



UNIVERSITÄT  
BAYREUTH

---

# Studies on the role of enhancer RNAs in the release of RNA polymerase II from promoter-proximal pausing

---

DISSERTATION

Submitted to the Bayreuth Graduate School Mathematical  
and Natural Sciences (BayNAT) of the University of Bayreuth for  
obtaining the academic grade of

*Doctor rerum naturalium (Dr. rer. nat.)*

by

**Vladyslava Gorbovytska**

from Karlivka, Ukraine

Bayreuth 2021

This doctoral thesis was prepared at the department of Biochemistry IV – RNA Biochemistry (former "Gene regulation by non-coding RNA") at the University of Bayreuth from July 2016 until April 2021 and was supervised by Dr. Claus-D. Kuhn.

This is a full reprint of the thesis submitted to obtain the academic degree of Doctor of Natural Sciences (Dr. rer. nat.) and approved by the Bayreuth Graduate School of Mathematical and Natural Sciences (BayNAT) of the University of Bayreuth.

Date of submission: 21.05.2021

Date of defense: 21.07.2021

Acting director: Prof. Dr. Markus Lippitz

Doctoral committee:

Prof. Dr. Claus-D. Kuhn	(reviewer)
Prof. Dr. Klaus Ersfeld	(reviewer)
Prof. Dr. Birte Höcker	(chairman)
Prof. Dr. Olaf Stemmann	

# Author contributions

A manuscript encompassing significant parts of the results of this thesis are in revision at *Nature Communications* and available as a preprint on bioRxiv:

Vladyslava Gorbovytska, Seung-Kyoon Kim, Filiz Kuybu, Michael Götze, Dahun Um, Keunsoo Kang, Andreas Pittroff, Lisa-Marie Schneider, Alexander Leitner, Tae-Kyung Kim, Claus-D. Kuhn - Enhancer RNAs stimulate Pol II pause release by harnessing multivalent interactions to NELF. bioRxiv 2021.04.25.441328; doi: <https://doi.org/10.1101/2021.04.25.441328>.

This thesis contains data from other scientists, as stated in the following:

- Filiz Kuybu, B.Sc. (Master's thesis) – performed the EMSA experiments under my supervision (experiments were designed by me)
- Dr. Alexander Leitner and Dr. Michael Götze – performed RNA-protein UV-crosslinking mass spectrometry (comprising the crosslinking, data acquisition and analysis of spectra)
- Dr. Seung-Kyoon Kim (GRO-seq data)
- Andreas Pittroff, M.Sc. – contributed to the bioinformatical analysis of the Exo-seq data

These and further contributors to this work, are explicitly mentioned in the results and methods chapter of this thesis.

## Summary

Enhancers are *cis*-regulatory elements on genomic DNA that bind activator proteins and facilitate transcription of target genes. About ten years ago, active enhancers were found to be transcribed, giving rise to long non-coding RNAs, termed enhancer RNAs (eRNAs). Enhancer transcription correlates with enhancer activation, hence further elevating the transcription rates of target genes. This finding revolutionized our view of enhancers, which now are seen to exploit several mechanisms to promote target gene transcription. The eRNA transcripts themselves were thereby reported to employ various mechanisms to activate target gene transcription, *e.g.*, in neurons eRNAs were suggested to facilitate the release of the RNA polymerase II (Pol II) from the promoter-proximally paused state. To that end, eRNAs are believed to compete off negative elongation factor (NELF) from paused Pol II at target genes, thereby facilitating its transition into productive elongation. As the molecular mechanism of eRNA function was unknown to date, this study was aimed at investigating, whether eRNAs share common secondary structures that enable them to perform their activatory function. Moreover, this work was focused on developing a biochemical *in vitro* framework to systematically test the impact of eRNAs on paused Pol II, and to thereby gain insights into their underlying mechanism of action.

Within the scope of this thesis, I determined the precise 5'-ends of mouse neuronal eRNAs using the Exo-seq protocol and I performed chemical structure probing for a set of 39 eRNA 5'-end fragments (first 200 nucleotides), employing the SHAPE-MaP protocol. The data I obtained revealed that eRNAs adopt a broad spectrum of structures without sharing specific structural motifs. However, with the aid of biochemical *in vitro* assays, we were able to demonstrate that, indeed, eRNAs can trigger the dissociation of NELF from the paused Pol II complex and that they thereby facilitate transcription. Furthermore, we found that the efficiency of an eRNA to detach NELF from paused Pol II correlates with its length and with the presence of unpaired guanosines only. The overall structure of an eRNA seems to only play a minor role. Last, by using a combination of biochemical assays, RNA-protein crosslinking coupled to mass spectrometry, and experiments with NELF mutants, we rationalized the strict length-dependence of eRNA-driven NELF dissociation: eRNAs make allosteric contacts with multiple, distant regions on the NELF complex, namely the RRM

---

domain of NELF-E and positively charged patches on the NELF-AC lobe. These multivalent interactions then likely trigger the dissociation of NELF from paused Pol II.

Taken together, by revealing the molecular determinants for eRNA function, this study mechanistically links eRNAs to Pol II pause release and to activity-induced transcription in neurons. Moreover, this study sets the stage for a further detailed investigation of the interaction of eRNAs with the paused elongation complex by providing an established set of biochemical assays, that can be developed further to better represent transcription conditions *in vivo*.

# Zusammenfassung

Enhancer sind *cis*-regulatorische Elemente auf der genomischen DNA, die Aktivatorproteine binden und darüber die Transkription von Zielgenen fördern. Vor etwa zehn Jahren wurde zum ersten Mal entdeckt, dass aktive Enhancer auch transkribiert werden und dabei lange nicht-kodierende RNAs entstehen, die sogenannten enhancer RNAs (eRNAs). Die Transkription von Enhancern ist ein Zeichen ihrer Aktivierung und steigert zusätzlich die Transkription von Zielgenen. Dieser Fund revolutionierte unseren Blick auf Enhancer, die offenbar parallel mehrere Mechanismen zur Aktivierung von Zielgenen verwenden. Dabei nutzen die eRNA-Transkripte selbst ebenfalls verschiedene Mechanismen zur Zielgenaktivierung. Beispielsweise wurde für die eRNAs in Neuronen ein Mechanismus vorgeschlagen bei dem die eRNAs die Freisetzung der RNA Polymerase II (Pol II) aus der promoter-proximalen Pause forcieren. Dabei wird angenommen, dass die eRNAs den negativen Elongationsfaktor (NELF), der an die pausierte Pol II gebunden ist und die Pause stabilisiert, verdrängen und dadurch den Übergang der Pol II in die Elongationsphase fördern. Der genaue molekulare Mechanismus war jedoch bisher unbekannt. Deshalb fokussiert sich diese Arbeit einerseits darauf zu erforschen, ob die neuronalen eRNAs eine einheitliche Sekundärstruktur aufweisen, die verantwortlich für ihre Funktionalität sein könnte. Ein weiterer Aspekt dieser Arbeit war die Entwicklung von biochemischen *in-vitro* Experimenten, um die Wirkung von eRNAs auf die pausierte Pol II systematisch testen zu können.

Im Rahmen dieser Forschungsarbeit, wurden die 5'-Enden neuronaler eRNAs aus der Maus mithilfe des Exo-seq Protokolls ermittelt. Für ein Set von 39 eRNAs wurde anschließend die Sekundärstruktur ihrer 5'-Enden mittels der SHAPE-MaP Methode bestimmt. Die Resultate zeigten, dass die getesteten eRNA ein breites Spektrum an Strukturen aufweisen und scheinbar kein einheitliches Strukturmotiv besitzen. Jedoch konnte mithilfe von biochemischen *in-vitro* Experimenten eindeutig gezeigt werden, dass eRNAs in der Lage sind die Dissoziation des Faktors NELF von der pausierten Pol II zu verursachen und darüber die Transkription vom Zielgenen zu beschleunigen. Darüber hinaus konnte gezeigt werden, dass der NELF-dissozierende Effekt von eRNAs nur von ihrer Länge und vom Vorhandensein ungepaarter Guanosin-Reste abhängig ist. Die exakte Sekundärstruktur der eRNAs scheint bei der Stärke ihres Effektes eine untergeordnete Rolle zu spielen.

Abschließend konnte die strikte Längenabhängigkeit der eRNA-induzierten Dissoziation von NELF mittels biochemischer Assays, RNA-Protein Crosslinking-Massenspektrometrie, und Experimenten mit Mutanten von NELF, rationalisiert werden. Denn eRNAs interagieren gleichzeitig mit mehreren Regionen des NELF-Komplexes, nämlich der RRM-Domäne und positiv-geladenen Patches an der Oberfläche des NELF-AC Proteinlappen. Diese multivalenten Wechselwirkungen zwischen den eRNAs und dem NELF triggern letztendlich dessen Ablösung von der pausierten Pol II.

Zusammenfassend ist festzuhalten, dass diese Forschungsarbeit, durch die Aufdeckung der molekularen Determinanten für die Funktion von eRNAs, einen mechanistischen Zusammenhang zwischen eRNAs und der Auflösung der promoter-proximalen Pause von Pol II sowie der aktivitätsinduzierten Transkription in Neuronen, herstellt. Darüber hinaus bieten die *in-vitro* Assays, die in dieser Studie etabliert wurden, eine Ausgangsbasis für eine weitergehende Erforschung der Interaktion von eRNAs mit dem pausierten Pol II Elongationskomplex an. Dabei können die Experimente weiterentwickelt werden, um sie mehr an die zellulären Rahmenbedingungen anzupassen.

