T. J. Nowakowski, H. Zhou, S. E. Godshalk, V. Lisi, A. R. Kriegstein, and K. S. Kosik, "A Primate lncRNA Mediates Notch Signaling during Neuronal Development by Sequestering miRNA," *Neuron*, vol. 90, no. 6, pp. 1174–1188, 2016.

[135] J. Rea, V. Menci, P. Tollis, T. Santini, A. Armaos, M. G. Garone, F. Iberite, A. Cipriano, G. G. Tartaglia, A. Rosa, M. Ballarino, P. Laneve, and E. Caffarelli, "HOTAIRM1 regulates neuronal differentiation by modulating NEUROGENIN 2 and the downstream neurogenic cascade," *Cell Death and Disease*, vol. 11, no. 7, 2020.

[136] N. Lin, K. Y. Chang, Z. Li, K. Gates, Z. Rana, J. Dang, D. Zhang, T. Han, C. S. Yang, T. J. Cunningham, S. Head, G. Duester, P. D. S. Dong, and T. M. Rana, "An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment," *Molecular Cell*, vol. 53, no. 6, pp. 1005–1019, 2014.

[137] S.-Y. Ng, R. Johnson, and L. W. Stanton, "Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors," *The EMBO Journal*, vol. 31, no. 3, pp. 522–533, 2011.

[138] J. Cox and M. Mann, "Computational principles of determining and improving mass precision and accuracy for proteome measurements in an Orbitrap.," *Journal of the American Society for Mass Spectrometry*, vol. 20, pp. 1477–85, aug 2009.

[139] R. Ihaka and R. Gentleman, "R: A language for data analysis and graphics," *Journal of Computational and Graphical Statistics*, vol. 5, pp. 299–314, Sept. 1996.

[140] B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with Bowtie 2," *Nature Methods*, vol. 9, no. 4, pp. 357–359, 2012.

[141] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras, "STAR: ultrafast universal RNA -seq aligner," *Bioinformatics*, pp. 1–7, 2013.

[142] B. Li and C. N. Dewey, "RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome," *BMC Bioinformatics*, vol. 12, Aug. 2011.

[143] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, "The Sequence Alignment/Map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.

[144] J. Feng, T. Liu, B. Qin, Y. Zhang, and X. S. Liu, "Identifying ChIP-seq enrichment using MACS," *Nature Protocols*, vol. 7, pp. 1728–1740, Aug. 2012.

[145] R. C. McLeay and T. L. Bailey, "Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data," *BMC Bioinformatics*, vol. 11, 2010.

[146] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass, "Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities," *Molecular Cell*, vol. 38, no. 4, pp. 576–589, 2010.

[147] D. Smedley, S. Haider, B. Ballester, R. Holland, D. London, G. Thorisson, and A. Kasprzyk, "BioMart - Biological queries made easy," *BMC Genomics*, vol. 10, pp. 1–12, 2009.

[148] S. Anders and W. Huber, "Differential expression analysis for sequence count data," *Genome Biology*, vol. 11, Oct. 2010.

[149] Y. Liao, G. K. Smyth, and W. Shi, "The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads," *Nucleic Acids Research*, vol. 47, no. 8, 2019.

[150] T. Alenghat, J. Yu, and M. A. Lazar, "The N-CoR complex enables chromatin remodeler SNF2H to enhance repression by thyroid hormone receptor," *EMBO Journal*, vol. 25, no. 17, pp. 3966–3974, 2006.

[151] S. A. Doyle, "High-throughput cloning for proteomics research.," *Methods in molecular biology (Clifton, N.J.)*, vol. 310, pp. 107–113, 2005.

[152] K. A. LeCuyer, L. S. Behlen, and O. C. Uhlenbeck, "Mutants of the Bacteriophage MS2 Coat Protein That Alter Its Cooperative Binding to RNA," *Biochemistry*, vol. 34, no. 33, pp. 10600–10606, 1995.

[153] K. T. Gagnon, L. Li, B. A. Janowski, and D. R. Corey, "Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading," *Nat Protoc*, vol. 9, no. 9, pp. 2045–2060, 2014.

[154] R. D. Morin, M. Bainbridge, A. Fejes, M. Hirst, M. Krzywinski, T. J. Pugh, H. McDonald, R. Varhol, S. J. Jones, and M. A. Marra, "Profiling the HeLa s3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing," *BioTechniques*, vol. 45, pp. 81–94, July 2008.

[155] J. D. Dignam, R. M. Lebovitz, and R. G. Roeder, "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei," *Nucleic Acids Research*, vol. 11, pp. 1475–1489, Mar. 1983.

[156] J. R. Wiśniewski, D. F. Zielinska, and M. Mann, "Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method," *Analytical Biochemistry*, vol. 410, no. 2, pp. 307–309, 2011.

[157] T. I. Lee, S. E. Johnstone, and R. A. Young, "Chromatin immunoprecipitation and microarray-based analysis of protein location," *Nature Protocols*, vol. 1, pp. 729–748, July 2006.

[158] Y. Zhang, K. Chen, S. a. Sloan, M. L. Bennett, A. R. Scholze, S. O. Keeffe, H. P. Phatnani, X. P. Guarnieri, C. Caneda, N. Ruderisch, S. Deng, S. a. Liddelow, C. Zhang, R. Daneman, T. Maniatis, X. B. a. Barres, and X. J. Q. Wu, "An RNA-Sequencing Transcriptome and Splicing Database of Glia , Neurons , and Vascular Cells of the Cerebral Cortex," *The Journal of Neuroscience*, vol. 34, no. 36, pp. 11929–11947, 2014.

[159] M. J. Ziller, R. Edri, Y. Yaffe, C. a. Gifford, J. Xing, H. Gu, O. Kohlbacher, and A. Gnirke, "Dissecting neural differentiation regulatory networks through epigenetic footprinting," *Nature*, vol. 518, no. 7539, pp. 355–359, 2015.

[160] E. V. Makeyev, J. Zhang, M. A. Carrasco, and T. Maniatis, "The MicroRNA miR-124 Promotes Neuronal Differentiation by Triggering Brain-Specific Alternative Pre-mRNA Splicing," *Molecular Cell*, vol. 27, no. 3, pp. 435–448, 2007.

[161] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, "Image processing with imageJ," *Biophotonics International*, vol. 11, no. 7, pp. 36–41, 2004.

[162] S. Elmore, "Apoptosis: A review of programmed cell death," *Toxicologic Pathology*, vol. 35, pp. 495–516, June 2007.

[163] H. Mi, A. Muruganujan, D. Ebert, X. Huang, and P. D. Thomas, "PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools," *Nucleic Acids Research*, vol. 47, pp. D419–D426, Nov. 2018.

[164] L. Naldini, U. Blomer, P. Gallay, D. Ory, R. Mulligan, F. H. Gage, I. M. Verma, and D. Trono, "In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector," *Science*, vol. 272, pp. 263–267, Apr. 1996.

[165] J. E. Donello, J. E. Loeb, and T. J. Hope, "Woodchuck hepatitis virus contains a tripartite post-transcriptional regulatory element," *Journal of Virology*, vol. 72, pp. 5085–5092, June 1998.

[166] S. Sharma, J.-L. Langhendries, P. Watzinger, P. Kötter, K.-D. Entian, and D. L. Lafontaine, "Yeast kre33 and human NAT10 are conserved 18s rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor tan1/THUMPD1," *Nucleic Acids Research*, vol. 43, pp. 2242–2258, Feb. 2015.

[167] J. Cox and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nature Biotechnology*, vol. 26, no. 12, pp. 1367–1372, 2008.

[168] M. A. Lazzaro and D. J. Picketts, "Cloning and characterization of the murine Imitation S witch (ISWI) genes: Differential expression patterns suggest distinct developmental roles for Snf2h and Snf2l," *Journal of Neurochemistry*, vol. 77, no. 4, pp. 1145–1156, 2001.

[169] T. Stopka and A. I. Skoultchi, "The ISWI ATPase Snf2h is required for early mouse development," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. SUPPL. 2, pp. 14097–14102, 2003.

[170] E. Shang, X. Wang, D. Wen, D. A. Greenberg, and D. J. Wolgemuth, "Double bromodomain-containing gene Brd2 is essential for embryonic development in Mouse," *Developmental Dynamics*, vol. 238, no. 4, pp. 908–917, 2009.

[171] C. Penas, M. E. Maloof, V. Stathias, J. Long, S. K. Tan, J. Mier, Y. Fang, C. Valdes, J. Rodriguez-Blanco, C. M. Chiang, D. J. Robbins, D. J. Liebl, J. K. Lee, M. E. Hatten, J. Clarke, and N. G. Ayad, "Time series modeling of cell cycle exit identifies Brd4 dependent regulation of cerebellar neurogenesis," *Nature Communications*, vol. 10, no. 1, pp. 1–11, 2019.

[172] A. Ghetti, S. Pinol-roma, W. M. Michael, C. Morandi, and G. Dreyfuss, "HnRNP 1, the polypyrimidine tract-binding protein: Distinct nuclear localization and association with hnRNAs," *Nucleic Acids Research*, vol. 20, no. 14, pp. 3671–3678, 1992.

[173] A. Joshi, M. B. Coelho, O. Kotik-Kogan, P. J. Simpson, S. J. Matthews, C. W. Smith, and S. Curry, "Crystallographic analysis of polypyrimidine tract-binding protein-raver1 interactions involved in regulation of alternative splicing," *Structure*, vol. 19, no. 12, pp. 1816–1825, 2011.

[174] N. S. Cetin, C. C. Kuo, T. Ribarska, R. Li, I. G. Costa, and I. Grummt, "Isolation and genome-wide characterization of cellular DNA:RNA triplex structures," *Nucleic Acids Research*, vol. 47, no. 5, pp. 2306–2321, 2019.

[175] A. Sulovari, R. Li, P. A. Audano, D. Porubsky, M. R. Vollger, G. A. Logsdon, W. C. Warren, A. A. Pollen, M. J. Chaisson, and E. E. Eichler, "Human-specific tandem repeat expansion and differential gene expression during primate evolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 46, pp. 23243–23253, 2019.

[176] P. E. Warburton, D. Hasson, F. Guillem, C. Lescale, X. Jin, and G. Abrusan, "Analysis of the largest tandemly repeated DNA families in the human genome," *BMC Genomics*, vol. 9, 2008.

[177] Y. Shen, F. Yue, D. F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenkov, and B. Ren, "A map of the cis-regulatory sequences in the mouse genome," *Nature*, vol. 488, pp. 116–120, July 2012.

[178] J. E. Phillips-Cremins, M. E. Sauria, A. Sanyal, T. I. Gerasimova, B. R. Lajoie, J. S. Bell, C.-T. Ong, T. A. Hookway, C. Guo, Y. Sun, M. J. Bland, W. Wagstaff, S. Dalton, T. C. McDevitt, R. Sen, J. Dekker, J. Taylor, and V. G. Corces, "Architectural protein subclasses shape 3d organization of genomes during lineage commitment," *Cell*, vol. 153, pp. 1281–1295, June 2013.

[179] R. N. Plasschaert, S. Vigneau, I. Tempera, R. Gupta, J. Maksimoska, L. Everett, R. Davuluri, R. Mamorstein, P. M. Lieberman, D. Schultz, S. Hannenhalli, and M. S. Bartolomei, "CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation," *Nucleic Acids Research*, vol. 42, pp. 774–789, Oct. 2013.

[180] D. S. Sams, S. Nardone, D. Getselter, D. Raz, M. Tal, P. R. Rayi, H. Kaphzan, O. Hakim, and E. Elliott, "Neuronal CTCF is necessary for basal and experience-dependent gene regulation, memory formation, and genomic structure of BDNF and arc," *Cell Reports*, vol. 17, pp. 2418–2430, Nov. 2016.

[181] R. J. Spencer, B. C. del Rosario, S. F. Pinter, D. Lessing, R. I. Sadreyev, and J. T. Lee, "A boundary element between tsix and xist binds the chromatin insulator ctcf and contributes to initiation of x-chromosome inactivation," *Genetics*, vol. 189, pp. 441–454, Aug. 2011.

[182] X. Wang, K. J. Goodrich, A. R. Gooding, H. Naeem, S. Archer, R. D. Paucek, D. T. Youmans, T. R. Cech, and C. Davidovich, "Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines," *Molecular Cell*, vol. 65, no. 6, pp. 1056–1067.e5, 2017.

[183] S. J. Tang, "Chromatin organization by repetitive elements (CORE): A genomic principle for the higher-order structure of chromosomes," *Genes*, vol. 2, no. 3, pp. 502–515, 2011.