# Table of Contents

<b>Summary</b> .....	<b>I</b>
<b>Zusammenfassung</b> .....	<b>III</b>
<b>Table of Contents</b> .....	<b>V</b>
<b>List of abbreviations</b> .....	<b>XI</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Transcription initiation and promoter-proximal pausing.....	1
1.2 Promoter-proximal pausing and its functions.....	4
1.3 Structure and function of DSIF, NELF and the PEC.....	5
1.3.1 DRB sensitivity-inducing factor (DSIF).....	5
1.3.2 NELF and the PEC .....	7
1.3.3 Structure function relationship of the PEC.....	9
1.4 Transcriptional regulation by enhancer RNAs.....	10
1.4.1 Canonical view of enhancers .....	10
1.4.2 Specific (chromatin) features of enhancers .....	11
1.5 Enhancers as non-coding RNA transcription units.....	12
1.5.1 Features of eRNAs and their transcription .....	13
1.5.2 Role of enhancer transcription and mechanisms of eRNA function in gene regulation .....	14
<b>2 Aims and scope of the thesis</b> .....	<b>19</b>
<b>3 Results</b> .....	<b>20</b>
3.1 Determination of eRNA 5'-ends by Exo-seq.....	20
3.1.1 Establishment of the Exo-Seq protocol .....	20
3.1.2 Analysis of Exo-seq data.....	23
3.2 Mapping of eRNA secondary structure.....	27
3.2.1 Structure probing of eRNAs by <i>in vitro</i> SHAPE-MaP.....	27
3.2.2 Production of eRNAs for <i>in vitro</i> SHAPE-MaP.....	29
3.2.3 Analysis of the SHAPE-MaP data.....	30
3.2.4 <i>In vivo</i> SHAPE-MaP .....	34
3.3 Functional impact of eRNAs on the PEC <i>in vitro</i> .....	37
3.3.1 Electrophoretic mobility shift assays .....	38
3.3.2 eRNAs can detach NELF from the PEC.....	40
3.3.3 The eRNA secondary structure is not essential for triggering NELF release .....	42

3.3.4	Unpaired guanosines play a critical role in the dissociative effect of eRNAs on NELF.....	43
3.3.5	eRNAs bind to a positive patch on NELF-C and to the NELF-E RRM domain.....	47
3.4	Protein-RNA UV-crosslinking mass spectrometry.....	52
3.5	<i>In vitro</i> Transcription assays.....	59
<b>4</b>	<b>Discussion.....</b>	<b>65</b>
4.1	eRNAs function is not associated with a common structural motif.....	65
4.2	eRNAs can detach NELF from the paused elongation complex <i>in vitro</i> .....	66
4.3	NELF release is triggered through multivalent interactions with an eRNA molecule.....	68
4.4	Reported preferences of the RRM domain and NELF-AC lobe for guanosines.....	70
4.5	A revised model for eRNA-triggered NELF dissociation.....	71
4.6	Outlook.....	73
<b>5</b>	<b>Materials and Methods.....</b>	<b>76</b>
5.1	Chemicals, Materials and Equipment.....	76
5.2	Bacterial strains and insect cells.....	76
5.3	Transformation of plasmid DNA into competent cells.....	76
5.4	cDNA synthesis.....	77
5.5	Polymerase Chain Reaction.....	77
5.6	Colony PCR.....	77
5.7	Agarose gel electrophoresis and gel extraction.....	78
5.8	Cloning of DNA by Gibson Assembly.....	79
5.9	<i>In vitro</i> production of RNAs.....	80
5.9.1	Production of DNA-Templates.....	80
5.9.2	<i>In vitro</i> Transcription (IVT).....	81
5.10	Ethanol precipitation.....	82
5.11	Urea Polyacrylamide gel electrophoresis (PAGE):.....	82
5.12	Mouse cortical neuron culture.....	84
5.13	rRNA and 7SL RNA depletion from total RNA.....	84
5.14	Exo-seq library preparation.....	85
5.15	Processing of 5' Exo-seq reads.....	86
5.16	RNA structure probing by SHAPE-MaP.....	87
5.16.1	<i>In vitro</i> SHAPE-MaP.....	87
5.16.2	<i>In vivo</i> SHAPE-MaP.....	88
5.17	Analysis of SHAPE-MaP data.....	89
5.18	Purification of endogenous Pol II.....	90
5.19	Expression of NELF variants and DSIF in insect cells.....	91
5.20	Purification of NELF variants.....	92

---

5.21	Purification of DSIF.....	93
5.22	SDS-PAGE .....	93
5.23	Electrophoretic mobility shift assays (EMSAs) .....	94
5.24	Transcription assay on magnetic beads .....	96
5.25	Protein-RNA Crosslinking coupled to mass spectrometry .....	98
5.25.1	Protein-RNA crosslinking and sample preparation.....	98
5.25.2	Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis and identification of protein-RNA crosslinks.....	99
5.26	quantitative PCR (qPCR).....	99
5.27	Data availability.....	100
<b>6</b>	<b>References.....</b>	<b>101</b>
<b>7</b>	<b>Supplement .....</b>	<b>118</b>
7.1	Materials .....	118
7.1.1	Chemicals and Reagents .....	118
7.1.2	Critical commercial kits for nucleic acid applications.....	118
7.1.3	Enzymes.....	119
7.1.4	Antibodies .....	119
7.1.5	Nucleic acid and Protein Standards .....	119
7.1.6	Materials and Consumables.....	120
7.1.7	Equipment list .....	121
7.2	Sequences .....	122
7.2.1	Oligos.....	122
7.3	Software.....	131
7.4	Supplementary Data .....	132
7.4.1	RNA sequences.....	132
7.4.2	pUC18 plasmid with cloned eRNA Insert.....	136
7.4.3	Exo-seq and SHAPE-MaP supplementary data.....	138
7.4.4	Protein-RNA crosslinking data .....	145
	<b>Acknowledgements.....</b>	<b>158</b>
	<b>(Eidesstattliche) Versicherungen und Erklärungen .....</b>	<b>159</b>

# List of Figures

Figure 1-1   Steps of transcription initiation.....	2
Figure 1-2   Promoter-proximal pausing and pause release.....	3
Figure 1-3   Domain organization of DSIF.....	6
Figure 1-4   Domain architecture of NELF subunits.....	7
Figure 1-5   Cryo-EM structure of the paused elongation complex (PEC).....	8
Figure 1-6   Long-range communication between enhancer and promoter regions.....	10
Figure 1-7   Chromatin features of active enhancers.....	12
Figure 1-8   Enhancer transcription activates the transcription of target genes.....	13
Figure 1-9   Different mechanisms of eRNAs function. eRNAs facilitate transcription of their target genes by acting mostly in cis and deploying multiple mechanisms.....	16
Figure 3-1   Schematic representation of the Exo-seq workflow.....	20
Figure 3-2   Characteristics of Exo-Seq libraries before and after optimization.....	22
Figure 3-3   Exo-seq allows for determining enhancer TSSs with single-nucleotide precision..	24
Figure 3-4   Exo-seq signal at active enhancers overlaps with ChIP-seq enhancer marks .....	26
Figure 3-5   Overview of the main SHAPE-MaP steps and subsequent analysis tools .....	28
Figure 3-6   Preparation of eRNAs for SHAPE-MaP. ....	29
Figure 3-7   SHAPE-MaP results for Arc eRNA. ....	31
Figure 3-8   eRNAs display a broad spectrum of structures.....	33
Figure 3-9   In vivo SHAPE-MaP yields low mutation rates over background.....	36
Figure 3-10   Purified protein and eRNA components required for in vitro assays.....	37
Figure 3-11   Workflow of the radioactive electrophoretic mobility shift assays (EMSAs). ....	39
Figure 3-12   eRNAs are able to dissociate NELF from the paused elongation complex in a length-dependent manner .....	41
Figure 3-13   EMSAs with eRNAs mutants.....	43
Figure 3-14   Guanosines play a critical role in the dissociation of NELF.....	44
Figure 3-15   G-less mutants of Nr4a1-(a) show a reduced dissociative effect on NELF .....	46
Figure 3-16   Charged patches on NELF-AC and the NELF-E RRM domain are both involved in eRNA-driven NELF dissociation. ....	50
Figure 3-17   The NELF double mutant cannot be displaced from the PEC by eRNAs. ....	51
Figure 3-18   Gelfiltration of eRNA-bound PEC complexes used for Protein-RNA crosslinking1. ....	53

---

Figure 3-19   Representation of the RNA-protein crosslinks between nascent RNA and subunits of Pol II and DSIF.....	54
Figure 3-20   RNA crosslinks to NELF. ....	55
Figure 3-21   eRNA crosslinks to the NELF-E and NELF-A tentacle as well as to the interface Pol II-NELF-AC interface. ....	57
Figure 3-22   3D Structure model of Arc eRNA 1-200 next to the PEC. ....	58
Figure 3-23   Schematic, stepwise representation of the in vitro transcription assay. ....	59
Figure 3-24   Presence of DSIF and NELF slows down Pol II transcription.....	60
Figure 3-25   eRNAs fasten the pause release of Pol II in the transcription assay. ....	61
Figure 3-26   Transcription assay with repetitive G-containing and G-less RNA .....	62
Figure 3-27   Reduced pause release with NELF-C patch and double mutant .....	64
Figure 4-1   Mechanistic model for the eRNA induced Pol II pause release.....	72
Figure 7-1   View of the Arc eRNA 1-200 insert sequence cloned into a pUC18 plasmid .....	137
Figure 7-2   View of the Arc eRNA 1-200 insert sequence cloned into a pUC18 plasmid; not for SHAPE-MaP purposes .....	138
Figure 7-3   Additional SHAPE-MaP structures for eRNAs of the immediate early genes c-Fos and Junb. ....	144
Figure 7-4   Additional SHAPE-MaP structures for selected eRNAs from the set of 39 eRNAs.....	145

# List of Tables

Table 7-1   Chemicals and reagents .....	118
Table 7-2   Commercial kits and beads used for purification of DNA and RNA and for depletion of rRNA. ....	118
Table 7-3   Enzymes used in this study .....	119
Table 7-4   Antibodies used in this study .....	119
Table 7-5   DNA, RNA and protein standards used in this study .....	120
Table 7-6   Materials and Consumables .....	120
Table 7-7   Equipment used in this work .....	121
Table 7-8   PCR primer used to produce linear PCR templates for T7-mediated IVT .....	122
Table 7-9   Biotinylated antisense oligos and qPCR primer for 7SL RNA depletion and depletion analysis .....	123
Table 7-10   Oligos used for Exo-seq library construction .....	123
Table 7-11   Oligos for <i>in vitro</i> SHAPE-MaP library construction .....	123
Table 7-12   Primer used for <i>in vivo</i> SHAPE-MaP libraries of <i>Gadd45b</i> , <i>Nr4a1-(a)</i> , <i>Fosb</i> and <i>Fos-e2</i> eRNA.....	124
Table 7-13   Oligos used for transcription bubble formation in EMSAs and Protein-RNA UV-crosslinking.....	125
Table 7-14   Oligos used for transcription bubble formation in <i>in vitro</i> transcription assays .....	125
Table 7-15   Repetitive, low complexity RNAs used for EMSAs .....	125
Table 7-16   DNA template oligos for T7-mediated <i>in vitro</i> synthesis of repetitive RNAs.....	126
Table 7-17   DNA template oligos for <i>in vitro</i> synthesis of <i>Nr4a1-(a)</i> G-to-A and G-to-C mutants. ....	127
Table 7-18   Vector- and insert-specific primers used to clone eRNAs (not for SHAPE-MaP).....	127
Table 7-19   Vector- and insert-specific primers used to clone eRNAs for SHAPE-MaP experiment..	128
Table 7-20   Software used in this study.....	131
Table 7-21   Sequences of the cloned set of 39 eRNA (1-200 nt) .....	132
Table 7-22   Sequences of the mutant eRNAs .....	135
Table 7-23   Final list of 39 eRNAs extracted from the Exo-seq data and the overlap with GRO-seq data.....	138
Table 7-24   <i>In vitro</i> SHAPE-MaP sequencing statistics .....	140
Table 7-25   <i>In vivo</i> SHAPE-MaP sequencing statistics .....	142
Table 7-26   List of Protein-RNA crosslinks for the experiment with <i>Arc</i> eRNA 1-200..	145
Table 7-27   List of Protein-RNA crosslinks for the experiment with <i>Nr4a1-(b)</i> eRNA 1-100.....	149