[184] B. J. Choi, J. H. Yoon, W. S. Choi, O. Kim, S. W. Nam, and W. S. Park, "Genetic association of KCNA5 and KCNJ3 polymorphisms in korean children with epilepsy," *Molecular & Cellular Toxicology*, vol. 10, pp. 223–228, June 2014.

[185] A. R. Aricescu, W.-C. Hon, C. Siebold, W. Lu, P. A. van der Merwe, and E. Y. Jones, "Molecular analysis of receptor protein tyrosine phosphatase -mediated cell adhesion," *The EMBO Journal*, vol. 25, pp. 701–712, Feb. 2006.

[186] S. Gnanapavan and G. Giovannoni, "Neural cell adhesion molecules in brain plasticity and disease," *Multiple Sclerosis and Related Disorders*, vol. 2, no. 1, pp. 13 – 20, 2013.

[187] R. Kawano, K. Ohta, and G. Lupo, "Cadherin-7 enhances sonic hedgehog signalling by preventing gli3 repressor formation during neural tube patterning," *Open Biology*, vol. 7, p. 170225, Dec. 2017.

[188] S. F. Rusin, K. A. Schlosser, M. E. Adamo, and A. N. Kettenbach, "Quantitative phosphoproteomics reveals new roles for the protein phosphatase PP6 in mitotic cells," *Science Signaling*, vol. 8, pp. rs12–rs12, Oct. 2015.

[189] H. Enomoto, "Regulation of neural development by glial cell line-derived neurotrophic factor family ligands," *Anatomical Science International*, vol. 80, pp. 42–52, Mar. 2005.

[190] M. Ka, D. A. Chopra, S. M. Dravid, and W. Y. Kim, "Essential roles for ARID1B in dendritic arborization and spine morphology of developing pyramidal neurons," *Journal of Neuroscience*, vol. 36, no. 9, pp. 2723–2742, 2016.

[191] E. Ensor, M. D. Smith, and D. S. Latchman, "The BRN-3a transcription factor protects sensory but not sympathetic neurons from programmed cell death/apoptosis," *Journal of Biological Chemistry*, vol. 276, pp. 5204–5212, Oct. 2000.

[192] C. D. Hudson, P. J. Morris, D. S. Latchman, and V. S. Budhram-Mahadeo, "Brn-3a transcription factor blocks p53-mediated activation of proapoptotic target genes noxa and bax in vitro and in vivo to determine cell fate," *Journal of Biological Chemistry*, vol. 280, pp. 11851–11858, Mar. 2005.

[193] R. N. Munji, Y. Choe, G. Li, J. A. Siegenthaler, and S. J. Pleasure, "Wnt signaling regulates neuronal differentiation of cortical intermediate progenitors," *Journal of Neuroscience*, vol. 31, no. 5, pp. 1676–1687, 2011.

[194] G. Vasileiou, A. B. Ekici, S. Uebe, C. Zweier, J. Hoyer, H. Engels, J. Behrens, A. Reis, and M. V. Hadjihannas, "Chromatin-Remodeling-Factor ARID1B Represses Wnt/β-Catenin Signaling," *American Journal of Human Genetics*, vol. 97, no. 3, pp. 445–456, 2015.

[195] T.-W. Yao, W.-S. Kim, D. M. Yu, G. Sharbeen, G. W. McCaughan, K.-Y. Choi, P. Xia, and M. D. Gorrell, "A novel role of dipeptidyl peptidase 9 in epidermal growth factor signaling," *Molecular Cancer Research*, vol. 9, pp. 948–959, May 2011.

[196] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity," *Science*, vol. 337, pp. 816–821, June 2012.

[197] R. Strohner, A. Nemeth, P. Jansa, U. Hofmann-Rohrer, R. Santoro, G. Längst, and I. Grummt, "NoRC - A novel member of mammalian ISWI-containing chromatin remodeling machines," *EMBO Journal*, vol. 20, no. 17, pp. 4892–4900, 2001.

[198] M. McArthur, S. Gerum, and G. Stamatoyannopoulos, "Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse -globin LCR 1 1edited by j. karn," *Journal of Molecular Biology*, vol. 313, pp. 27–34, Oct. 2001.

[199] H. Wang, M. Nakamura, T. R. Abbott, D. Zhao, K. Luo, C. Yu, C. M. Nguyen, A. Lo, T. P. Daley, M. L. Russa, Y. Liu, and L. S. Qi, "CRISPR-mediated live imaging of genome editing and transcription," *Science*, vol. 365, pp. 1301–1305, Sept. 2019.

9 Appendix

9.1 ImageJ PlugIn development

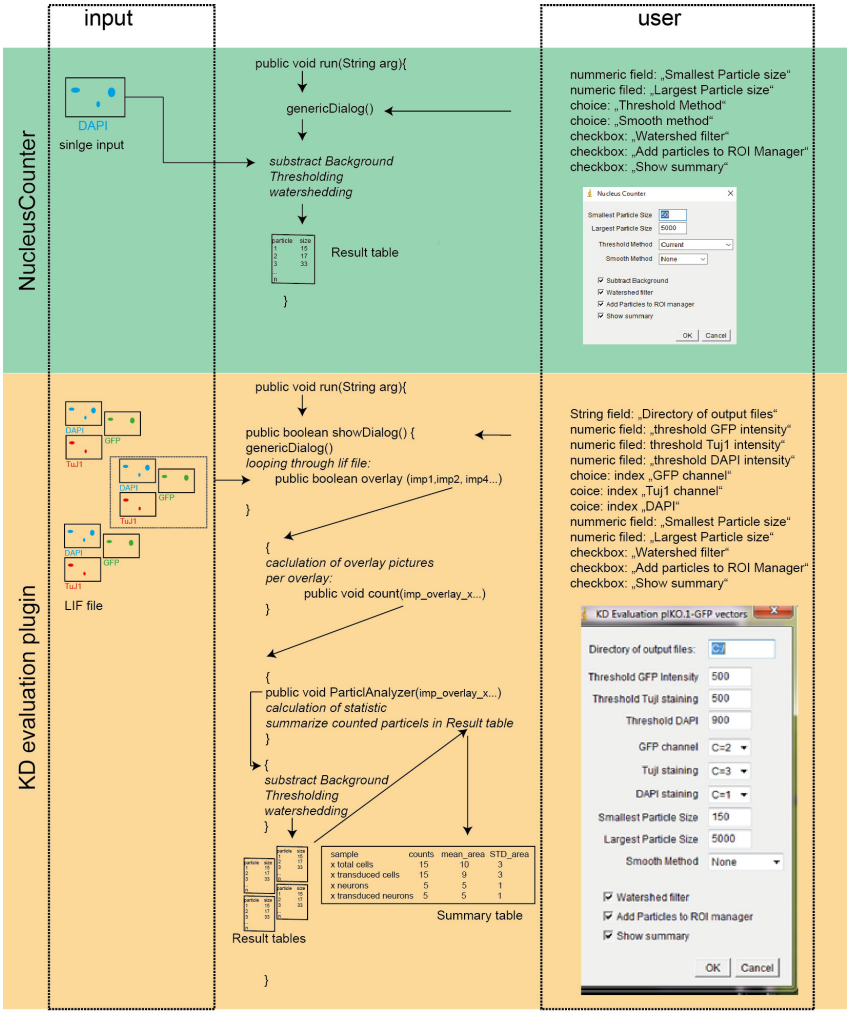


Figure 48: strategy to convert the public available NucleusCounter plugin to a full automated cell type counting plugin

To analyze the number of different stained cells, I modified the NucleusCounter plugin. I implemented a Boolean class to calculate overlay images of the three microscopy channels: DAPI (blue), GFP (green), and β tubulin III (red). Thereby, the overlay between individual channels representing images of the following cell nuclei of:

1. DAPI and GFP = transduced cells
2. DAPI and β tubulin III = neurons
3. DAPI, GFP, and β tubulin III = transduced neurons

Image J handles images either as binary ImageJProcessor objects containing all the pixel values and necessary meta-data or numeric one-dimensional pixel arrays. The number-arrays make it easy to distinguish whether a signal is present in another channel and can be quickly done by a for loop going through all pixels (p(x)) and comparing the value in all three channels. Since we recorded our microscope images at a conventional fluorescence microscope, we defined the overlay with a user-specified signal to noise threshold as followed:

$$p(x)_{transduced-cells} = \begin{cases} p(x)_{DAPI} & \text{if } p(x)_{GFP} \geq threshold_{GFP} \\ 0 \end{cases} \tag{1}$$

$$p(x)_{neurons} = \begin{cases} p(x)_{DAPI} & \text{if } p(x)_{\beta-tubulin} \geq threshold_{\beta-tubulin} \\ 0 \end{cases} \tag{2}$$

$$p(x)_{transduced-neurons} = \begin{cases} p(x)_{DAPI} & \text{if } p(x)_{\beta-tubulin} \geq threshold_{\beta-tubulin} \ \& \ p(x)_{GFP} \geq threshold_{GFP} \\ 0 \end{cases} \tag{3}$$

Furthermore, we also defined to remove the noise in the DAPI channel to calculate the total cell number by the same method:

$$p(x)_{transduced-cells} = \begin{cases} p(x)_{DAPI} & \text{if } p(x)_{DAPI} \geq threshold_{DAPI} \\ 0 \end{cases} \tag{4}$$

Next, I modified and transferred the GenericDialog method into a Boolean showDialog() class. This dialog pops up when the plugin gets started. I added our threshold boxes for DAPI, GFP, and β tubulin III. I also modified the plugin to calculate different cells over an entire opened lif file containing multiple microscope images. I defined a simple for-loop going through the lif file, extracting all three channels, and running the overlay method per image. The different channels are assigned by indices that the user must define by a choice box. Since the plugin calculates a plethora of overlay-, thresholded- and outline-images, we were interested in saving all processed image files. Therefore, I added a String box to specify the directory where the SaveImage method saves all processed images. After calculating all three overlay images and modified DAPI images, the overlay class submits each picture, including the title, separately to the count class. The overlay class uses the microscope image's original name and tag it with the overlay's respective name for the title. The count class is a newly created void class and runs the ParticAnalyzer class. The ParticlAnalyzer class corresponds to the main run class of the original NucleusCounter plugin. I have chosen as threshold-method: "Otsu thresholding 16Bit" and as watershed-filter: "circularity = 0.5-1" by default. For thresholding, I set 255 as the maximum value by default. The count class provides the minimal value for thresholding. Therefore, the count class calculates the minimum of all threshold values used to calculate the overlay. Since the microscope records 12-bit

pictures, the count class converts the 12-bit value to an 8-bit value. After running the ParticlAnalyzer class, the count class extracts from the public result table counted particles and areas. With those extracted values, the count class calculates to the total number of counted particles - representing the number of cells of the respective overlay - the mean area and STD area of all counted particles as a measure of quantification accuracy and adds all three values to a summary table. The plugin's run class, which resembles the main class of java programs, solely launch the boolean showDialog().

Listing 1: Source Code of the developed plugin

```
// this plugin implements Particle Analyzer, the watersheding and tresholding
    tool, similar to Nucleus counter
// By stating threshold values the plugin compares whether a pixels is green
    and red, green and blue, red and blue or all and generates overlay pictures
    .
// plugin works on series
import java.awt.*;
import java.awt.event.*;
import java.io.*;
import java.util.*;
import java.awt.List;
import ij.*;
import ij.process.*;
import ij.gui.*;
import java.lang.*;

import ij.measure.ResultsTable;
import ij.plugin.PlugIn;
import ij.plugin.filter.PlugInFilter;
import ij.plugin.filter.Analyzer;
import ij.plugin.frame.PlugInFrame;
import ij.*;
import ij.process.*;
import ij.io.FileSaver;
import java.awt.image.*;
import java.awt.*;
import java.io.*;
import ij.io.OpenDialog;
import java.util.Iterator;
import javax.imageio.*;
import javax.imageio.stream.*;

public class KD_evaluation_GFPv8 extends java.lang.Object implements PlugIn {

// imp1 = green channel
// imp2 = red channel
// imp3 = overlay: green/red = transduced neurons
// imp4 = blue channel
// imp5 =  overlay: blue and green = nuclei of transduced cells. Cells are
    counted by particle analyzer
// imp6 = overlay: red and blue = nuclei of neurons. Cells are counted by
    particle analyzer
// imp7 = overlay: red/green and blue = nuclei of transduced neurons. Cells are
     counted by particle analyzer.
// imp8 = modified blue channel
```

```
    private ImagePlus imp1;

  // imp1 = DAPI channels
    private ImagePlus imp2;
  private ImagePlus imp3;
  private ImagePlus imp4;
  private ImagePlus imp5;
  private ImagePlus imp6;
  private ImagePlus imp7;
  public  ImagePlus imp8;
  public float[] fl;
  public double [] areas;
  public double meanA = 0;
  public double STDa = 0;
  public String Dapi;
  public String GFP;
  public String TujI;
  public String directory = (String)Prefs.get("", "C:/");
  public ResultsTable rt;
  public ResultsTable summaryTable = new ResultsTable();

  public String path = "";

  public int maxG = (int)Prefs.get("NC_max.intA", 500);
  public int maxR = (int)Prefs.get("NC_max.intB", 500);
  public int maxB = (int)Prefs.get("NC_max.intC", 900);
  // area ints count green, red and both colored pixels, respectively , but are
     not used further
  public int area_green;
  public int area_red;
  public int area_overlay;
  public String Titel;
  public String Titel1;
  public String Titel2;
  public String Titel3;
  public String Titel4;
// nucleus counter
// int max = largest particle size
// int minP = smalles particle size
  public int max = (int)Prefs.get("NC_max.int", 5000);
  public int minP = (int)Prefs.get("NC_min.intP", 150);


//  public int threshIndex=(int)Prefs.get("NC_threshIndex.int",0);
  public int smoothIndex=(int)Prefs.get("NC_smoothIndex.int",0);
  public boolean watershed=Prefs.get("NC_watershed.boolean",true);
  public boolean summarize=Prefs.get("NC_summarize.boolean",true);
  public boolean add=Prefs.get("NC_record.boolean",true);
  public boolean fullStats=Prefs.get("NC_fullStats.boolean",true);
// int minthreshold2 = 12bit value
  public int minthreshold2;
// int minthreshold = not used anymore
  public int minthreshold;
// int minThreshold = 8bit value
  public int minThreshold;
// int maxThreshold = 8bit value
  public int maxThreshold = 255;
```

```
// int maxhreshold = not used anymore
  public int maxthreshold;

  public double ln1;
  public double ln2;
  public double ln3;
  public double ratio;
  public int choice;
public boolean overlay(ImagePlus imp1, ImagePlus imp2, ImagePlus imp4, int maxG
    , int maxR, int maxB, String Titel, String Titel2, String Titel3, String
    Titel4) {
  // imp1 = GFP channel, imp2 = TujI channel, imp4 = Dapi channel
  summaryTable.incrementCounter();
  int w = imp1.getWidth();
  int h = imp1.getHeight();
  //initializing imp 5 = overlay between DAPI and GFP = tranduced cells
  imp5 = NewImage.createShortImage(Titel2, w, h,
        1, NewImage.FILL_BLACK);
  //initializing imp 6 = overlay between DAPI and TujI = neurons
  imp6 = NewImage.createShortImage(Titel3, w, h,
        1, NewImage.FILL_BLACK);
  //initializing imp 6 = overlay between DAPI, GFP and TujI = transduced
    neurons
  imp7 = NewImage.createShortImage(Titel4, w, h,
        1, NewImage.FILL_BLACK);
  //initializing imp 8 = copied DAPI channel = total cells
  imp8 = NewImage.createShortImage("DAPI modified", w, h,
        1, NewImage.FILL_BLACK);
  ImageProcessor IP1 = imp1.getProcessor();
  ImageProcessor IP2 = imp2.getProcessor();
  ImageProcessor IP4 = imp4.getProcessor();
  ImageProcessor IP5 = imp5.getProcessor();
  ImageProcessor IP6 = imp6.getProcessor();
  ImageProcessor IP7 = imp7.getProcessor();
  ImageProcessor IP8 = imp8.getProcessor();

  short[] pixels1 = (short[])IP1.getPixels();
  short[] pixels2 = (short[])IP2.getPixels();
  short[] pixels3 = (short[])IP3.getPixels();

  short[] pixels4 = (short[])IP4.getPixels();
  short[] pixels5 = (short[])IP5.getPixels();
  short[] pixels6 = (short[])IP6.getPixels();
  short[] pixels7 = (short[])IP7.getPixels();
  short[] pixels8 = (short[])IP8.getPixels();

// calculation of overlay images

  for (int i =0; i < pixels1.length; i++){

      int p1 = pixels1[i];
      int p2 = pixels2[i];
      int p4 = pixels4[i];
      int p5 = 0;
      int p6 = 0;
      int p7 = 0;
      int p8 = 0;

      if (p1 >= maxG && p2 >= maxR && p4 >= maxB) {
```

```
            p7 = p4;
          }
          if (p1 >= maxG && p4 >= maxB) {
            p5 = p4;
          }
          if (p2 >= maxR && p4 >= maxB) {
            p6 = p4;
          }
          if (p4 >= maxB){
            p8 = p4;
          }
          pixels3[i] = (short)(p3);
          pixels5[i] = (short)(p5);
          pixels6[i] = (short)(p6);
          pixels7[i] = (short)(p7);
          pixels8[i] = (short)(p8);
        }


    // saving overlay pictures as jpg files
      imp3.show();
      imp3.updateAndDraw();
      path = directory + "/" + Titel4 + "2" + ".jpg";
      path = path.replace("," , "-" );
      saveAsJpeg(imp3, path, 75);
      this.imp3.close();

      imp5.show();
      imp5.updateAndDraw();
      path = directory+ "/"  + Titel2 + ".jpg";
      path = path.replace("," , "-" );
      saveAsJpeg(imp5, path, 75);


      imp6.show();
      imp6.updateAndDraw();
      path = directory+ "/"  + Titel3+ ".jpg";
      path = path.replace("," , "-" );
      saveAsJpeg(imp6, path, 75);



      imp7.show();
      imp7.updateAndDraw();
      path = directory + "/" + Titel4+ ".jpg";
      path = path.replace("," , "-" );
      saveAsJpeg(imp7, path, 75);


      imp8.show();
      imp8.updateAndDraw();
      path = directory + "/" + Titel1 + ".jpg";
      path = path.replace("," , "-" );
      saveAsJpeg(imp8, path, 75);

    // running of Particle Analyzer method for each overlay:

    // running of Particle Analyzer method for each overlay:
```

```
// total cells, DAPI modified
  int[] minU1 = new int [1];
  minU1[0] = maxB;
  count(minU1, this.imp8, Titel1, true);
  this.imp8.close();

// transduced cells
  int[] minU2 = new int [2];
  minU2[0] = maxB;
  minU2[1] = maxG;
  count(minU2, this.imp5, Titel2, true);
  this.imp5.close();

// neurons
  int[] minU4 = new int [2];
  minU4[0] = maxB;
  minU4[1] = maxR;
  count(minU4,  this.imp6, Titel3, true);
  this.imp6.close();

// KD neurons
  int[] minU5 = new int [3];
  minU5[0] = maxB;
  minU5[1] = maxG;
  minU5[2] = maxR;
  count(minU5, this.imp7, Titel4, false);
  this.imp7.close();




  summaryTable.show("summary KD evaluation");
//  IJ.showMessage("area of green pixels = " + area_green + "\n" + "area of red
     pixels = " + area_red);
  return true;
  }


public void ParticlAnalyzer(ImagePlus imp, String Titel, int minThreshold, int
    minthreshold2){
// many parts copied from nucleus counter, default threshold values were
    changed, respectively.... Default Thresholding method = "OtsuThresholding
    16Bit"

  ImageProcessor ip1 = imp.getProcessor();
  ImageProcessor ip2 = ip1.duplicate();
  new ImagePlus("Analysis", ip2).show();
  IJ.run("Grays");

  ImagePlus imp2 = WindowManager.getCurrentImage();
  ImageWindow winimp2=imp2.getWindow();
//duplicate this for image thresholding
  ImageProcessor ip3 = ip2.duplicate();
  ip3.setThreshold(minThreshold,maxThreshold,4);
  new ImagePlus("Threshold", ip3).show();
```

```
  ImagePlus imp3 = WindowManager.getCurrentImage();
  ImageWindow winimp3=imp3.getWindow();
  ImagePlus imp5=null;
  ImageWindow winimp5=null;
  //IJ.showMessage("thres="+threshIndex);

//set threshold

    if (smoothIndex==1) IJ.run("Mean...", "radius=2 separable");
    else if(smoothIndex==2) IJ.run("Mean...", "radius=3 separable");
    else if(smoothIndex==3) IJ.run("Median...", "radius=2");
    else if(smoothIndex==4) IJ.run("Median...", "radius=3");
    WindowManager.setCurrentWindow(winimp3);
    IJ.run("OtsuThresholding 16Bit");
/
    IJ.run("Convert to Mask");
//watershed
    if (watershed) IJ.run("Watershed");
    IJ.setThreshold(minThreshold,maxThreshold);
    ImagePlus mask = WindowManager.getCurrentImage();

//analyze particles
    String analyseStr = "size=" + minP +"-" + max +" circularity=0.5-1.0 show="
    ;
    analyseStr+="Outlines";
    //analyseStr+=" display";
    analyseStr+=" exclude clear";

    if (summarize) analyseStr+=" summarize";
    if (add) analyseStr+=" add";
    //if (fullStats) analyseStr+=" size";

//IJ.showMessage(analyseStr);

    IJ.run("Analyze Particles...",   analyseStr+ " add");


//get Outline image
    ImagePlus imp4 = WindowManager.getCurrentImage();
    ImageProcessor ip4 = imp4.getProcessor();
    ImageWindow winimp4 = imp4.getWindow();
    IJ.run("Rename...", "title=boundaries");
    path = directory + "/" + Titel+ "-outlines.jpg";
    path = path.replace("," , "-" );
    saveAsJpeg(imp4, path, 75);

  imp2.changes = false; winimp2.close();
  imp4.changes = false;

  winimp4.close();

    mask.changes=false;

    mask.close();
}
public void run(String arg) {

  this.showDialog();
```

```
  }
// Generic Dialog Box
public boolean showDialog() {
  /**
 * Generate Dialog box
 * Get set image processor
 * overlay:
 */

  String[] channel = new String[4];
  channel[0] = "C=0";
  channel[1] = "C=1";
  channel[2] = "C=2";
  channel[3] = "C=3";

  int index1 = 2;
  int index2 = 3;
  int index3 = 1;

// Nucleus counter
//  String []   thresholds  = {"Current", "Otsu", "Max. Entropy", "Mixture
    Modelling", "k-means Clustering", "Adaptive"};
  String []   smooths= {"None", "Mean 3x3", "Mean 5x5", "Median 3x3", "Median 5
    x5"};
//  overlay
  GenericDialog gd = new GenericDialog("KD Evaluation plKO.1-GFP vectors");
  gd.addStringField("Directory of output files: ", directory);
  gd.addNumericField("Threshold GFP Intensity", maxG,0);
  gd.addNumericField("Threshold TujI staining", maxR,0);
  gd.addNumericField("Threshold DAPI", maxB,0);
    gd.addChoice("GFP channel", channel, channel[index1]);
    gd.addChoice("TujI staining", channel, channel[index2]);
  gd.addChoice("DAPI staining", channel, channel[index3]);


//  Nucleus counter
  gd.addNumericField("Smallest Particle Size", minP,0);
  gd.addNumericField("Largest Particle Size", max,0);
//  gd.addChoice("Threshold Method", thresholds,thresholds[threshIndex]);
  gd.addChoice("Smooth Method", smooths,smooths[smoothIndex]);

//  gd.addCheckbox("Smooth prior to segmentation",smooth);
  gd.addCheckbox("Watershed filter",watershed);
  gd.addCheckbox("Add Particles to ROI manager",add);
//  gd.addCheckbox("Show full statistics",fullStats);
  gd.addCheckbox("Show summary",summarize);


  gd.showDialog();
  if (gd.wasCanceled()) {
          return false;
  }

// overlay
    index1 = gd.getNextChoiceIndex();
    index2 = gd.getNextChoiceIndex();
```

```
  index3 = gd.getNextChoiceIndex();
  directory = gd.getNextString();
  GFP = channel[index1];
    TujI = channel[index2];
  Dapi = channel[index3];
  maxG=(int)gd.getNextNumber();
  maxR=(int)gd.getNextNumber();
  maxB=(int)gd.getNextNumber();

// Nucleus counter
  minP=(int)gd.getNextNumber();
  max=(int)gd.getNextNumber();
//  threshIndex = 1;
  smoothIndex=gd.getNextChoiceIndex();


  watershed=gd.getNextBoolean();
  add=gd.getNextBoolean();
//  fullStats=gd.getNextBoolean();
  summarize=gd.getNextBoolean();
   ImagePlus imagePlus;
   int[] arrn = WindowManager.getIDList();
  String[] arrstring = new String[arrn.length];
    int n = 0;
    while (n < arrn.length) {

        imagePlus = WindowManager.getImage((int)arrn[n]);
        if (imagePlus != null) {
             arrstring[n] = imagePlus.getTitle();

        } else {
        arrstring[n] = "";
        }
        ++n;
        }
  String imA = "";
  String imB = "";
  String imC = "";
  String imD = "";

// looping through opened lif file

  for (int i= 0; i < (arrstring.length)-3; i++){
//  IJ.showMessage("click");
    if (arrstring[i].contains("=") && arrstring[i+1].contains("=") && arrstring
    [i+2].contains("=") && arrstring[i+3].contains("=")){
//  IJ.showMessage("clickagain");
      imA = arrstring[i];
      imB = arrstring[i+1];
      imC = arrstring[i+2];
      imD = arrstring[i+3];
      String[] ImA = imA.split("-\\s+");
      String[] ImB = imB.split("-\\s+");
      String[] ImC = imC.split("-\\s+");
      String[] ImD = imD.split("-\\s+");
        if(ImA[1].equals(ImB[1]) &&  ImC[1].equals(ImD[1]) && ImD[1].equals(ImA
    [1])){

//        IJ.showMessage("toomanyClicks");
```

```
            this.imp1 = WindowManager.getImage((int)arrn[i+ index1]);
            IJ.run("Subtract Background...", "rolling=50");
            this.imp2 = WindowManager.getImage((int)arrn[i+ index2]);
            IJ.run("Subtract Background...", "rolling=50");
            this.imp4 = WindowManager.getImage((int)arrn[i+ index3]);
            IJ.run("Subtract Background...", "rolling=50");
            Titel1 = ImA[1] + "," + "total cells";
            Titel2 = ImA[1] + "," + "transduced cells";
            Titel3 = ImA[1] + "," + "neurons";
            Titel4 = ImA[1] + "," + "transduced neurons";
            this.overlay(this.imp1, this.imp2, this.imp4, maxG, maxR,maxB, Titel,
        Titel2, Titel3, Titel4);


            }

        }
      }
      return true;

    }
    String saveAsJpeg(ImagePlus imp, String path, int quality) {
            int width = imp.getWidth();
            int height = imp.getHeight();
            int biType = BufferedImage.TYPE_INT_RGB;
            boolean overlay = imp.getOverlay()!=null && !imp.getHideOverlay();
            if (imp.getProcessor().isDefaultLut() && !imp.isComposite() && !overlay
        )
                biType = BufferedImage.TYPE_BYTE_GRAY;
            BufferedImage bi = new BufferedImage(width, height, biType);
            String error = null;
            try {
                Graphics g = bi.createGraphics();
                Image img = imp.getImage();
                if (overlay)
                    img = imp.flatten().getImage();
                g.drawImage(img, 0, 0, null);
                g.dispose();
                Iterator iter = ImageIO.getImageWritersByFormatName("jpeg");
                ImageWriter writer = (ImageWriter)iter.next();
                File f = new File(path);
                String originalPath = null;
                boolean replacing = f.exists();
                if (replacing) {
                    originalPath = path;
                    path += ".temp";
                    f = new File(path);
                }
                ImageOutputStream ios = ImageIO.createImageOutputStream(f);
                writer.setOutput(ios);
                ImageWriteParam param = writer.getDefaultWriteParam();
                param.setCompressionMode(param.MODE_EXPLICIT);
                param.setCompressionQuality(quality/100f);
                if (quality == 100)
                    param.setSourceSubsampling(1, 1, 0, 0);
                IIOImage iioImage = new IIOImage(bi, null, null);
                writer.write(null, iioImage, param);
                ios.close();
                writer.dispose();
```

```
            if (replacing) {
                File f2 = new File(originalPath);
                boolean ok = f2.delete();
                if (ok) f.renameTo(f2);
            }
        } catch (Exception e) {
            error = ""+e;
            IJ.error("Jpeg Writer", ""+error);
        }
        return error;
    }
public void count (int[] min, ImagePlus impA, String Titel, boolean
    incrementCol){


    try {
    minthreshold = getMinValue(min);
    ln1 = Math.log(choice);
    ratio = ln1 / 12;
    ln2 = 8*ln2;
    ln3 = Math.exp(ln2);
    minThreshold = (int)ln3;
    ParticlAnalyzer(impA, Titel, minThreshold, minthreshold);
    rt = Analyzer.getResultsTable();
    fl = rt.getColumn(0);
    int l4 = fl.length;
    areas = rt.getColumnAsDoubles(0);
    meanA = mean_area(areas);
    STDa = std_area(meanA, areas);
    summaryTable.addValue("sample", Titel);
    summaryTable.addValue("counts", l4);
    summaryTable.addValue("mean_area", meanA);
    summaryTable.addValue("std_area", STDa);
    if (incrementCol == true){
      summaryTable.incrementCounter();
      }

    impA.close();
  }   catch (Exception e) {
      //      e.printStackTrace();
    summaryTable.addValue("sample", Titel);
    summaryTable.addValue("counts", 0);
    summaryTable.addValue("mean_area", 0);
    summaryTable.addValue("std_area", 0);
    if (incrementCol == true){
      summaryTable.incrementCounter();
      }
    impA.close();
 //     return false;
        }

}
public double mean_area (double [] areas){
  int length = areas.length;
  double sum = 0;
  for (int i =0; i < length; i ++){
    sum = sum + areas[i];
  }
  return sum/length;
```

```
  }

  public double std_area (double mean, double[] areas){

    int length = areas.length;
    double sum = 0;
    for (int i =0; i < length; i ++){
      sum = sum + ((areas[i] - mean)*(areas[i] - mean));
    }
    double var = sum/(length-1);
    double std = Math.sqrt(var);
    return std;

  }
  public int getMinValue(int[] min){
    int Min = 0;
    Min = min[0];
    for (int i = 0; i < min.length; i++){
      if (min[i] <= Min){
        Min = min[i];
      }
    }
    return Min;
  }
  }
```