# List of Abbreviations

1M7	1-methyl-7-nitroisatoic anhydride
APS	Ammonium Persulfate
bp	base pair(s)
C	cytosine
CBP	CREB binding protein
CTCF	CCCTC-binding factor
ChIP	Chromatin immunoprecipitation
CREB	cAMP response element-binding protein
CTD	C-terminal domain
CV	column volume(s)
<i>D. melanogaster</i>	Drosophila melanogaster
DEPC	Diethyl pyrocarbonate
DEPC-H <sub>2</sub> O	DEPC treated water
DNA	Desoxyribonucleic acid
DSIF	DRB sensitivity-inducing factor
DMSO	Dimethyl sulfoxide
<i>E. coli</i>	Escherichia coli
EMSA	Electrophoretic mobility shift assay
eRNA	enhancer RNA
ID	identification (number/name)
IVT	in vitro transcription
mRNA	messenger RNA
MWCO	molecular weight cutoff
NELF	Negative elongation factor
lncRNA	long non-coding RNA
nt	nucleotide(s)
NT-DNA	Non-template DNA
G	guanosine
G4	G-quadruplex
GTF	General transcription factor
HAT	Histone acetyltransferase
HMT	Histone methyltransferase
PAGE	Polyacrylamide gel electrophoresis
PEC	Paused elongation complex
Pol II	RNA polymerase II
P-TEFb	Positive transcription elongation factor b

---

poly(A)	polyadenylated
RNA	Ribonucleic acids
rpm	revolutions per minute
SHAPE-MaP	Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP)
T-DNA	Template DNA
TEMED	N,N,N',N'-Tetramethylethylenediamin
TF	Transcription factor
TSS	Transcription start site

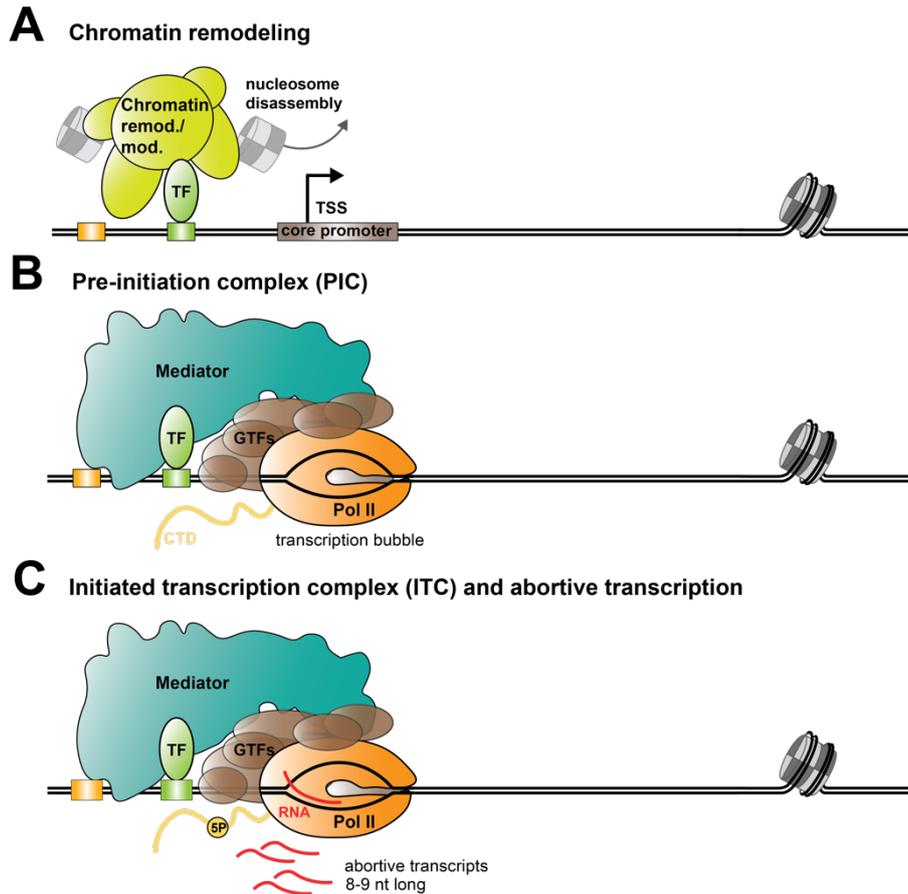
# 1 Introduction

## 1.1 Transcription initiation and promoter-proximal pausing

The accurate spatiotemporal control of gene expression is crucial for the viability of all organisms. This is in particular the case for eukaryotic cells, where gene expression is a highly intricate process that encompasses a network of regulatory mechanisms that act at different stages of the gene expression process. The tight regulation of gene expression forms the basis of cell differentiation during the development of multi-cellular organisms and in the response and adaption of cells to environmental stimuli. Within the gene expression framework transcription itself is the central process. It takes place in the nucleus and is responsible for copying the genetically encoded information from the DNA into protein-coding messenger RNA (mRNA). In eukaryotic cells, RNA polymerase II (Pol II) is the central enzyme catalyzing mRNA synthesis.

Eukaryotic transcription can be divided into three main steps: initiation, elongation, and termination. The initiation step of transcription (Figure 1-1) is highly regulated by a plethora of protein factors, beyond others transcription factors (TFs), co-activators as well as by regulatory DNA elements, like promoters and enhancers (see section 1.4). The establishment of open chromatin at the regulatory elements is a prerequisite for the initiation step (Li et al., 2007). It is accomplished by chromatin remodelers, modifiers and co-activators, like histone methyltransferases (HMTs) and histone acetyl transferases (HATs).

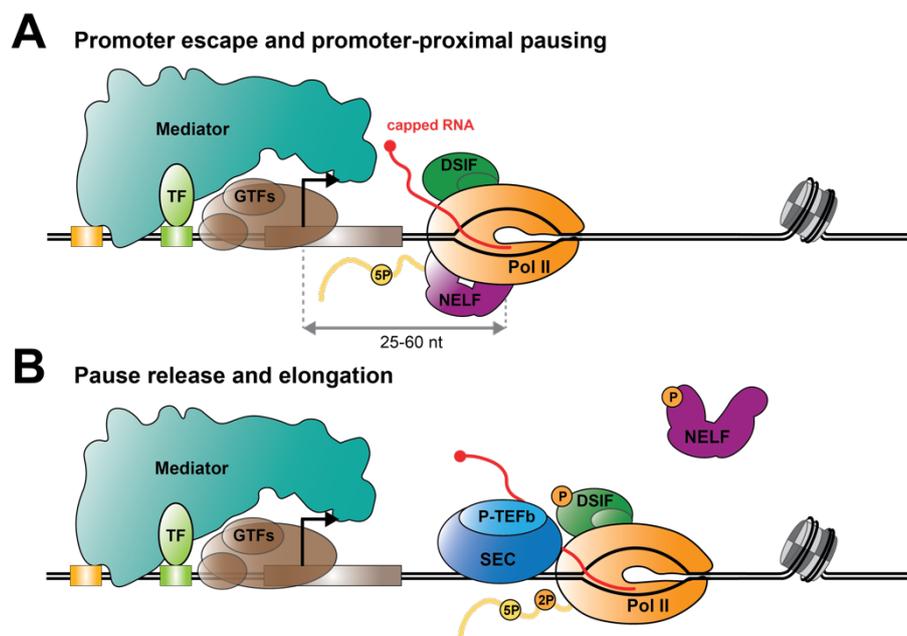
During transcription initiation, Pol II is recruited to the gene promoter with the help of general transcription factors (GTFs) and the huge co-activator complex Mediator, together forming the pre-initiation complex (PIC) (Murakami et al., 2013; Schier & Taatjes, 2020). The GTF that binds last to the PIC, TFIID, possesses helicase activity to drive promoter DNA melting. The template strand of the so formed transcription bubble then engages then with the active site of Pol II (Dienemann et al., 2019; Schilbach et al., 2017). Subsequently, the PIC scans for the transcription start site (TSS), where the synthesis of the nascent RNA starts *de novo*, leading to the formation of the initially transcribing complex (ITC) (Cheung et al., 2011; Sainsbury et al., 2013). After a few cycles of abortive transcription (Goldman et al., 2009; Kapanidis et al., 2006), Pol II escapes the promoter.



**Figure 1-1 | Steps of transcription initiation.** (A) Transcription initiation is preceded by the establishment of open chromatin at the gene's promoter region aided by chromatin remodeling and modifying complexes, and by binding of transcription factors (TFs) to *cis*-regulatory elements (like enhancer, green and orange box). (B) The formation of the pre-initiation complex (PIC) at the core promoter, comprising Pol II in its unphosphorylated form (no phosphorylation of the C-terminal domain (CTD)), general transcription factors (GTFs) and the huge Mediator complex. (C) Within the initiated transcription complex (ITC) the CTD of Pol II becomes phosphorylated at the serine-5 residues (labeled 5P) and Pol II performs multiple rounds of abortive transcription before Pol II escapes the promoter (see Figure 1-2).