9.2 Nucleotide Sequences

9.2.1 pcDNA5-5xMS2-FRT

```
LOCUS       pcDNA5_5xMS2_FRT        5160 bp ds-DNA     circular     27-JAN-2021
DEFINITION  synthetic circular DNA
COMMENT     ApEinfo:methylated:1
FEATURES             Location/Qualifiers
     primer_bind     complement(44..63)
                     /note="pRS vectors, use to sequence yeast selectable
                     marker"
                     /locus_tag="pRS-marker"
                     /ApEinfo_fwdcolor="#14c0bd"
                     /ApEinfo_revcolor="#4ec02b"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     primer_bind     863..882
                     /locus_tag="T7"
                     /ApEinfo_fwdcolor="cyan"
                     /ApEinfo_revcolor="green"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     primer_bind     complement(1113..1130)
                     /note="Bovine growth hormone terminator, reverse primer.
                     Also called BGH reverse"
                     /locus_tag="BGH-rev"
                     /ApEinfo_fwdcolor="#14c0bd"
                     /ApEinfo_revcolor="#4ec02b"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     polyA_signal    1119..1343
                     /note="bovine growth hormone polyadenylation signal"
```

```
                /locus_tag="bGH poly(A) signal"
                /ApEinfo_fwdcolor="#ff3eee"
                /ApEinfo_revcolor="#ff3eee"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
primer_bind     complement(1476..1495)
                /note="F1 origin, reverse primer"
                /locus_tag="F1ori-R"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
misc_feature    938..954
                /locus_tag="MS2-loop1"
                /ApEinfo_fwdcolor="cyan"
                /ApEinfo_revcolor="green"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
protein_bind    1626..1673
                /bound_moiety="FLP recombinase from the Saccharomyces
                cerevisiae 2u plasmid"
                /note="FLP-mediated recombination occurs in the 8-bp core
                sequence TCTAGAAA (Turan and Bode, 2011)."
                /locus_tag="FRT"
                /ApEinfo_fwdcolor="#ff0080"
                /ApEinfo_revcolor="pink"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
misc_feature    1002..1018
                /locus_tag="MS2-loop3"
                /ApEinfo_fwdcolor="cyan"
                /ApEinfo_revcolor="green"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
CDS             1681..2703
                /codon_start=1
                /gene="aph(4)-Ia"
                /product="aminoglycoside phosphotransferase from E. coli"
                /note="confers resistance to hygromycin"
                /translation="KKPELTATSVEKFLIEKFDSVSDLMQLSEGEESRAFSFDVGGRG
                Y
                VLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLTYCISRRAQGVTLQDLP
                E
                TELPAVLQPVAEAMDAIAAADLSQTSGFGPFGPQGIGQYTTWRDFICAIADPHVYHWQ
                T
                VMDDTVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDNGRITAVIDWSEAMF
                G
                DSQYEVANIFFWRPWLACMEQQTRYFERRHPELAGSPRLRAYMLRIGLDQLYQSLVDG
                N
                FDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCVEVLADSGNRRPSTRPRAK
                E "
                /locus_tag="HygR"
                /ApEinfo_fwdcolor="#00ff00"
                /ApEinfo_revcolor="#e9d024"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
misc_feature    1034..1050
                /locus_tag="MS2-loop4"
                /ApEinfo_fwdcolor="cyan"
```

```
                /ApEinfo_revcolor="green"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     polyA_signal    2833..2954
                /note="SV40 polyadenylation signal"
                /locus_tag="SV40 poly(A) signal"
                /ApEinfo_fwdcolor="#ff3eee"
                /ApEinfo_revcolor="#ff3eee"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     promoter        236..852
                /locus_tag="CMV_immearly_promoter"
                /ApEinfo_fwdcolor="#ff0000"
                /ApEinfo_revcolor="#ff0000"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     primer_bind     complement(2870..2889)
                /note="SV40 polyA, reverse primer"
                /locus_tag="SV40pA-R"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     primer_bind     2924..2943
                /note="SV40 polyA terminator, reverse primer"
                /locus_tag="EBV-rev"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     primer_bind     complement(3003..3019)
                /note="common sequencing primer, one of multiple similar
                variants"
                /locus_tag="M13 rev"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     primer_bind     complement(3003..3019)
                /note="In lacZ gene. Also called M13-rev"
                /locus_tag="M13 Reverse"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     primer_bind     complement(3016..3038)
                /note="In lacZ gene"
                /locus_tag="M13/pUC Reverse"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     protein_bind    3027..3043
                /bound_moiety="lac repressor encoded by lacI"
                /note="The lac repressor binds to the lac operator to
                inhibit transcription in E. coli. This inhibition can be
                relieved by adding lactose or
                isopropyl-beta-D-thiogalactopyranoside (IPTG)."
                /locus_tag="lac operator"
```

```
                /ApEinfo_fwdcolor="pink"
                /ApEinfo_revcolor="pink"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
promoter        complement(3051..3081)
                /note="promoter for the E. coli lac operon"
                /locus_tag="lac promoter"
                /ApEinfo_fwdcolor="#346ee0"
                /ApEinfo_revcolor="#346ee0"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
protein_bind    3096..3117
                /bound_moiety="E. coli catabolite activator protein"
                /note="CAP binding activates transcription in the presence
                of cAMP."
                /locus_tag="CAP binding site"
                /ApEinfo_fwdcolor="pink"
                /ApEinfo_revcolor="pink"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
primer_bind     complement(3234..3251)
                /note="L4440 vector, forward primer"
                /locus_tag="L4440"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
rep_origin      complement(3405..3993)
                /direction=LEFT
                /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                replication"
                /locus_tag="ori"
                /ApEinfo_fwdcolor="#999999"
                /ApEinfo_revcolor="#999999"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
primer_bind     complement(3485..3504)
                /note="pBR322 origin, forward primer"
                /locus_tag="pBR322ori-F"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
CDS             complement(4164..5024)
                /codon_start=1
                /gene="bla"
                /product="beta-lactamase"
                /note="confers resistance to ampicillin, carbenicillin,
                and related antibiotics"
                /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                I
                ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEY
                S
                PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDR
                W
                EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRS
                A
                LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGA
                S LIKHW"
```

```
                     /locus_tag="AmpR"
                     /ApEinfo_fwdcolor="#e9d024"
                     /ApEinfo_revcolor="#e9d024"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     primer_bind     4787..4806
                     /note="Ampicillin resistance gene, reverse primer"
                     /locus_tag="Amp-R"
                     /ApEinfo_fwdcolor="#14c0bd"
                     /ApEinfo_revcolor="#4ec02b"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     promoter        complement(5025..5129)
                     /gene="bla"
                     /locus_tag="AmpR promoter"
                     /ApEinfo_fwdcolor="#346ee0"
                     /ApEinfo_revcolor="#346ee0"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     misc_feature    3016..3038
                     /locus_tag="#1854"
                     /ApEinfo_fwdcolor="cyan"
                     /ApEinfo_revcolor="green"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     misc_feature    970..986
                     /locus_tag="MS2-loop2"
                     /ApEinfo_fwdcolor="cyan"
                     /ApEinfo_revcolor="green"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     misc_feature    1065..1081
                     /locus_tag="MS2-loop5"
                     /ApEinfo_fwdcolor="cyan"
                     /ApEinfo_revcolor="green"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     misc_feature    908..913
                     /locus_tag="LIC_cloning_site"
                     /ApEinfo_fwdcolor="cyan"
                     /ApEinfo_revcolor="green"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
ORIGIN
        1 gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg
       61 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg
      121 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc
      181 ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatACG CGTTGACATT
      241 GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA
      301 TGGAGTTCCG CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC
      361 CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA GGGACTTTCC
      421 ATTGACGTCA ATGGGTGGAG TATTTACGGT AAACTGCCCA CTTGGCAGTA CATCAAGTGT
      481 ATCATATGCC AAGTACGCCC CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT
      541 ATGCCCAGTA CATGACCTTA TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA
      601 TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG
      661 ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC
      721 AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC CCCATTGACG CAAATGGGCG
      781 GTAGGCGTGT ACGGTGGGAG GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA
      841 CTGCTTACTG GCTTATCGAA ATTAATACGA CTCACTATAG GGAGACCCAA GCTGGCTAGC
```

```
 901 GTTTAAACTT AAGCTTGGTA CCGAGCTCGG ATCCGGCGTA CACCATCAGG GtACGAGCTA
 961 GCCCATGGCG TACACCATCA GGGTACGACT CAGGATCCGG CGTACACCAT CAGGGTACGA
1021 GCTAGCCCAT GGCGTACACC ATCAGGGTAC GACTAGTAGA TCTCGTACAC CATCAGGGTA
1081 CTCTAGAGGG CCCGTTTAAA CCCGCTGATC AGcctcgact gtgccttcta gttgccagcc
1141 atctgttgtt tgcccctccc ccgtgccttc cttgaccctg gaaggtgcca ctcccactgt
1201 cctttcctaa taaaatgagg aaattgcatc gcattgtctg agtaggtgtc attctattct
1261 gggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata gcaggcatgc
1321 tggggatgcg gtgggctcta tggcttctga ggcggaaaga accagctggg gctctagggg
1381 gtatccccac gcgccctgta gcggcgcatt aagcgcggcg ggtgtggtgg ttacgcgcag
1441 cgtgaccgct acacttgcca gcgccctagc gcccgctcct ttcgctttct tcccttcctt
1501 tctcgccacg ttcgccggct ttccccgtca agctctaaat cgggggtccc tttagggttc
1561 cgatttagtg ctttacggca cctcgacccc aaaaaacttg attagggtga tggttcacgt
1621 acctagaagt tcctattccg aagttcctat tctctagaaa gtataggaac ttccttggcc
1681 aaaaagcctg aactcaccgc gacgtctgtc gagaagtttc tgatcgaaaa gttcgacagc
1741 gtctccgacc tgatgcagct ctcggagggc gaagaatctc gtgctttcag cttcgatgta
1801 ggagggcgtg gatatgtcct gcgggtaaat agctgcgccg atggtttcta caaagatcgt
1861 tatgtttatc ggcactttgc atcggccgcg ctcccgattc cggaagtgct tgacattggg
1921 gaattcagcg agagcctgac ctattgcatc tcccgccgtg cacagggtgt cacgttgcaa
1981 gacctgcctg aaaccgaact gcccgctgtt ctgcagccgg tcgcggaggc catggatgcg
2041 atcgctgcgg ccgatcttag ccagacgagc gggttcggcc cattcggacc gcaaggaatc
2101 ggtcaataca ctacatggcg tgatttcata tgcgcgattg ctgatcccca tgtgtatcac
2161 tggcaaactg tgatggacga caccgtcagt gcgtccgtcg cgcaggctct cgatgagctg
2221 atgctttggg ccgaggactg ccccgaagtc cggcacctcg tgcacgcgga tttcggctcc
2281 aacaatgtcc tgacggacaa tggccgcata acagcggtca ttgactggag cgaggcgatg
2341 ttcggggatt cccaatacga ggtcgccaac atcttcttct ggaggccgtg gttggcttgt
2401 atggagcagc agacgcgcta cttcgagcgg aggcatccgg agcttgcagg atcgccgcgg
2461 ctccgggcgt atatgctccg cattggtctt gaccaactct atcagagctt ggttgacggc
2521 aatttcgatg atgcagcttg ggcgcagggt cgatgcgacg caatcgtccg atccggagcc
2581 gggactgtcg ggcgtacaca aatcgcccgc agaagcgcgg ccgtctggac cgatggctgt
2641 gtagaagtac tcgccgatag tggaaaccga cgccccagca ctcgtccgag ggcaaaggaa
2701 tagcacgtac tacgagattt cgattccacc gccgccttct atgaaaggtt gggcttcgga
2761 atcgttttcc gggacgccgg ctggatgatc ctccagcgcg gggatctcat gctggagttc
2821 ttcgcccacc ccaacttgtt tattgcagct tataatggtt acaaataaag caatagcatc
2881 acaaatttca caaataaagc attttttttca ctgcattcta gttgtggttt gtccaaactc
2941 atcaatgtat cttatcatgt ctgtataccg tcgacctcta gctagagctt ggcgtaatca
3001 tggtcatagc tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacga
3061 gccggaagca taaagtgtaa agcctggggt gcctaatgag tgagctaact cacattaatt
3121 gcgttgcgct cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga
3181 atcggccaac gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc
3241 actgactcgc tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg
3301 gtaatacggt tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc
3361 cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc
3421 ccccctgacg agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga
3481 ctataaagat accaggcgtt tcccccctgga agctccctcg tgcgctctcc tgttccgacc
3541 ctgccgctta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat
3601 agctcacgct gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg
3661 cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc
3721 aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga
3781 gcgaggtatg taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact
3841 agaaggacag tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt
3901 ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag
3961 cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg
4021 tctgacgctc agtggaacga aaactcacgt taagggattt tggtcatgag attatcaaaa
4081 aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata
4141 tatgagtaaa cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg
4201 atctgtctat ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata
4261 cgggagggct taccatctgg ccccagtgct gcaatgatac cgcgagaccc acgctcaccg
4321 gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct
4381 gcaactttat ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt
```

```
     4441 tcgccagtta atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc
     4501 tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga
     4561 tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt
     4621 aagttggccg cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc
     4681 atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa
     4741 tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca
     4801 catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca
     4861 aggatcttac cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct
     4921 tcagcatctt ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc
     4981 gcaaaaaagg gaataagggc gacacggaaa tgttgaatac tcatactctt ccttttcaa
     5041 tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt
     5101 tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc
//
```

9.2.2 UbcP-mCherry-2TA-Puro-WPRE bidirectional vector

```
LOCUS       recue_pLenti_Ubc        9763 bp ds-DNA     circular     27-JAN-2021
COMMENT
COMMENT
COMMENT     ApEinfo:methylated:1
FEATURES             Location/Qualifiers
     promoter        complement(2619..3235)
                     /locus_tag="CMV_promoter"
                     /ApEinfo_fwdcolor="#ff0000"
                     /ApEinfo_revcolor="#ff80ff"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     source          1..2211
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
                     /locus_tag="source:synthetic DNA construct"
                     /ApEinfo_fwdcolor="pink"
                     /ApEinfo_revcolor="pink"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
                     /ApEinfo_hidden
     source          complement(5115..5473)
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
                     /locus_tag="source:synthetic DNA construct(1)"
                     /ApEinfo_label="source:synthetic DNA construct"
                     /ApEinfo_fwdcolor="pink"
                     /ApEinfo_revcolor="pink"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
                     /ApEinfo_hidden
     primer_bind     complement(2589..2608)
                     /locus_tag="T7"
                     /ApEinfo_fwdcolor="cyan"
                     /ApEinfo_revcolor="green"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     source          3897..5108
                     /organism="synthetic DNA construct"
                     /mol_type="other DNA"
                     /locus_tag="source:synthetic DNA construct(2)"
                     /ApEinfo_label="source:synthetic DNA construct"
                     /ApEinfo_fwdcolor="pink"
```

```
                /ApEinfo_revcolor="pink"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
protein_bind    113..134
                /bound_moiety="E. coli catabolite activator protein"
                /note="CAP binding activates transcription in the presence
                of cAMP."
                /locus_tag="CAP binding site"
                /ApEinfo_fwdcolor="pink"
                /ApEinfo_revcolor="pink"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
source          complement(2212..2217)
                /organism="synthetic DNA construct"
                /mol_type="other DNA"
                /locus_tag="source:synthetic DNA construct(3)"
                /ApEinfo_label="source:synthetic DNA construct"
                /ApEinfo_fwdcolor="pink"
                /ApEinfo_revcolor="pink"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
promoter        149..179
                /note="promoter for the E. coli lac operon"
                /locus_tag="lac promoter"
                /ApEinfo_fwdcolor="#346ee0"
                /ApEinfo_revcolor="#346ee0"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
misc_feature    2446..2448
                /locus_tag="BGH_rev_primer"
                /ApEinfo_fwdcolor="#7eff74"
                /ApEinfo_revcolor="#7eff74"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
protein_bind    187..203
                /bound_moiety="lac repressor encoded by lacI"
                /note="The lac repressor binds to the lac operator to
                inhibit transcription in E. coli. This inhibition can be
                relieved by adding lactose or
                isopropyl-beta-D-thiogalactopyranoside (IPTG)."
                /locus_tag="lac operator"
                /ApEinfo_fwdcolor="pink"
                /ApEinfo_revcolor="pink"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     192..214
                /note="In lacZ gene"
                /locus_tag="M13/pUC Reverse"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     211..227
```

```
                /note="common sequencing primer, one of multiple similar
                variants"
                /locus_tag="M13 rev"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
     primer_bind     211..227
                /note="In lacZ gene. Also called M13-rev"
                /locus_tag="M13 Reverse"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
     promoter        complement(2218..2445)
                /locus_tag="bGH-polyA(signal)"
                /ApEinfo_fwdcolor="#ff0000"
                /ApEinfo_revcolor="#ff0000"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     CDS             5474..5831
                /codon_start=1
                /product="monomeric derivative of DsRed fluorescent
                protein (Shaner et al., 2004)"
                /note="mammalian codon-optimized"
                /translation="MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYE
                G
                TQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMN
                F
                EDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGAL
                K
                GEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYER
                A EGRHSTGGMDELYK"
                /locus_tag="mCherry"
                /ApEinfo_fwdcolor="#ff0000"
                /ApEinfo_revcolor="#e9d024"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
     primer_bind     246..266
                /note="T3 promoter, forward primer"
                /locus_tag="T3"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
     promoter        248..266
                /note="promoter for bacteriophage T3 RNA polymerase"
                /locus_tag="T3 promoter"
                /ApEinfo_fwdcolor="#346ee0"
                /ApEinfo_revcolor="#346ee0"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
     promoter        294..520
                /note="Rous sarcoma virus enhancer/promoter"
```

```
                /locus_tag="RSV promoter"
                /ApEinfo_fwdcolor="#346ee0"
                /ApEinfo_revcolor="#346ee0"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     primer_bind     5537..5556
                /note="mCherry, forward primer"
                /locus_tag="mCherry-F"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
     LTR             521..701
                /note="truncated 5' long terminal repeat (LTR) from HIV-1"
                /locus_tag="5' LTR (truncated)"
                /ApEinfo_fwdcolor="#86f71d"
                /ApEinfo_revcolor="#d7336f"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     primer_bind     5734..5754
                /note="DsRed1, forward primer"
                /locus_tag="DsRed1-C"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
     misc_feature    748..873
                /note="packaging signal of human immunodeficiency virus
                type 1"
                /locus_tag="HIV-1 Psi"
                /ApEinfo_fwdcolor="#7eff74"
                /ApEinfo_revcolor="#7eff74"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     CDS             5838..5891
                /codon_start=1
                /product="2A peptide from Thosea asigna virus capsid
                protein"
                /note="Eukaryotic ribosomes fail to insert a peptide bond
                between the Gly and Pro residues, yielding separate
                polypeptides."
                /translation="EGRGSLLTCGDVEENPGP"
                /locus_tag="T2A"
                /ApEinfo_fwdcolor="#ff0080"
                /ApEinfo_revcolor="#e9d024"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     misc_feature    1366..1599
                /note="The Rev response element (RRE) of HIV-1 allows for
                Rev-dependent mRNA export from the nucleus to the
                cytoplasm."
                /locus_tag="RRE"
                /ApEinfo_fwdcolor="#7eff74"
                /ApEinfo_revcolor="#7eff74"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
     CDS             5895..6491
```

```
                /codon_start=1
                /gene="pac from Streptomyces alboniger"
                /product="puromycin N-acetyltransferase"
                /note="confers resistance to puromycin"
                /translation="TEYKPTVRLATRDDVPRAVRTLAAAFADYPATRHTVDPDRHIER
                V
                TELQELFLTRVGLDIGKVWVADDGAAVAVWTTPESVEAGAVFAEIGPRMAELSGSRLA
                A
                QQQMEGLLAPHRPKEPAWFLATVGVSPDHQGKGLGSAVVLPGVEAAERAGVPAFLETS
                A PRNLPFYERLGFTVTADVEVPEGPRTWCMTRKPGA"
                /locus_tag="PuroR"
                /ApEinfo_fwdcolor="#e9d024"
                /ApEinfo_revcolor="#e9d024"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
misc_feature    2095..2211
                /note="central polypurine tract and central termination
                sequence of HIV-1"
                /locus_tag="cPPT/CTS"
                /ApEinfo_fwdcolor="#7eff74"
                /ApEinfo_revcolor="#7eff74"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
primer_bind     6388..6408
                /note="Puromycin resistance gene, forward primer"
                /locus_tag="Puro-F"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
LTR             6575..6808
                /note="self-inactivating 3' long terminal repeat (LTR)
                from HIV-1"
                /locus_tag="3' LTR (Delta-U3)"
                /ApEinfo_fwdcolor="#86f71d"
                /ApEinfo_revcolor="#d7336f"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
polyA_signal    6880..7001
                /note="SV40 polyadenylation signal"
                /locus_tag="SV40 poly(A) signal"
                /ApEinfo_fwdcolor="#ff3eee"
                /ApEinfo_revcolor="#ff3eee"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
primer_bind     complement(6917..6936)
                /note="SV40 polyA, reverse primer"
                /locus_tag="SV40pA-R"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     6971..6990
                /note="SV40 polyA terminator, reverse primer"
                /locus_tag="EBV-rev"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
```

```
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
misc_feature    3256..3844
                /note="woodchuck hepatitis virus posttranscriptional
                regulatory element"
                /locus_tag="WPRE"
                /ApEinfo_fwdcolor="#7eff74"
                /ApEinfo_revcolor="#7eff74"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
rep_origin      7041..7176
                /note="SV40 origin of replication"