Before the transition of Pol II into productive elongation, in which it transcribes at high speed through the gene body, Pol II mostly pauses at a region 20-60 nucleotides (nt) downstream of the TSS. Pausing at this so-called early elongation step of Pol II is referred to as promoter-proximal pausing (Core & Adelman, 2019; Kwak & Lis, 2013). In contrast to other pausing events in the gene body, the promoter-proximal pause is very stable, with an average half-life of paused Pol II spanning from ~2 to 30 min. Initially, promoter-proximal pausing was considered to be a phenomenon restricted to a subset of genes. However, today's evidence suggests that promoter-proximal pausing is a mandatory, and rate-limiting step in the transcription process experienced by Pol II at most, if not all, genes (Core & Adelman, 2019). Promoter-proximal pausing goes hand in hand with the formation of a dedicated Pol II

complex, the paused elongation complex (PEC) that consists of Pol II, the DRB sensitivity-inducing factor (DSIF) (Wada et al., 1998), and the negative elongation factor (NELF) (Yamaguchi et al., 1999, 2002) (Figure 1-2). Pol II is released into productive elongation by the activity of the positive transcription elongation factor b (P-TEFb) (Marshall & Price, 1992, 1995). P-TEFb is a cyclin-dependent kinase (CDK) that consist of CDK9 and cyclin T1 and is often found to be part of the super elongation complex (SEC) (Luo et al., 2012). P-TEFb phosphorylates serine-2 residues of Pol II C-terminal domain (CTD) as well as DSIF and NELF, leading to the dissociation of NELF and turning DSIF into a positive elongation factor. In contrast to NELF, DSIF remains bound to Pol II throughout the elongation phase and increases Pol II processivity.



**Figure 1-2 | Promoter-proximal pausing and pause release.** (A) After several rounds of abortive transcription (see Figure 1-1C), Pol II escapes the promoter and pauses about 25-60 nt downstream of the transcription start site (TSS; depicted by an arrow). The formed paused elongation complex (PEC) comprises Pol II and the pausing factors DSIF and NELF. (B) Pol II is released from the pause into productive elongation by the action of the kinase P-TEFb, which is often part of the super elongation complex (SEC). P-TEFb phosphorylates Pol II CTD at serine-2 residues (labeled 2P), phosphorylates NELF and DSIF. While NELF dissociates, DSIF turns into a positive elongation factor and remains bound to Pol II.

Each step of transcription is linked to a particular phosphorylation pattern of the CTD, which represents the disordered, C-terminal extension of the largest subunit of Pol II, RPB1. The CTD comprises numerous repeats of the consensus sequence  $Y_1S_2P_3T_4S_5P_6S_7$ , varying in number from 26 or 27 repeats in budding yeast to 52 repeats in mammalian Pol II (P. Liu et al., 2010).

During Pol II recruitment and the formation of the PIC, the CTD is unphosphorylated, allowing interactions with the Mediator complex. During early elongation, the CTD is becoming phosphorylated by the CDK7 kinase, which is part of the general transcription factor TFIIF, at the serine-5 (Ser5) residues of the repeats. The transition into productive elongation is facilitated by the phosphorylation of Ser2 residues by the kinase P-TEFb, as described above. These post-transcriptional CTD modifications stimulate interactions of Pol II with stage-appropriate TFs and RNA processing factors (Buratowski, 2009; Harlen & Churchman, 2017).

## 1.2 Promoter-proximal pausing and its functions

Paused RNA Pol II was initially described as a phenomenon in the promoter-proximal region of the *hsp70* heat shock gene in *Drosophila melanogaster* S2 cells (Gilmour & Lis, 1986). There, Pol II was observed to be enriched at the -12 to +65 promoter region in uninduced cells already before the application of a heat shock. Later, high throughput methods, such as chromatin immunoprecipitation (ChIP) combined with deep sequencing (ChIP-seq) and Global run-on sequencing (GRO-seq), provided evidence for the accumulation of transcriptionally engaged Pol II close to the vast majority of promoters in metazoan cells (Core et al., 2008; Guenther et al., 2007; Kim et al., 2005; Muse et al., 2007; Zeitlinger et al., 2007). In subsequent work, promoter-proximal pausing was established to be a regulatory fine-tuning step rather than an "on-off switch" for gene expression. It was found at highly expressed genes as well as at those showing very low transcription levels, but not at constitutively inactive genes (Core & Adelman, 2019).

Pol II pausing allows for the integration of regulatory signals opposing or reinforcing the transcription elongation. The exact combination of transcription factors that either favor the engagement of Pol II at the promoter or that favor pause release fine-tunes gene expression in a dynamic way (Adelman & Lis, 2012; Core & Adelman, 2019). Thereby, Pol II pausing offers the basis for a rapid and/or synchronous activation of transcription in signaling systems, such as the immune response (Adelman et al., 2009; Gilchrist et al., 2012), hormone signaling (Gupte et al., 2013; Hah et al., 2011) and during early development (Boettiger & Levine, 2009; Lagha et al., 2013; Saunders et al., 2013). The depletion of NELF in mouse embryonic stem cells or the loss of SPT5 in zebrafish cause defects in differentiation, stressing the critical role of pausing during development (Amleh et al., 2009; Guo et al., 2000). Another

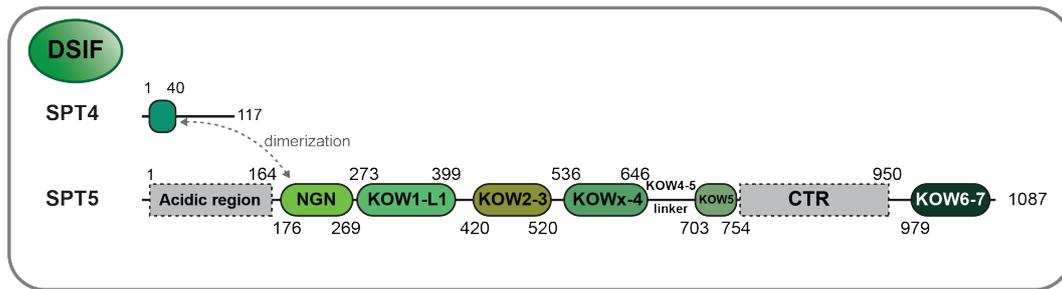
function of promoter-proximally paused Pol II is the maintenance of nucleosome-deprived promoters, as shown genome-wide in *Drosophila*, by building a physical barrier to the nucleosome assembly (Gilchrist et al., 2008, 2010). Interestingly, the DNA sequences at the promoters of paused genes are predicted to favor nucleosome assembly and probably require paused Pol II to maintain opened chromatin at the promoter. Keeping chromatin accessible for TFs and co-activators is a prerequisite for the rapid activation of genes, as open chromatin at promoters primes these for bursts of transcription activation in response to external cues (Adelman & Lis, 2012).

Furthermore, promoter-proximal pausing could provide a checkpoint for coupling elongation and RNA processing. This notion is conceivable, as the pause provides a time-window for proper 5' capping of nascent RNA before productive elongation occurs. Indications for a link between the pausing and RNA capping are that DSIF was shown to interact with and stimulate the capping enzyme (Mandal et al., 2004; Moore & Proudfoot, 2009) and that nascent transcripts associated with paused Pol II are already 5' capped (Nechaev et al., 2010; Rasmussen & Lis, 1993). Besides this, the release of paused Pol II is coupled to the phosphorylation of Ser2 residues on the Pol II CTD by the kinase P-TEFb. This modification provides a binding platform for complexes that perform RNA 3' end processing (Buratowski, 2009). Promoter-proximal pausing thus could prevent Pol II from proceeding into the gene body before it is appropriately modified to bind RNA processing factors (Adelman & Lis, 2012).

### 1.3 Structure and function of DSIF, NELF and the PEC

#### 1.3.1 DRB sensitivity-inducing factor (DSIF)

Human DSIF is a heterodimer that consists of SPT5 (hSpt4; 120 kDa) and SPT4 (hSpt4; 14 kDa), both of which are conserved in *Saccharomyces cerevisiae* (Spt4/Spt5) (Hartzog et al., 1998; Wada et al., 1998). The large subunit SPT5 comprises a disordered N-terminal acidic region, followed by an NGN (NusG N-terminal domain) (Ponting, 2002), multiple KOW (Kyrpides, Ouzounis, Woese) motifs (Kyrpides et al., 1996), and a mobile C-terminal-repeat region (CTR). The CTR is phosphorylated by P-TEFb during pause release, an event that converts DSIF into a positive transcription elongation factor (Peterlin & Price, 2006) (Figure 1-3).

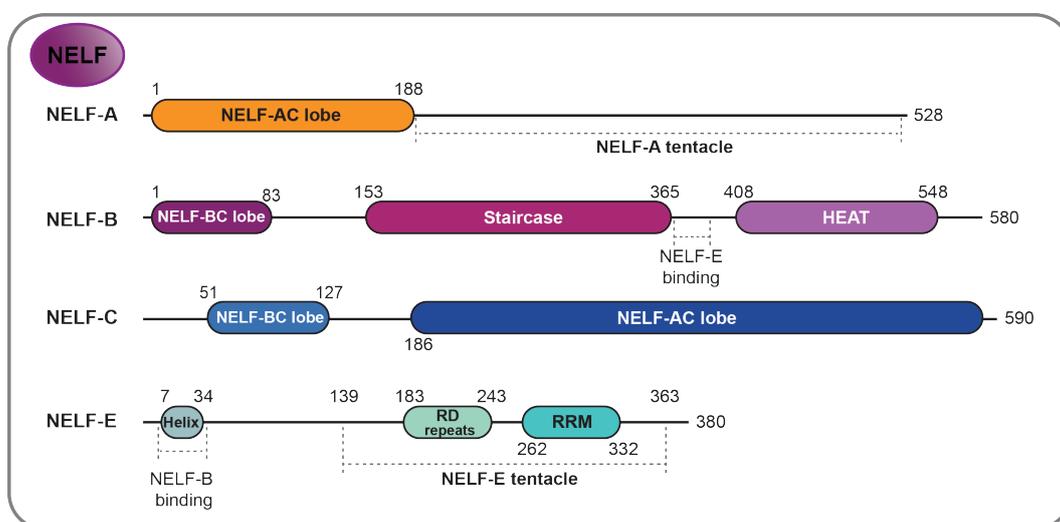


**Figure 1-3 | Domain organization of DSIF.** Schematic representation of human SPT4/SPT5 (DSIF) domain organization. Numbers along the scheme depict the first and last amino acid residues of each domain. Adapted from Bernecky et al., 2017.

SPT5 is the only transcription factor that is universally conserved in all three domains of life. This is, however, only true for its role as an elongation factor, not for its part in pausing (Werner, 2012). The bacterial homolog of DSIF, NusG, exists as a monomer without SPT4 and comprises only the NGN and one KOW domain. SPT5 dimerizes with SPT4 via the NGN domain (Figure 1-3) (Guo et al., 2008). The structure of the human Pol II-DSIF complex was recently solved (Bernecky et al., 2017). In this complex the NGN–SPT4 and the KOW1–L1 domains are forming a "DNA-clamp" that bridges the Pol II active center cleft and that positions upstream DNA (Figure 1-5C). The NGN-SPT4 dimer mainly contributes to the maintenance of a closed active center cleft. This is important for the function of DSIF in transcription elongation, as the "DSIF clamp" increases the processivity of Pol II by locking it in a closed, pause-resistant state (Martinez-Rucobo et al., 2011). Furthermore, DSIF also forms an "RNA clamp" at the RNA exit tunnel of Pol II by docking its KOWx-4 and KOW5 domains to Pol II (Figure 1-5C). In more detail, both domains directly contact nascent RNA at register -15 to -18 relative to the Pol II active site (register +1) (Bernecky et al., 2017). This finding explained previous reports that demonstrated that stable binding of DSIF to Pol II requires interactions between DSIF and the emerging nascent RNA transcript. This requirement causes DSIF to associate with Pol II only when the RNA transcript reaches a length of > 18-25 nt (Cheng & Price, 2008; Missra & Gilmour, 2010; Palangat et al., 2005). Moreover, DSIF was reported to recruit the capping enzyme (CE) and stimulate capping of nascent RNA in humans and yeast (Lindstrom et al., 2003; Pei & Shuman, 2002; Wen & Shatkin, 1999). Indeed, nascent RNA is co-transcriptionally capped (Moteki & Price, 2002) as it reaches a length of ~19-24 nt (Mandal et al., 2004; Rasmussen & Lis, 1993, 1995; Tome et al., 2018).

### 1.3.2 NELF and the PEC

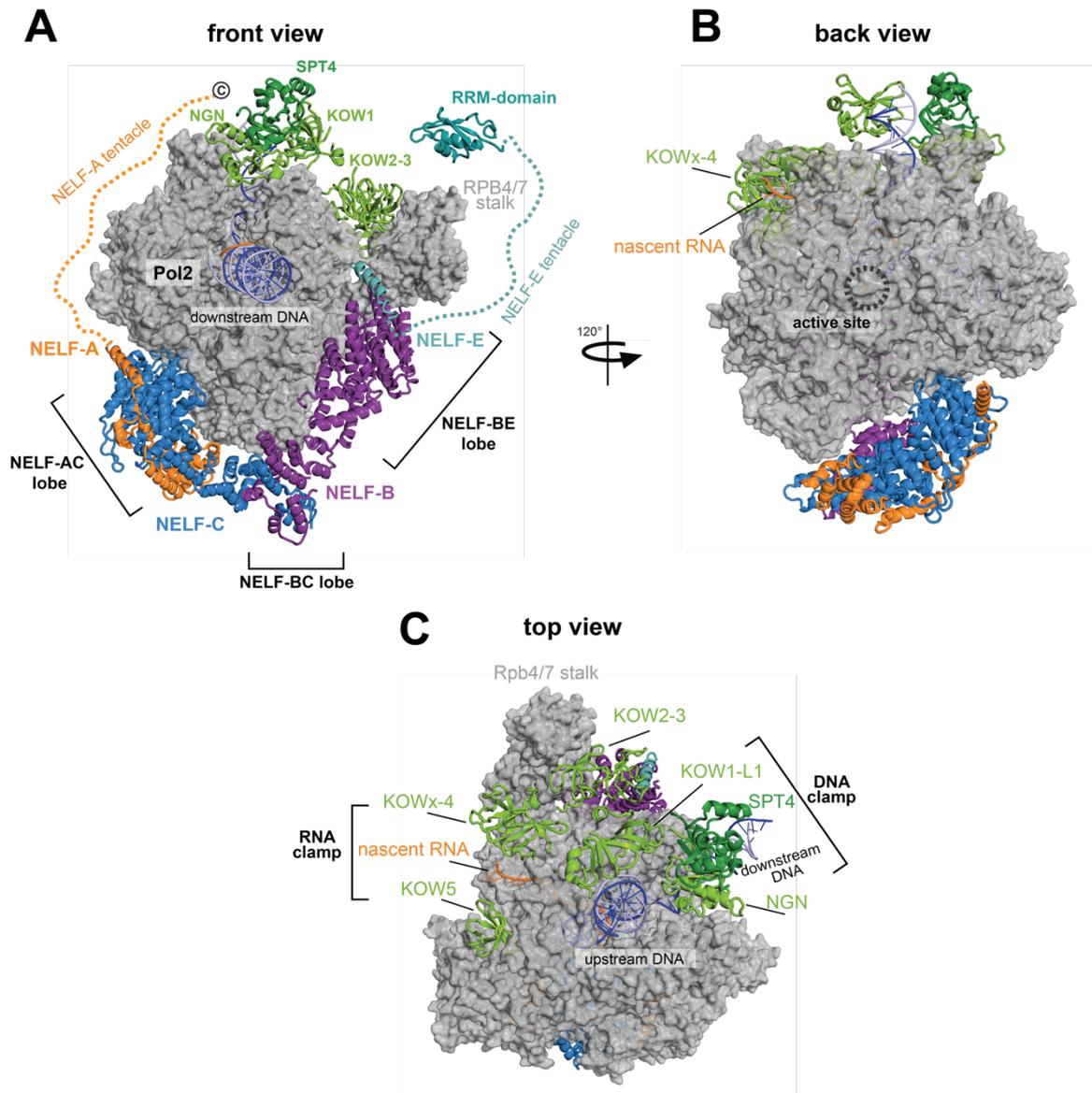
NELF is a complex of four subunits: NELF-A (57 kDa), -B (66 kDa), -C/D (66 kDa), and E (43 kDa) (Figure 1-4). The tetrameric complex contains either NELF-C or -D, which are isoforms of the same protein generated by alternative translation start sites. NELF-D is the shorter isoform lacking the first nine residues of NELF-C (aa 1-9) (Narita et al., 2003). The overall structure of the NELF complex is intrinsically flexible due to many unstructured regions. Interestingly, there are no homologs of NELF in yeast, *C. elegans* and plants, organisms that also lack *bona fide* promoter-proximal pausing (Gaertner & Zeitlinger, 2014).



**Figure 1-4 | Domain architecture of NELF subunits.** Schematic representation of human NELF domain organization. Numbers along the scheme depict the first and last amino acid residues of each domain. Adapted from Vos et al., 2018.

The Pol II-DSIF-NELF complex structure was recently solved by cryo-electron microscopy (cryo-EM) to a resolution of 3.2 Å. The complex structure revealed for the first time the exact binding mode of NELF on Pol II (Vos et al., 2018). NELF adopts a three-lobed structure composed of the NELF-AC, NELF-BC, and NELF-BE lobes, largely confirming previous characterizations of the NELF subunit arrangement (Narita et al., 2003; Vos et al., 2016). NELF binds to the bottom of Pol II beneath the entry site of the downstream DNA, on the face opposite of the cleft (Figure 1-5A), the binding interface of NELF-Pol II is highly charged. The NELF-AC and -BE lobes do not contact each other and bind independently to two mobile regions on Pol II, whereas the NELF-BC lobe does not bind to Pol II (Figure 1-5). The NELF-AC lobe consists of the previously crystalized NELF-AC dimer (NELF-A: 6-188 and NELF-C: 183-590) (Vos et al., 2016) and binds to the mobile core and shelf module of Pol II (Cramer et al.,

2001). The NELF-BE lobe comprises a 'staircase' and a HEAT repeat domain (Figure 1-4) that anchor the N-terminal helix  $\alpha$ -1 of NELF-E in between them. The NELF-B staircase domain primarily contacts Pol II subunits RPB5 and RPB6, both of which are part of the mobile shelf module (Cramer et al., 2001).



**Figure 1-5 | Cryo-EM structure of the paused elongation complex (PEC).** Shown is the Cryo-EM structure of the paused elongation complex (PDB code 6GML; Vos et al., 2018) from three different perspectives: front (A), back (B) and top view (C). Pol II structure is shown as surface (gray), NELF and DSIF are shown as ribbon models. Individual subunits of NELF and DSIF, the nascent RNA and DNA are labeled. The front view (A) additionally shows the NMR-solved structure of the NELF-E RRM-domain (PDB code 2JX2; Rao et al., 2008) and schematically illustrates the NELF-A and NELF-E tentacle reaching to top of the PEC.

In addition to the well-structured lobes, NELF features two flexible regions that were not visible in the cryo-EM structure but were tracked on the Pol II surface by lysine-lysine

crosslinking coupled to mass spectrometry (XL-MS). The region extending from the NELF-AC lobe is referred to as the NELF-A tentacle (residues 189-528), and from the NELF-BE lobe as NELF-E tentacle (residues 139 – 363) (Figures 1-4 and 1-5) (Vos et al., 2018). While the structured domains of NELF are not contacting DSIF, both tentacles seem to reach DSIF. This might explain why NELF binding to Pol II requires prior DSIF binding (Misra & Gilmour, 2010; Yamaguchi et al., 1999).

NELF-E is the smallest NELF subunit and bears a C-terminal RNA recognition motif (RRM) within the flexible NELF-E tentacle, whose structure was previously solved by NMR (Rao et al., 2006, 2008). The human RRM domain was shown to bind different RNA sequences and structures, and binding of RRM to the nascent RNA transcript was proposed to play a role in stabilizing the promoter-proximal pause (Yamaguchi et al., 2002). Interestingly, the homologous NELF-E/RRM protein in *Drosophila* was reported to have a high binding specificity to a certain sequence-structure motif, referred to as the NELF binding element (NBE) (Pagano et al., 2014). However, the binding specificity of the RRM domain and its relevance for pausing is still under debate, as contradictory results are speaking for and against the involvement of the RRM domain in promoter-proximal pausing (Schaukowitch et al., 2014; Vos et al., 2018; Yamaguchi et al., 2002). Notably, besides the RRM domain, NELF features positively charged patches on its surface, which were shown to bind nucleic acids. However, the role of these patches in pausing is unclear (Vos et al., 2016).