                /locus_tag="SV40 ori"
                /ApEinfo_fwdcolor="#999999"
                /ApEinfo_revcolor="#999999"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     complement(3309..3329)
                /note="WPRE, reverse primer"
                /locus_tag="WPRE-R"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     7103..7122
                /note="SV40 promoter/origin, forward primer"
                /locus_tag="SV40pro-F"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
CDS             complement(3727..3738)
                /codon_start=1
                /product="Factor Xa recognition and cleavage site"
                /translation="IEGR"
                /locus_tag="Factor Xa site"
                /ApEinfo_fwdcolor="#e9d024"
                /ApEinfo_revcolor="#e9d024"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     complement(7126..7137)
                /note="SV40 promoter/origin, forward primer"
                /locus_tag="SV40pro-F(1)"
                /ApEinfo_label="SV40pro-F"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     complement(7196..7215)
                /note="T7 promoter, forward primer"
                /locus_tag="T7(1)"
                /ApEinfo_label="T7"
                /ApEinfo_fwdcolor="#14c0bd"
```

```
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
promoter        complement(7197..7215)
                /note="promoter for bacteriophage T7 RNA polymerase"
                /locus_tag="T7 promoter"
                /ApEinfo_fwdcolor="#346ee0"
                /ApEinfo_revcolor="#346ee0"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     complement(7225..7242)
                /note="In lacZ gene. Also called M13-F20 or M13 (-21)
                Forward"
                /locus_tag="M13 Forward"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     complement(7225..7241)
                /note="common sequencing primer, one of multiple similar
                variants"
                /locus_tag="M13 fwd"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     complement(7234..7256)
                /note="In lacZ gene"
                /locus_tag="M13/pUC Forward"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
rep_origin      7383..7838
                /direction=RIGHT
                /note="f1 bacteriophage origin of replication; arrow
                indicates direction of (+) strand synthesis"
                /locus_tag="f1 ori"
                /ApEinfo_fwdcolor="#999999"
                /ApEinfo_revcolor="#999999"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     complement(7470..7489)
                /note="F1 origin, reverse primer"
                /locus_tag="F1ori-R"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     7680..7701
                /note="F1 origin, forward primer"
                /locus_tag="F1ori-F"
```

```
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
promoter        7864..7968
                /gene="bla"
                /locus_tag="AmpR promoter"
                /ApEinfo_fwdcolor="#346ee0"
                /ApEinfo_revcolor="#346ee0"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
CDS             7969..8829
                /codon_start=1
                /gene="bla"
                /product="beta-lactamase"
                /note="confers resistance to ampicillin, carbenicillin,
                and related antibiotics"
                /translation="MSIQHFRVALIPFFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
                I
                ELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEY
                S
                PVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDR
                W
                EPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVAGPLLRS
                A
                LPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIGA
                S LIKHW"
                /locus_tag="AmpR"
                /ApEinfo_fwdcolor="#e9d024"
                /ApEinfo_revcolor="#e9d024"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     complement(8187..8206)
                /note="Ampicillin resistance gene, reverse primer"
                /locus_tag="Amp-R"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
rep_origin      9000..9588
                /direction=RIGHT
                /note="high-copy-number ColE1/pMB1/pBR322/pUC origin of
                replication"
                /locus_tag="ori"
                /ApEinfo_fwdcolor="#999999"
                /ApEinfo_revcolor="#999999"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     9489..9508
                /note="pBR322 origin, forward primer"
                /locus_tag="pBR322ori-F"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
```

```
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     9742..9759
                /note="L4440 vector, forward primer"
                /locus_tag="L4440"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
source          5109..5114
                /organism="synthetic DNA construct"
                /mol_type="other DNA"
                /locus_tag="source:synthetic DNA construct(4)"
                /ApEinfo_label="source:synthetic DNA construct"
                /ApEinfo_fwdcolor="pink"
                /ApEinfo_revcolor="pink"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
promoter        3897..5108
                /note="human ubiquitin C promoter"
                /locus_tag="UbC promoter"
                /ApEinfo_fwdcolor="#346ee0"
                /ApEinfo_revcolor="#346ee0"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
primer_bind     4945..4964
                /note="Human Ubiquitin C (UbC) promoter, forward primer"
                /locus_tag="hUBCpro-F"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
source          6498..9763
                /organism="synthetic DNA construct"
                /mol_type="other DNA"
                /locus_tag="source:synthetic DNA construct(5)"
                /ApEinfo_label="source:synthetic DNA construct"
                /ApEinfo_fwdcolor="pink"
                /ApEinfo_revcolor="pink"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     complement(5319..5333)
                /note="DsRed1, reverse primer"
                /locus_tag="DsRed1-N"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
primer_bind     complement(5269..5287)
                /note="mCherry, reverse primer"
                /locus_tag="mCherry-R"
                /ApEinfo_fwdcolor="#14c0bd"
                /ApEinfo_revcolor="#4ec02b"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
```

```
                width 5 offset 0"
                /ApEinfo_hidden
source          complement(3891..3896)
                /organism="synthetic DNA construct"
                /mol_type="other DNA"
                /locus_tag="source:synthetic DNA construct(6)"
                /ApEinfo_label="source:synthetic DNA construct"
                /ApEinfo_fwdcolor="pink"
                /ApEinfo_revcolor="pink"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
misc_feature    3891..3896
                /locus_tag="ClaI"
                /ApEinfo_fwdcolor="cyan"
                /ApEinfo_revcolor="green"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
misc_feature    3244..3267
                /locus_tag="New Feature"
                /ApEinfo_fwdcolor="cyan"
                /ApEinfo_revcolor="green"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
misc_feature    2212..2217
                /locus_tag="ClaI(1)"
                /ApEinfo_label="ClaI"
                /ApEinfo_fwdcolor="cyan"
                /ApEinfo_revcolor="green"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
misc_feature    3845..3868
                /locus_tag="New Feature(1)"
                /ApEinfo_label="New Feature"
                /ApEinfo_fwdcolor="cyan"
                /ApEinfo_revcolor="green"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
misc_feature    4945..4964
                /locus_tag="#2772"
                /ApEinfo_fwdcolor="cyan"
                /ApEinfo_revcolor="green"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
misc_feature    6208..6226
                /locus_tag="#996"
                /ApEinfo_fwdcolor="cyan"
                /ApEinfo_revcolor="green"
                /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0"
                /ApEinfo_hidden
misc_feature    5124..5831
                /locus_tag="mCherry(1)"
                /ApEinfo_label="mCherry"
```

```
                     /ApEinfo_fwdcolor="#ff0000"
                     /ApEinfo_revcolor="green"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
     misc_feature    2468..2473
                     /locus_tag="LIC_cloning_site"
                     /ApEinfo_fwdcolor="cyan"
                     /ApEinfo_revcolor="green"
                     /ApEinfo_graphicformat="arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0"
ORIGIN
        1 cggaagagcg cccaatacgc aaaccgcctc tccccgcgcg ttggccgatt cattaatgca
       61 gctggcacga caggtttccc gactggaaag cgggcagtga gcgcaacgca attaatgtga
      121 gttagctcac tcattaggca ccccaggctt tacactttat gcttccggct cgtatgttgt
      181 gtggaattgt gagcggataa caatttcaca caggaaacag ctatgaccat gattacgcca
      241 agcgcgcaat taaccctcac taaagggaac aaaagctgga gctgcaagct taatgtagtc
      301 ttatgcaata ctcttgtagt cttgcaacat ggtaacgatg agttagcaac atgccttaca
      361 aggagagaaa aagcaccgtg catgccgatt ggtggaagta aggtggtacg atcgtgcctt
      421 attaggaagg caacagacgg gtctgacatg gattggacga accactgaat tgccgcattg
      481 cagagatatt gtatttaagt gcctagctcg atacataaac gggtctctct ggttagacca
      541 gatctgagcc tgggagctct ctggctaact agggaaccca ctgcttaagc ctcaataaag
      601 cttgccttga gtgcttcaag tagtgtgtgc ccgtctgttg tgtgactctg gtaactagag
      661 atccctcaga ccctttttagt cagtgtggaa aatctctagc agtggcgccc gaacagggac
      721 ttgaaagcga aagggaaacc agaggagctc tctcgacgca ggactcggct tgctgaagcg
      781 cgcacggcaa gaggcgaggg gcggcgactg gtgagtacgc caaaaatttt gactagcgga
      841 ggctagaagg agagagatgg gtgcgagagc gtcagtatta agcgggggag aattagatcg
      901 cgatgggaaa aaattcggtt aaggccaggg ggaaagaaaa aatataaatt aaaacatata
      961 gtatgggcaa gcagggagct agaacgattc gcagttaatc ctggcctgtt agaaacatca
     1021 gaaggctgta gacaaatact gggacagcta caaccatccc ttcagacagg atcagaagaa
     1081 cttagatcat tatataatac agtagcaacc ctctattgtg tgcatcaaag gatagagata