### 1.3.3 Structure function relationship of the PEC

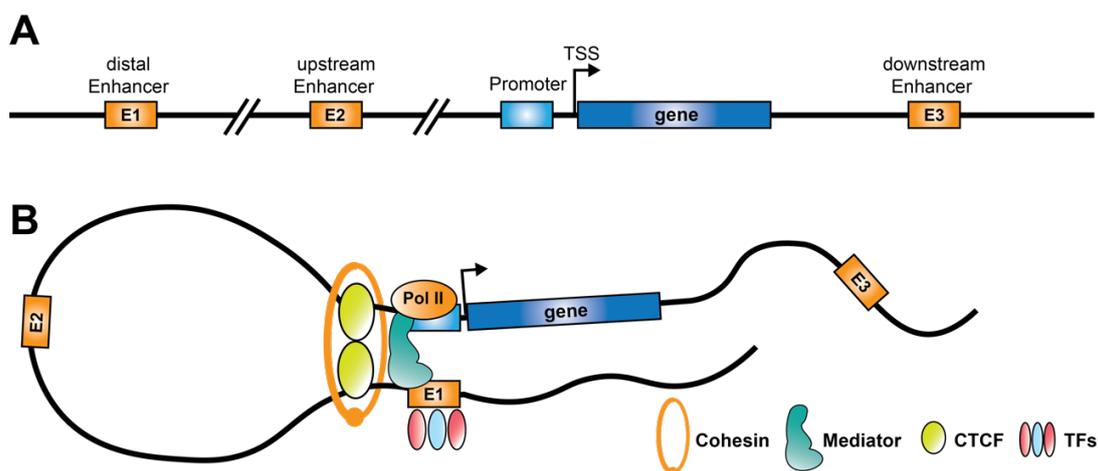
Multiple aspects of the PEC structure explain how this complex's formation may induce promoter-proximal pausing at the molecular level. First, the binding of NELF seems to induce a tilted conformation of the DNA-RNA hybrid in the active site of Pol II, compromising the binding of a nucleoside triphosphate (NTP) (Vos et al., 2018). Such a tilted state conformation was previously observed for backtracked and short DNA-RNA hybrids (Cheung et al., 2011; Cheung & Cramer, 2011) and even for paused bacterial elongation complexes (ECs) (Guo et al., 2018; Kang et al., 2018). This finding suggests that the induction of a tilted DNA-RNA hybrid could generally underlie paused states of multi-subunit RNA polymerases. Second, NELF bridges two mobile polymerase modules, the core and the shelf module (Cramer et al., 2001), thereby restraining Pol II's movement in general. A third aspect contributing to NELF-

mediated pause stabilization is that NELF appears to occupy the binding region of the anti-pausing factor TFIIIS on Pol II. TFIIIS promotes the reactivation of backtracked Pol II by stimulating cleavage of the 3' end of the nascent RNA (Reines et al., 1992). Promoter-proximal pausing is usually accompanied by backtracking, and efficient pause release relies not solely on P-TEFb but also on the action of TFIIIS (Adelman et al., 2005; Nechaev et al., 2010).

## 1.4 Transcriptional regulation by enhancer RNAs

### 1.4.1 Canonical view of enhancers

Eukaryotic transcription is a highly complex and elaborate process that requires the orchestration of interactions between proteins like Pol II and transcription factors and regulatory regions on the DNA. Enhancers are cis-regulatory DNA sequences positively driving the transcription of target genes. They are key determinants of cell type-specific gene expression patterns (Bulger & Groudine, 2011; Levine, 2010). Enhancers were initially characterized about 40 years ago as 200-500 bp long DNA sequences that contain clustered binding sites for multiple transcription factors. Enhancers activate transcription of target genes independent of their distance and orientation with respect to the promoter (Banerji et al., 1981).



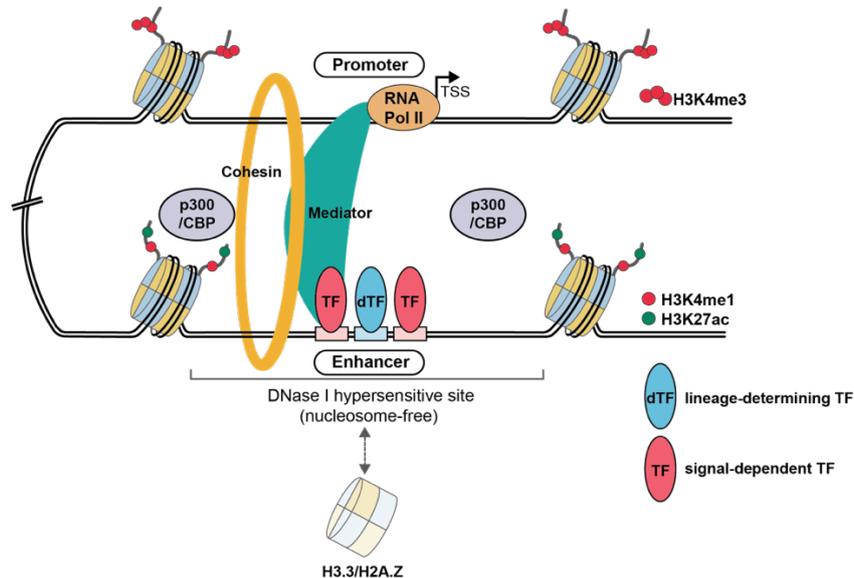
**Figure 1-6 | Long-range communication between enhancer and promoter regions.** (A) Schematic representation of a promoter, gene, and multiple scattered enhancer elements (boxed) along a linear view of chromosomal DNA (black line). (B) Chromatin loop promoted by factors like CTCF and Cohesin brings the distal enhancer (E1) in close proximity to the promoter. The Enhancer and promoter are further bridged by the Mediator complex, a transcription co-activator that activates transcription of target genes by Pol II.

Enhancers and target promoters can be located more than 1 Mb apart. To exert their activating function, they are brought into close proximity by looping out the interjacent DNA (Dekker, 2008) (Figure 1-6). The formation of these so-called chromatin loops (or enhancer-promoter loopings) is facilitated and stabilized mainly by two factors that are in general responsible for the organization of global chromatin architecture: the CCCTC-binding factor (CTCF) (Phillips & Corces, 2009) and Cohesin (Mishiro et al., 2009; Wendt et al., 2008), a large ring complex known to mediate sister chromatid cohesion (Michaelis et al., 1997). Besides these factors, the large Mediator complex (see section 1.1; Figure 1-1) was shown to play an important role for the enhancer-promoter looping and to interact with Cohesin (Kagey et al., 2010) (Figure 1-6).

#### 1.4.2 Specific (chromatin) features of enhancers

Enhancers are characterized by distinct chromatin features (Figure 1-7). Using high-throughput methods, these features are widely employed to identify putative enhancers (Heintzman et al., 2007). Enhancers are largely nucleosome-free regions displaying DNase I hypersensitivity (Gross & Garrard, 1988), yet they can be bound by dynamic, hyper-mobile, nucleosomes. Such dynamic nucleosomes contain the unstable histone variants H3.3 and H2A.Z and exhibit high histone replacement rates (Barski et al., 2007; Jin & Felsenfeld, 2007; Mito et al., 2007). In contrast to promoter regions, which are flanked by nucleosomes with trimethylated lysine 4 residues at histone H3 (H3K4me3), enhancer regions are marked by monomethylated lysine 4 residues (H3K4me1) and acetylated lysine 27 (H3K27ac) (Heintzman et al., 2007; Zentner et al., 2011). The H3K27ac modification is deposited by the co-activator p300 or the cyclic AMP-responsive element binding (CREB) protein (CBP), which are closely related proteins and whose presence is another prominent signature of enhancer regions (Birney et al., 2007; Heintzman et al., 2007; Visel et al., 2009). During early developmental stages, enhancer regions get primed for their subsequent activation by binding of so-called pioneer factors, the presence of H3K4me1, and hypomethylated DNA (Calo & Wysocka, 2013). During differentiation, lineage-determining TFs and signal-dependent TFs bind to the primed enhancers and recruit co-activators such as p300/CBP, which places the H3K27ac mark and finally activates the enhancer (Calo & Wysocka, 2013; Heinz et al., 2015). Also, activated enhancers were reported to exhibit an increased proportion of the promoter-

specific H3K4me3 modification, while the H3K4me1 mark is less prominent (Henriques et al., 2018).

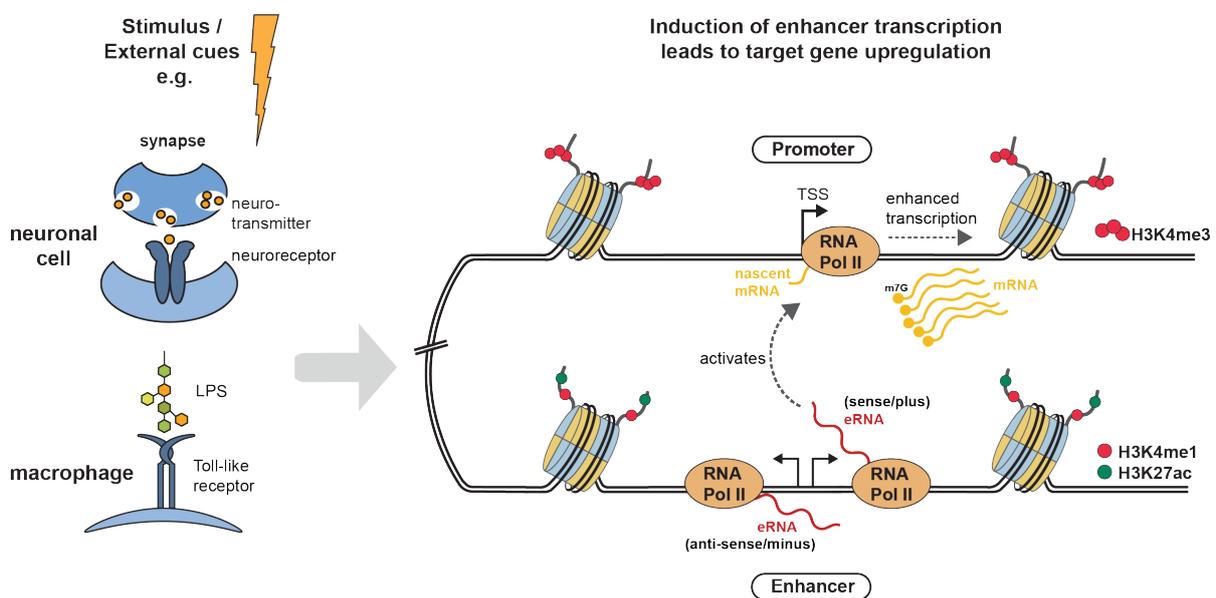


**Figure 1-7 | Chromatin features of active enhancers.** Enhancers are DNase I hypersensitive regions, free of nucleosomes (or bound by highly dynamic nucleosomes with H3.3/H2A.Z histone variants). They comprise binding sites for multiple TFs, which recruit further co-activators, like the Mediator complex. Histones flanking the nucleosome-free region of enhancers are modified with the H3K4me1, and the H3K27ac mark that characteristic for an active enhancer. The H3K4me3 modification marks histones at promoter regions. The co-activator p300/CBP, which writes the activating H3K27ac mark, is another prominent enhancer feature. The Cohesin ring complex is bridging the enhancer and the target promoter and stabilizes the enhancer-promoter looping.