     1141 aaagacacca aggaagcttt agacaagata gaggaagagc aaaacaaaag taagaccacc
     1201 gcacagcaag cggccgctga tcttcagacc tggaggagga gatatgaggg acaattggag
     1261 aagtgaatta tataaatata aagtagtaaa aattgaacca ttaggagtag cacccaccaa
     1321 ggcaaagaga agagtggtgc agagagaaaa aagagcagtg ggaataggag ctttgttcct
     1381 tgggttcttg ggagcagcag gaagcactat gggcgcagcg tcaatgacgc tgacggtaca
     1441 ggccagacaa ttattgtctg gtatagtgca gcagcagaac aatttgctga gggctattga
     1501 ggcgcaacag catctgttgc aactcacagt ctggggcatc aagcagctcc aggcaagaat
     1561 cctggctgtg gaaagatacc taaaggatca acagctcctg gggatttggg gttgctctgg
     1621 aaaactcatt tgcaccactg ctgtgccttg gaatgctagt tggagtaata aatctctgga
     1681 acagatttgg aatcacacga cctggatgga gtgggacaga gaaattaaca attacacaag
     1741 cttaatacac tccttaattg aagaatcgca aaaccagcaa gaaaagaatg aacaagaatt
     1801 attggaatta gataaatggg caagtttgtg gaattggttt aacataacaa attggctgtg
     1861 gtatataaaa ttattcataa tgatagtagg aggcttggta ggtttaagaa tagtttttgc
     1921 tgtactttct atagtgaata gagttaggca gggatattca ccattatcgt ttcagaccca
     1981 cctcccaacc ccgaggggac ccgacaggcc cgaaggaata gaagaagaag gtggagagag
     2041 agacagagac agatccattc gattagtgaa cggatctcga cggtatcggt taacttttaa
     2101 aagaaaaggg gggattgggg ggtacagtgc aggggaaaga atagtagaca taatagcaac
     2161 agacatacaa actaaagaat tacaaaaaca aattacaaaa ttcaaaattt tatcgatCCA
     2221 TAGAGCCCAC CGCATCCCCA GCATGCCTGC TATTGTCTTC CCAATCCTCC CCCTTGCTGT
     2281 CCTGCCCCAC CCCACCCCCC AGAATAGAAT GACACCTACT CAGACAATGC GATGCAATTT
     2341 CCTCATTTTA TTAGGAAAGG ACAGTGGGAG TGGCACCTTC CAGGGTCAAG GAAGGCACGG
     2401 GGGAGGGGCA AACAACAGAT GGCTGGCAAC TAGAAGGCAC AGTCGAGGCT GATCAGCGGG
     2461 TTTAAACGGG CCCTCTAGAC TCGAGCGGCC GCCACTGTGC TGGATATCTG CAGAATTCCA
     2521 CCACACTGGA CTAGTGGATC CGAGCTCGGT ACCAAGCTTA AGTTTAAACG CTAGCCAGCT
     2581 TGGGTCTCCC CTATAGTGAG TCGTATTAAT TTCGATAAGC CAGTAAGCAG TGGGTTCTCT
     2641 AGTTAGCCAG AGAGCTCTGC TTATATAGAC CTCCCACCGT ACACGCCTAC CGCCCATTTG
     2701 CGTCAATGGG GCGGAGTTGT TACGACATTT TGGAAAGTCC CGTTGATTTT GGTGCCAAAA
     2761 CAAACTCCCA TTGACGTCAA TGGGGTGGAG ACTTGGAAAT CCCCGTGAGT CAAACCGCTA
     2821 TCCACGCCCA TTGATGTACT GCCAAAACCG CATCACCATG GTAATAGCGA TGACTAATAC
```

```
2881 GTAGATGTAC TGCCAAGTAG GAAAGTCCCA TAAGGTCATG TACTGGGCAT AATGCCAGGC
2941 GGGCCATTTA CCGTCATTGA CGTCAATAGG GGGCGTACTT GGCATATGAT ACACTTGATG
3001 TACTGCCAAG TGGGCAGTTT ACCGTAAATA CTCCACCCAT TGACGTCAAT GGAAAGTCCC
3061 TATTGGCGTT ACTATGGGAA CATACGTCAT TATTGACGTC AATGGGCGGG GGTCGTTGGG
3121 CGGTCAGCCA GGCGGGCCAT TTACCGTAAG TTATGTAACG CGGAACTCCA TATATGGGCT
3181 ATGAACTAAT GACCCCGTAA TTGATTACTA TTAATAACTA GTCAATAATC AATGTCAACG
3241 CGTtcgccag tcgacaatca acctctggat tacaaaattt gtgaaagatt gactggtatt
3301 cttaactatg ttgctccttt tacgctatgt ggatacgctg ctttaatgcc tttgtatcat
3361 gctattgctt cccgtatggc tttcattttc tcctccttgt ataaatcctg gttgctgtct
3421 ctttatgagg agttgtggcc cgttgtcagg caacgtggcg tggtgtgcac tgtgtttgct
3481 gacgcaaccc ccactggttg gggcattgcc accacctgtc agctcctttc cgggactttc
3541 gctttccccc tccctattgc cacggcggaa ctcatcgccg cctgccttgc ccgctgctgg
3601 acaggggctc ggctgttggg cactgacaat tccgtggtgt tgtcggggaa gctgacgtcc
3661 tttccatggc tgctcgcctg tgttgccacc tggattctgc gcgggacgtc cttctgctac
3721 gtcccttcgg ccctcaatcc agcggacctt ccttcccgcg gcctgctgcc ggctctgcgg
3781 cctcttccgc gtcttcgcct tcgccctcag acgagtcgga tctccctttg ggccgcctcc
3841 ccgcctggaa ttcgagctcg gtacctttAC GCGTATATCT GGCCCGTACA atcgatggcc
3901 tccgcgccgg gttttggcgc ctcccgcggg cgccccctc ctcacggcga gcgctgccac
3961 gtcagacgaa gggcgcagcg agcgtcctga tccttccgcc cggacgctca ggacagcggc
4021 ccgctgctca taagactcgg ccttagaacc ccagtatcag cagaaggaca ttttaggacg
4081 ggacttgggt gactctaggg cactggtttt ctttccagag agcggaacag gcgaggaaaa
4141 gtagtccctt ctcggcgatt ctgcggaggg atctccgtgg ggcggtgaac gccgatgatt
4201 atataaggac gcgccgggtg tggcacagct agttccgtcg cagccgggat ttgggtcgcg
4261 gttcttgttt gtggatcgct gtgatcgtca cttggtgagt agcgggctgc tgggctggcc
4321 ggggctttcg tggccgccgg gccgctcggt gggacggaag cgtgtggaga gaccgccaag
4381 ggctgtagtc tgggtccgcg agcaaggttg ccctgaactg ggggttgggg ggagcgcagc
4441 aaaatggcgg ctgttcccga gtcttgaatg gaagacgctt gtgaggcggg ctgtgaggtc
4501 gttgaaacaa ggtgggggcc atggtgggcg gcaagaaccc aaggtcttga ggccttcgct
4561 aatgcgggaa agctcttatt cgggtgagat gggctggggc accatctggg gaccctgacg
4621 tgaagtttgt cactgactgg agaactcggt ttgtcgtctg ttgcgggggc ggcagttatg
4681 gcggtgccgt tgggcagtgc acccgtacct ttgggagcgc gcgccctcgt cgtgtcgtga
4741 cgtcacccgt tctgttggct tataatgcag ggtggggcca cctgccggta ggtgtgcggt
4801 aggcttttct ccgtcgcagg acgcagggtt cgggcctagg gtaggctctc ctgaatcgac
4861 aggcgccgga cctctggtga ggggagggat aagtgaggcg tcagtttctt tggtcggttt
4921 tatgtaccta tcttcttaag tagctgaagc tccggttttg aactatgcgc tcggggttgg
4981 cgagtgtgtt ttgtgaagtt ttttaggcac cttttgaaat gtaatcattt gggtcaatat
5041 gtaattttca gtgttagact agtaaattgt ccgctaaatt ctggccgttt ttggcttttt
5101 tgttagacgg atccgccgcc accatggtga gcaagggcga ggaggataac atggccatca
5161 tcaaggagtt catgcgcttc aaggtgcaca tggagggctc cgtgaacggc cacgagttcg
5221 agatcgaggg cgagggcgag ggccgcccct acgagggcac ccagaccgcc aagctgaagg
5281 tgaccaaggg tggccccctg cccttcgcct gggacatcct gtcccctcag ttcatgtacg
5341 gctccaaggc ctacgtgaag caccccgccg acatccccga ctacttgaag ctgtccttcc
5401 ccgagggctt caagtgggag cgcgtgatga acttcgagga cggcggcgtg gtgaccgtga
5461 cccaggactc ctccctgcag gacggcgagt tcatctacaa ggtgaagctg cgcggcacca
5521 acttcccctc cgacggcccc gtaatgcaga agaagaccat gggctgggag gcctcctccg
5581 agcggatgta ccccgaggac ggcgccctga agggcgagat caagcagagg ctgaagctga
5641 aggacggcgg ccactacgac gctgaggtca agaccaccta caaggccaag aagcccgtgc
5701 agctgcccgg cgcctacaac gtcaacatca agttggacat cacctcccac aacgaggact
5761 acaccatcgt ggaacagtac gaacgcgccg agggccgcca ctccaccggc ggcatggacg
5821 agctgtacaa gccccgggag ggcagaggaa gtcttctaac atgcggtgac gtggaggaga
5881 atcccggccc tcgaaccgag tacaagccca cggtgcgcct cgccacccgc gacgacgtcc
5941 ccagggccgt acgcaccctc gccgccgcgt tcgccgacta ccccgccacg cgccacaccg
6001 tcgatccgga ccgccacatc gagcgggtca ccgagctgca agaactcttc ctcacgcgcg
6061 tcgggctcga catcggcaag gtgtgggtcg cggacgacgg cgccgcggtg gcggtctgga
6121 ccacgccgga gagcgtcgaa gcgggggcgg tgttcgccga gatcggcccg cgcatggccg
6181 agttgagcgg ttcccggctg gccgcgcagc aacagatgga aggcctcctg gcgccgcacc
6241 ggcccaagga gcccgcgtgg ttcctggcca ccgtcggcgt ctcgcccgac caccagggca
6301 agggtctggg cagcgccgtc gtgctccccg gagtggaggc ggccgagcgc gccggggtgc
6361 ccgccttcct ggagacctcc gcgccccgca acctcccctt ctacgagcgg ctcggcttca
```

```
6421 ccgtcaccgc cgacgtcgag gtgcccgaag gaccgcgcac ctggtgcatg acccgcaagc
6481 ccggtgccta agaattcgag ctcggtacct ttaagaccaa tgacttacaa ggcagctgta
6541 gatcttagcc acttttaaa agaaaagggg ggactggaag ggctaattca ctcccaacga
6601 agacaagatc tgctttttgc ttgtactggg tctctctggt tagaccagat ctgagcctgg
6661 gagctctctg gctaactagg gaacccactg cttaagcctc aataaagctt gccttgagtg
6721 cttcaagtag tgtgtgcccg tctgttgtgt gactctggta actagagatc cctcagaccc
6781 ttttagtcag tgtggaaaat ctctagcagt agtagttcat gtcatcttat tattcagtat
6841 ttataacttg caaagaaatg aatatcagag agtgagagga acttgtttat tgcagcttat
6901 aatggttaca aataaagcaa tagcatcaca aatttcacaa ataaagcatt tttttcactg
6961 cattctagtt gtggtttgtc caaactcatc aatgtatctt atcatgtctg gctctagcta
7021 tcccgccccc aactccgccc atcccgcccc taactccgcc cagttccgcc cattctccgc
7081 cccatggctg actaattttt tttatttatg cagaggccga ggccgcctcg gcctctgagc
7141 tattccagaa gtagtgagga ggcttttttg gaggcctagg gacgtaccca attcgcccta
7201 tagtgagtcg tattacgcgc gctcactggc cgtcgtttta caacgtcgtg actgggaaaa
7261 ccctggcgtt acccaactta atcgccttgc agcacatccc cctttcgcca gctggcgtaa
7321 tagcgaagag gcccgcaccg atcgcccttc ccaacagttg cgcagcctga atggcgaatg
7381 ggacgcgccc tgtagcggcg cattaagcgc ggcgggtgtg gtggttacgc gcagcgtgac
7441 cgctacactt gccagcgccc tagcgcccgc tcctttcgct ttcttccctt cctttctcgc
7501 cacgttcgcc ggctttcccc gtcaagctct aaatcggggg ctccctttag ggttccgatt
7561 tagtgcttta cggcacctcg accccaaaaa acttgattag ggtgatggtt cacgtagtgg
7621 gccatcgccc tgatagacgg tttttcgccc tttgacgttg gagtccacgt tctttaatag
7681 tggactcttg ttccaaactg gaacaacact caaccctatc tcggtctatt cttttgattt
7741 ataagggatt ttgccgattt cggcctattg gttaaaaaat gagctgattt aacaaaaatt
7801 taacgcgaat tttaacaaaa tattaacgct tacaatttag gtggcacttt tcggggaaat
7861 gtgcgcggaa cccctatttg tttatttttc taaatacatt caaatatgta tccgctcatg
7921 agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat gagtattcaa
7981 catttccgtg tcgcccttat tccctttttt gcggcatttt gccttcctgt ttttgctcac
8041 ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac
8101 atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga agaacgtttt
8161 ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccg tattgacgcc
8221 gggcaagagc aactcggtcg ccgcatacac tattctcaga atgacttggt tgagtactca
8281 ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg cagtgctgcc
8341 ataaccatga gtgataacac tgcggccaac ttacttctga caacgatcgg aggaccgaag
8401 gagctaaccg ctttttgca caacatgggg gatcatgtaa ctcgccttga tcgttgggaa
8461 ccggagctga atgaagccat accaaacgac gagcgtgaca ccacgatgcc tgtagcaatg
8521 gcaacaacgt tgcgcaaact attaactggc gaactactta ctctagcttc ccggcaacaa
8581 ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc ggcccttccg
8641 gctggctggt ttattgctga taaatctgga gccggtgagc gtgggtctcg cggtatcatt
8701 gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt
8761 caggcaacta tggatgaacg aaatagacag atcgctgaga taggtgcctc actgattaag
8821 cattggtaac tgtcagacca agtttactca tatatacttt agattgattt aaaacttcat
8881 ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac caaaatccct
8941 taacgtgagt tttcgttcca ctgagcgtca gaccccgtag aaaagatcaa aggatcttct
9001 tgagatcctt tttttctgcg cgtaatctgc tgcttgcaaa caaaaaaacc accgctacca
9061 gcggtggttt gtttgccgga tcaagagcta ccaactcttt ttccgaaggt aactggcttc
9121 agcagagcgc agataccaaa tactgttctt ctagtgtagc cgtagttagg ccaccacttc
9181 aagaactctg tagcaccgcc tacatacctc gctctgctaa tcctgttacc agtggctgct
9241 gccagtggcg ataagtcgtg tcttaccggg ttggactcaa gacgatagtt accggataag
9301 gcgcagcggt cgggctgaac gggggttcg tgcacacagc ccagcttgga gcgaacgacc
9361 tacaccgaac tgagatacct acagcgtgag ctatgagaaa gcgccacgct tcccgaaggg
9421 agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa caggagagcg cacgagggag
9481 cttccagggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca cctctgactt
9541 gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc tatggaaaaa cgccagcaac
9601 gcggcctttt tacggttcct ggccttttgc tggccttttg ctcacatgtt ctttcctgcg
9661 ttatcccctg attctgtgga taaccgtatt accgcctttg agtgagctga taccgctcgc
9721 cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg aag
//
```