## 1.5 Enhancers as non-coding RNA transcription units

About 30 years after discovering enhancer sequences, in the upcoming era of genome-wide RNA sequencing, enhancer regions were found to be transcribed by Pol II, producing eRNAs, as a novel class of long non-coding RNAs. The initial two studies reported Pol II peaks and low-abundance transcripts at enhancer regions upon neuronal stimulation by potassium chloride (KCl) and macrophage activation by lipopolysaccharides (LPS), respectively (De Santa et al., 2010; Kim et al., 2010). Since then, activity-induced enhancer transcription has been reported for various cell lineages and in response to different stimuli (Andersson et al., 2014a). In general, enhancer transcription positively correlates with the transcription of the cognate target gene (Figure 1-8). Over time, enhancer transcription became even a more reliable marker for functional enhancers and is nowadays more frequently used to annotate enhancers than solely epigenetic marks.

For the *Arc* enhancer/promoter pair in mouse cortical neurons, it was shown that eRNA transcription is dependent on the presence of the cognate promoter region. In contrast, the sole Pol II binding to the enhancer seems to be promoter-independent (Kim et al., 2010). However, there are also opposing results on the dependence of eRNA transcription on the enhancer-promoter interactions (Ling et al., 2005). Multiple single-locus and global studies on transcriptional kinetics indicate that enhancers are the first units to be transcribed upon a stimulus and their activation likely precedes the transcription of their target genes (Arner et al., 2015; Hsieh et al., 2014; Kim et al., 2015; Schaukowitch et al., 2014).



**Figure 1-8 | Enhancer transcription activates the transcription of target genes.** Transcription of enhancers is often induced upon a stimulus such as the depolarization of neurons through neurotransmitter signaling or LPS-induced activation of macrophages. The transcription of enhancers is carried out by RNA Pol II and is a sign of a fully activated enhancer. The transcription at enhancers is typically bidirectional, producing a sense and anti-sense eRNA. Further, the induction of enhancer transcription positively correlates with the induction of target gene transcription. Enhancer transcription precedes the gene upregulation and activates the gene transcription through different mechanisms (see section 1.5.2).

### 1.5.1 Features of eRNAs and their transcription

Several studies support the idea that transcription initiation at promoters and enhancers work similarly, while the elongation and termination steps differ. Referring to this, it was shown that GTFs and the serine 5-phosphorylated form of Pol II (Ser5p), which are associated with initiation and early elongation at gene promoters, are also recruited to enhancers (Koch et al., 2011). However, the serine 2-phosphorylated form of Pol II (Ser2p) and the chromatin

mark H3K36me3, associated with productive elongation, are less abundant at enhancers (Koch et al., 2011). Transcription termination at enhancers is dependent on the Integrator complex (Lai et al., 2015), known initially from the termination of small nuclear RNAs (Baillat et al., 2005), and the function of the WD repeat-containing protein 82 (WDR82) (Austena et al., 2015).

Besides numerous similarities between the transcription of genes and enhancers, eRNAs also exhibit distinct features. eRNAs are mostly present in the nuclear and chromatin-bound fractions. eRNAs possess a 5' 7-methylguanosine (m<sup>7</sup>G) cap just like mRNAs and lncRNAs (Core et al., 2014; Lam et al., 2013). RNA capping generally happens shortly after initiation, and capping factors are recruited and stimulated by the Ser5p form of Pol II, which is present at promoters and enhancers (Bentley, 2014; Ho & Shuman, 1999). However, unlike mRNAs and lncRNAs, enhancer RNAs are typically not spliced and not polyadenylated. This is the consequence of a lack of U1 splice sites and a high frequency of downstream poly(A) cleavage sites, leading to early termination and nuclear RNA exosome-mediated decay (Andersson et al., 2014b; Lubas et al., 2015). The early termination and rapid degradation by the exosome explain why eRNAs are unstable and lowly abundant. Another signature of eRNA transcription is bidirectionality (Figure 1-8), though one of both strands is often preferentially transcribed. Bidirectional transcription initiation seems to be a general feature of transcription applying to enhancers and promoters. Genes appear unidirectionally transcribed only since their antisense transcripts, referred to as promoter upstream transcripts (PROMPTs), are post-transcriptionally degraded by the exosome. eRNAs appear as truly bidirectional transcripts only because both the sense and antisense RNA are sensitive to exosome-mediated decay (Andersson et al., 2014b).

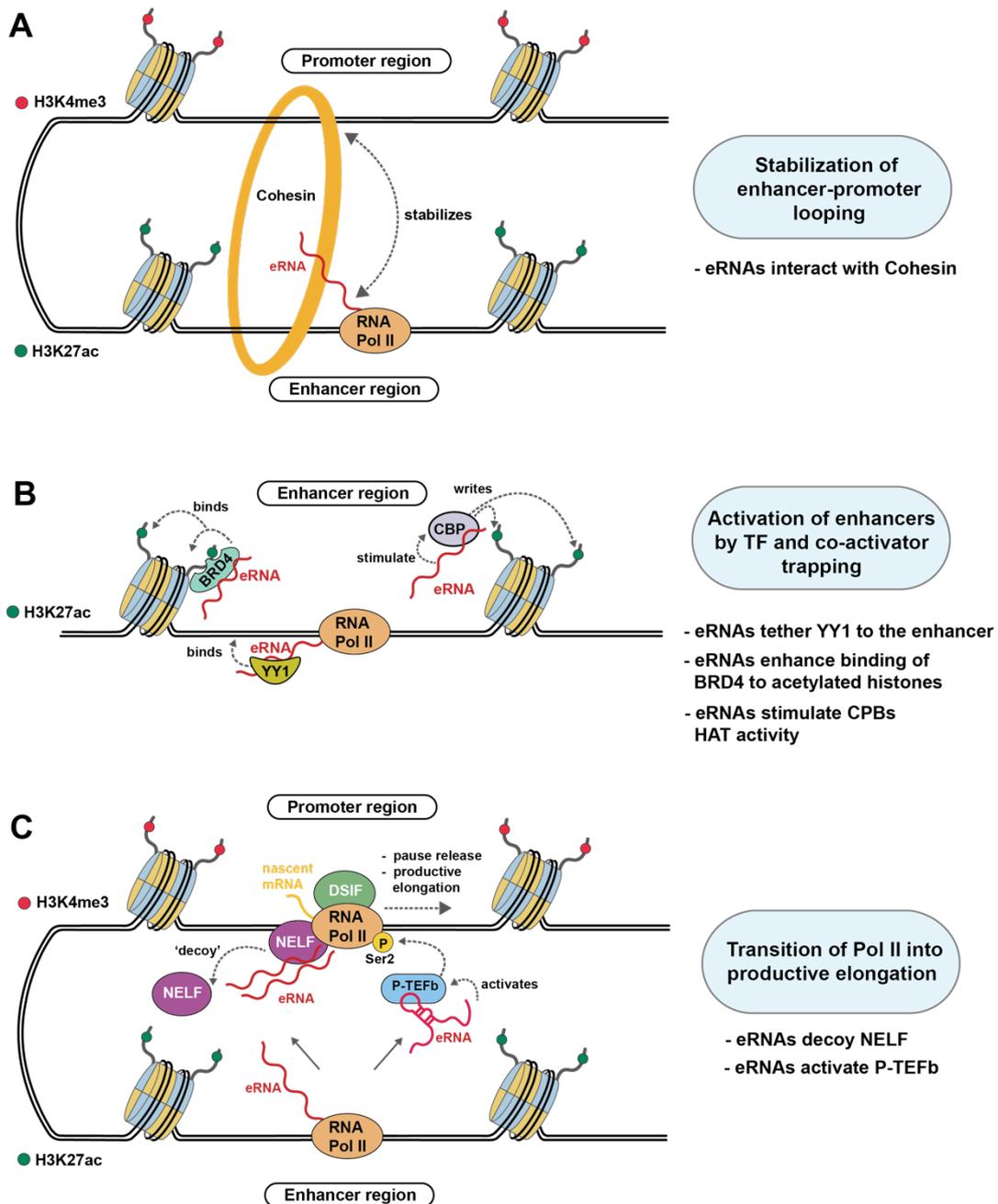
### **1.5.2 Role of enhancer transcription and mechanisms of eRNA function in gene regulation**

Enhancer transcription positively correlates with enhancer activity and the transcription of the cognate target gene. Three mutually non-exclusive scenarios are possible for the functional role of enhancer transcription: 1) Only the act of transcription itself, 2) the actual enhancer RNA molecules, or 3) both are responsible for the activation of target gene transcription. The act of transcription at enhancers very likely impacts the transcription from gene promoters. The presence of Pol II at enhancers influences local chromatin architecture

by recruiting co-activators like acetyltransferases or methyltransferases (Gribnau et al., 2000; Ho et al., 2006; Ling et al., 2005). However, reports that support the importance of the transcription process fail to disentangle the impact of transcription from the impact of the produced eRNA transcripts. Multiple eRNA knockdown studies throughout different cell types, using short-hairpin RNAs (shRNA), small interfering RNAs (siRNAs), or locked nucleic acids (LNA) attested to the functionality of eRNA molecules (Hsieh et al., 2014; Lam et al., 2013; Li et al., 2013b; Mousavi et al., 2013; Schaukowitch et al., 2014).