9.2.3 RUS sequences

The following full-length lncRNAs and mutant sequences were cloned into pcDNA.5-FRT-5xMS2 & UbcP-mCherry-2TA-Puro-WPRE:

```
>full-length isoform 1
AGAGCATTTGGGCTTAAGCCGGTTCAGCTCAGAGGTGGGTTTCGGAAGTCTGTGGCGGAC
CACCGCAGCAAGTCATCACTCAGCGTACAAGCCTTTCCATTTCTGAAAATATAAAGAACT
TAAGCATGCGGCGTTCGGAGTGTAGATTGCTTTGAGGTATTTCAAAGCAGCGCTGATCAT
GGGCAGAGGAGCTGATATGTGCAACAGAAACTACTCACTGGAACTCTTCTGGACGTCTGT
CCTCTGTAATGCAAGTGTCTGCCCTCACTCTGGCGACTGTGCTGTTGGAACGCTGGTGCT
CATCTTTCACTGTCCTCCACAAAGCATGCCTCAGGACTGTAGGCAGCCTGTTTATGCCAT
TGTAAGACTACCCCTGGGAAGGAGGGACAGACCTAAAAAATTCACTAACAATTCTGACTA
AACTGACCTGTCTCCTAACATAGAGAAGATTGGCAGCCAGAAAGACCTGGATATCCCACC
TCCATTGCCCACTGCACTGTGGTTACAGCATGAAGCCATAACTGGCTCTTTAGTTCTATG
CCATGAAATTCAACTCTCATTTTCATGCTTGTGCACTAATCCTAAAAGTATCATATTCCC
AAATGGTTATTTTGTTTTCTGTATTTGCAAATACTTGTTTTCTTTTCTTAATGGAGTTTG
TAGGCCCATAAGTTACTTCTACTCAATTTTATCATTTCGATATGATTCTTTCATGTTCTG
AGGATTGGTTGGAAAAAGATCCTAGTCATTAGTGTACAGGTTCCAATCTCAAATATAAAA
TCTATACATGGAATCCTTACTAATTTTTTTTTACCTGTATAATTTCTTTACTATGGCTCC
TACCACCTTCACTTCATTTATGTTAAAATCAACAATATTATTCATATTAAAAACATATGC
TCTATAGCTCCC
>5'domain
AGAGCATTTGGGCTTAAGCCGGTTCAGCTCAGAGGTGGGTTTCGGAAGTCTGTGGCGGAC
CACCGCAGCAAGTCATCACTCAGCGTACAAGCCTTTCCATTTCTGAAAATATAAAGAACT
TAAGCATGCGGCGTTCGGAGTGTAGATTGCTTTGAGGTATTTCAAAGCAGCGCTGATCAT
GGGCAGAGGAGCTGATATGTGCAACAGAAACTACTCACTGGAACTCTTCTGGACGTCTGT
CCTCTGTAATGCAAGT
>deletion 5'
ACGTCTGTCCTCTGTAATGCAAGTGTCTGCCCTCACTCTGGCGACTGTGCTGTTGGAACG
CTGGTGCTCATCTTTCACTGTCCTCCACAAAGCATGCCTCAGGACTGTAGGCAGCCTGTT
TATGCCATTGTAAGACTACCCCTGGGAAGGAGGGACAGACCTAAAAAATTCACTAACAAT
TCTGACTAAACTGACCTGTCTCCTAACATAGAGAAGATTGGCAGCCAGAAAGACCTGGAT
ATCCCACCTCCATTGCCCACTGCACTGTGGTTACAGCATGAAGCCATAACTGGCTCTTTA
GTTCTATGCCATGAAATTCAACTCTCATTTTCATGCTTGTGCACTAATCCTAAAAGTATC
ATATTCCCAAATGGTTATTTTGTTTTCTGTATTTGCAAATACTTGTTTTCTTTTCTTAAT
GGAGTTTGTAGGCCCATAAGTTACTTCTACTCAATTTTATCATTTCGATATGATTCTTTC
ATGTTCTGAGGATTGGTTGGAAAAAGATCCTAGTCATTAGTGTACAGGTTCCAATCTCAA
ATATAAAATCTATACATGGAATCCTTACTAATTTTTTTTTACCTGTATAATTTCTTTACT
ATGGCTCCTACCACCTTCACTTCATTTATGTTAAAATCAACAATATTATTCATATTAAAA
ACATATGCTCTATAGCTCCC
>deletion 3'
AGAGCATTTGGGCTTAAGCCGGTTCAGCTCAGAGGTGGGTTTCGGAAGTCTGTGGCGGAC
CACCGCAGCAAGTCATCACTCAGCGTACAAGCCTTTCCATTTCTGAAAATATAAAGAACT
TAAGCATGCGGCGTTCGGAGTGTAGATTGCTTTGAGGTATTTCAAAGCAGCGCTGATCAT
GGGCAGAGGAGCTGATATGTGCAACAGAAACTACTCACTGGAACTCTTCTGGACGTCTGT
CCTCTGTAATGCAAGTGTCTGCCCTCACTCTGGCGACTGTGCTGTTGGAACGCTGGTGCT
CATCTTTCACTGTCCTCCACAAAGCATGCCTCAGGACTGTAGGCAGCCTGTTTATGCCAT
TGTAAGACTACCCCTGGGAAGGAGGGACAGACCTAAAAAATTCACTAACAATTCTGACTA
AACTGACCTGTCTCCTAA
```

9.3 Acknowledgement

Hiermit möchte ich mich bei Prof. Dr. Dr. Johanna Scheuermann für die Vergabe dieser Arbeit und für all die Freiheiten bedanken, sich an neuen Methoden auszuprobieren. Ich möchte ihr meine besten Genesungswünsche ausrichten. Ich hoffe, es geht dir bald wieder besser!

Mein besonderer Dank gilt der LMU, die mich in dieser herausfordernden Situation nicht im Stich gelassen hat. Insbesondere möchte ich mich an dieser Stelle bei Prof. Dr. Dr. Christian Haass und Prof. Dr. Peter Becker bedanken. Vielen Dank, dass Sie mich in dieser Situation nicht nur hilfreich beraten haben, sondern mir auch tatkräftig aus der Not geholfen haben. Ich halte das in höchsten Ehren und mir ist bewusst, dass das nicht selbstverständlich ist. Vielen Dank Herr Prof. Dr. Christian Haass, dass Sie für mich als neuer Doktorvater eingesprungen sind und mir die Möglichkeit gegeben haben, die Doktorarbeit erfolgreich zu Ende zu bringen. Vielen Dank Herr Prof. Dr. Peter Becker, dass Sie mich in der Schlussphase intensiv beraten und Korrektur gelesen haben.

Einen besonderen Dank möchte ich ebenso unserer technischen Assistentin Veronika Müller aussprechen: Dein Talent für preparative Arbeiten, deine Hingabe und dein Fleiß haben diese Arbeit erst möglich gemacht. Weiter möchte ich den ehemaligen Mitgliedern der AG-Scheuermann danken: Danke Dr. Antje Fischer und Dr. Miriam Vogt für die Einführung in qantitative RT-PCR Analysen, Fluoreszenzmikroskopie und Zellkulturarbeiten sowie die Versorgung von Maus-Gewebe. Ebenso ein Dank an Matin Moschref und Matina-Jasemi Loukeri, die mich im Rahmen eines Praktikums unterstützt haben.
Vielen Dank an Prof. Dr. Stefan Lichtentahler, Dr. Stephan Müller und Anna Berghofer für die Probenvorbereitungen und die LC-MS Messungen.
Many thanks to the EMBL-Gene core facility: Dr. Vladimir Benes and Bianka Baying for sample preparation and high throughput sequencing.
Vielen Dank an Dr. Tobias Straub für die Beratung bei NGS-Datenanalysen und für das zur Verfügungstellen von R-Skripten, pipelines und Referenzgenomen.
I thank the research group of Dr. Sabine Tahirovic for providing their covaris M220, the research group of Prof. Dr. Dieter Edbauer for providing their BioRad CFX 384 qPCR system and UV crosslinker, and the research group of Prof. Dr. Harald Steiner for providing their UV-lamp. I thank Prof. Dr. Peter Becker's group for gifting the plasmids: pMal-MS2MBP and AdMl3. I thank the entire research department of Prof. Dr. Christian Haass for advising me and sharing reagents. Vielen Dank an Dr. Frits Kamp und Brigitte Nuscher für CD-Messungen (nicht Teil der Dissertation). Vielen Dank an Dr. Anja Capell für die Einführung in die konfokale Mikroskopie. Vielen Dank an Gabrielle Basset für die freundliche Einführung in den Branson-Sonicator.
Vielen Dank an Sabine Odoy für die Bestellung wichtiger Reagenzien und benötigter Verbrauchsmaterialien.

I thank all my TAC members: Prof. Dr. Dr. Johanna Scheuermann, Prof. Dr. Peter Becker, Prof. Dr. Dieter Edbauer, and Prof. Dr. Jovica Ninkovic, for scientific advice and fruitful discussions.
I thank the IMPRS-LS school and coordination office for admission, the kind entrance, and the offer of all the helpful courses.
Zu guter Letzt möchte ich vor allem auch meiner Familie danken. Vielen Dank Mama und Papa, dass ihr mir das Studium ermöglicht habt. Vielen Dank an meine Schwestern Ines Preuß, Ivonne Wehrl, Isabelle Schneider, Jasmin Schmidt und an meine Frau Djana Schneider für eure Unterstützung. Es ist schön, sich auf seine Familie immer voll und ganz verlassen zu können. Das hat mir in meiner Situation sehr geholfen und mir den notwendigen Rückhalt geboten. Vor allem vielen Dank an Isabelle Schneider und Wolfgang Wehrl für das Korrekturlesen und die hilfreichen Vorschläge. Danke Ivonne, dass du bei Bewerbungen immer geholfen hast. Vielen Dank an Djana und unseren Sohn Maximilian, dass ihr meine PHD Zeit so versüßt habt. Vielen Dank auch an meine Freunde für die anregenden Diskussionen und die moralische Unterstützung.

9.4 Affidavit

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN | Promotionsbüro Medizinische Fakultät | MMRS

Eidesstattliche Versicherung

Schneider, Marius

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

The long non-coding RNA "upstream to Slitrk3" (RUS) affects chromatin organization by binding to Brd2 and Smarca5.

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 24.03.2022
Ort, Datum

Marius F. Schneider
Unterschrift Doktorandin bzw. Doktorand