In general, the knockdown of eRNAs resulted in a reduction of target gene transcription. Complementary experiments using an RNA-tethering strategy further underlined that enhancer RNAs exert a function independent of the act of transcription (Lam et al., 2013; Li et al., 2013b; Melo et al., 2013; Shechner et al., 2015). eRNAs could act either as trappers for transcription factors or as a decoy for transcriptional repressors. Furthermore, eRNAs could be involved in enhancer-promoter looping or the tuning of chromatin accessibility (Li et al., 2016). Indeed, there are numerous reports for subsets of eRNAs supporting each of the hypothetical mechanisms. This highlights the diversity of enhancer RNA functions, acting on different steps of the transcription process (Figure 1-9). Some eRNAs were reported to facilitate the formation or stabilization of enhancer-promoter looping as they interact with subunits of the Cohesin (Li et al., 2013) and the Mediator complex (Hsieh et al., 2014; Lai et al., 2013) (Figure 1-9A). However, this mechanism does not apply to all eRNAs, as shown by studies in neurons and breast cancer cells, in which chromatin loops were not impaired by eRNA knockdowns (Hah et al., 2011; Schaukowitch et al., 2014). In mouse embryonic stem cells (mESCs) nascent enhancer RNAs, but also nascent mRNAs, were found to bind to the ubiquitously expressed transcription factor Yin-Yang 1 (YY1) (Sigova et al., 2015), that later turned out to be a structural regulator of chromatin loops (Weintraub et al., 2017). A CRISPR-Cas9 mediated RNA-tethering assay confirmed that nascent eRNAs could specifically increase YY1 binding to a respective enhancer, suggesting that the promotion of TF binding to the enhancer and promoter region could be a general role of eRNAs. Another study described the binding of enhancer RNAs to the CBP in *cis* (Bose et al., 2017) (Figure 1-9B).

(Rahnamoun et al., 2018)



**Figure 1-9 | Different mechanisms of eRNAs function.** eRNAs facilitate transcription of their target genes by acting mostly in *cis* and deploying multiple mechanisms. **(A)** Oestrogen-dependent eRNAs in human breast cancer cells were shown to facilitate or stabilize enhancer-promoter looping by interacting with the Cohesin complex (Li et al., 2013). **(B)** Other eRNAs were reported to increase the enhancer occupancy of certain transcription factors like YY1 (Sigova et al., 2015) or the co-activator BRD4. The binding of eRNAs, possibly in their nascent form, tethers/traps the factors to the chromatin or stimulates their activity as in the case of CBP (Bose et al., 2017). eRNA interaction with BRD4 enhances its binding to acetylated histones and binding to CBP enhances its histone acetyltransferase activity thereby contributing to the activation of enhancer and promoter region. **(C)** eRNAs were further suggested to promote target gene transcription by facilitating the promoter-proximal pause release of Pol II. Activity-induced neuronal eRNAs are interacting with NELF and are supposed to decoy NELF from the PEC. Another example is an eRNA that was shown to bind and stimulate P-TEFb. Both

mechanisms are potentially leading to an enhanced pause release of Pol II. Adapted from Sartorelli & Lauberth, 2020.

CBP is a co-activator of transcription and a prominent mark of enhancers (see section 1.4.2). Its histone acetyltransferase (HAT) activity can be stimulated by eRNA binding, which leads to the displacement of an autoinhibitory loop from the active site. In turn, the stimulation of CBP facilitates positioning of the activating H3K27ac mark at the enhancer and probably also at the associated promoter region. The H3K27ac modification is known to be bound by the bromodomain-containing protein 4 (BRD4) (Dhalluin et al., 1999; Junwei & Vakoc, 2014), another co-activator of transcription, also shown to interact with eRNAs (Rahnamoun et al., 2018). eRNAs bind to the bromodomains (BDs) of BRD4 and promote its binding to enhancer regions (Figure 1-9B).

A study from 2014 suggested that neuronal eRNAs have a role in the abrogation of promoter-proximal pausing (Schaukowitch et al., 2014). Activity-induced eRNAs, expressed upon KCl-mediated membrane depolarization in mouse cortical neurons, were shown to interact with the pausing factor NELF (see sections 1.2 and 1.3). The knockdown of several neuronal activity-induced eRNAs precluded the release of NELF from the cognate gene promoters leading to reduced transcription of these genes. The authors hypothesized that eRNAs could decoy NELF from the paused Pol II at the promoter-proximal regions, thereby facilitating the pause release (Figure 1-9C). The examined target genes of this study were mainly so-called immediate early genes (IEGs), like *Arc*, *Gadd45b*, and *c-Fos*, whose expression was previously reported to rely on promoter-proximal pausing (Saha et al., 2011). IEGs are rapidly transcribed upon neuronal activity within a short timeframe of only a few minutes. Their expression does not require *de novo* synthesis of TFs, thereby facilitating the immediate signal response of these genes (Bahrami & Drabløs, 2016). The rapid induction of IEGs plays an essential role in synaptic plasticity, which is the basis of learning and memory (Minatohara et al., 2015).

Another study suggests that eRNAs could be involved in the promoter-proximal pause release through binding to P-TEFb. In castration-resistant prostate cancer (CRPC) cells, at least one androgen receptor (AR)-induced eRNA was shown to bind and activate P-TEFb, contributing to the transition of Pol II into productive elongation (Zhao et al., 2016) (Figure 1-9C). eRNA binding to the cyclin T1 subunit of P-TEFb was mediated by an HIV TAR element like (TAR-L) sequence-structure motif. So far, it was the only report linking eRNA's function to

a particular sequence-structure motif. The reports of eRNA interactions with CBP (Bose et al., 2017) and BRD4 (Rahnamoun et al., 2018) indicated that eRNAs function in a non-sequence-specific manner. Last, enhancer RNAs were suggested to be involved in the formation of phase-separated condensates (Nair et al., 2019), which are thought to be a general mechanism for compartmentalization and facilitation of reactions within cells (Hnisz et al., 2017; Polymenidou, 2018). eRNAs seem to play a critical role in regulating the fluid properties of these condensates by promoting a more dynamic and liquid-like state, that is essential for transcriptional activity (Nair et al., 2019). In particular, eRNAs and eRNA binding partners with intrinsically disordered regions tend to form ribonucleoprotein (eRNP) complexes at enhancers. In this context, the NELF-A and NELF-E tentacle, both intrinsically disordered regions of NELF (see section 1.3.2), were recently reported to drive the stress-induced formation of condensates by liquid-liquid phase separation (Rawat et al., 2021). However, the reported sequestering of NELF in those condensates is independent of any RNA and leads to downregulation of transcription.

## 2 Aims and scope of the thesis

Enhancer transcription and the resulting eRNAs were shown to activate target gene expression by deploying various mechanisms. A study of eRNA transcription in mouse cortical neurons hypothesized that activity-induced eRNAs promote target gene transcription, especially of neuronal immediate early genes, by facilitating the release of Pol II from the promoter-proximal paused state. Several eRNAs were found to bind the factor NELF, which is part of the paused elongation complex. The knockdown of these eRNAs impeded the dissociation of NELF from the promoter-proximal regions of the cognate target genes, causing a decrease in their transcription (Schaukowitch et al., 2014). Based on these data it was proposed that eRNAs decoy NELF from the PEC and thereby facilitate the transition of the paused Pol II at the target gene promoters into productive elongation. This mechanism could represent an alternative or additional mechanism to the canonical P-TEFb dependent pause release. The molecular determinants of the eRNA interaction with the PEC remained unsolved. In general, enhancer and their corresponding enhancer RNAs display rather low sequence conservation (Villar et al., 2015; Yang et al., 2015). However, besides the primary RNA sequence, the secondary structure, determining the shape of RNAs, is known to be a major driver for protein-RNA interactions (Li et al., 2014; Sanchez de Groot et al., 2019). Neuronal eRNAs could share a common structural motif responsible for their function.

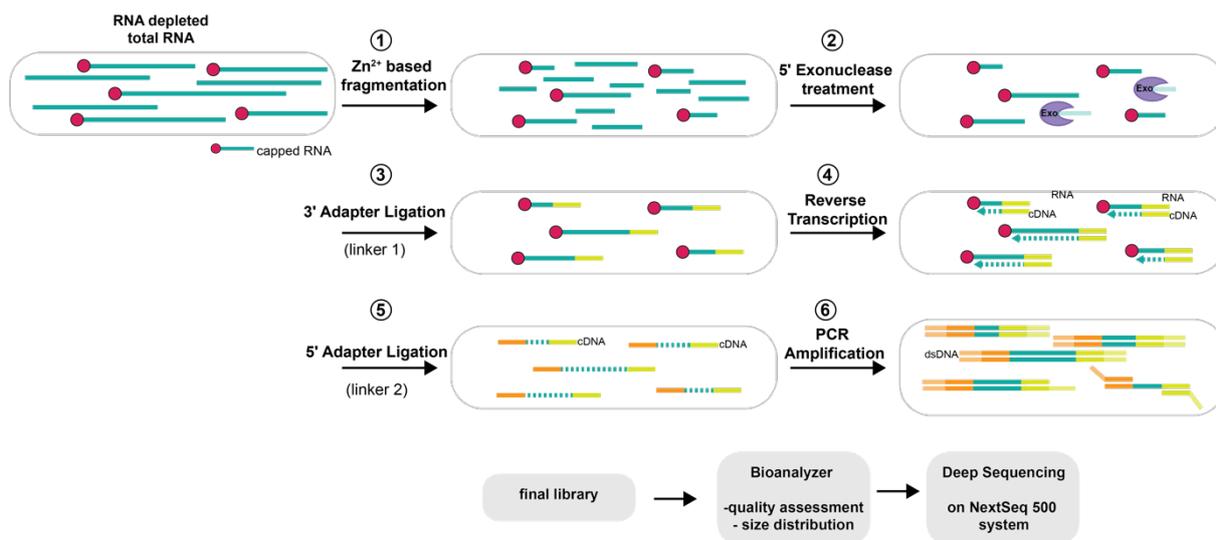
This thesis aimed to decipher the molecular mechanism of the eRNA function and to establish a direct physical link between eRNAs and the Pol II transcription apparatus. To this end, I planned to use the next generation sequencing method Exo-seq to identify first the exact 5' starts (TSSs) of mouse neuronal eRNAs. In a second step, I sought to determine the secondary structure of the obtained eRNA sequences by employing the SHAPE-MaP (2'-Selective hydroxy acylation analyzed by primer extension and mutational profiling) protocol. In addition, I aimed to establish biochemical *in vitro* assays, precisely an EMSA and a transcription assay, to test whether eRNAs can displace NELF from the PEC, thereby promoting transcription. To this end, I intended to reconstitute the PEC *in vitro* from its purified components Pol II, NELF, and DSIF. The combination of the secondary structure probing and *in vitro* assays results should serve to uncover the molecular determinants for the function of eRNAs in promoter-proximal pause release.

## 3 Results

### 3.1 Determination of eRNA 5'-ends by Exo-seq

#### 3.1.1 Establishment of the Exo-Seq protocol

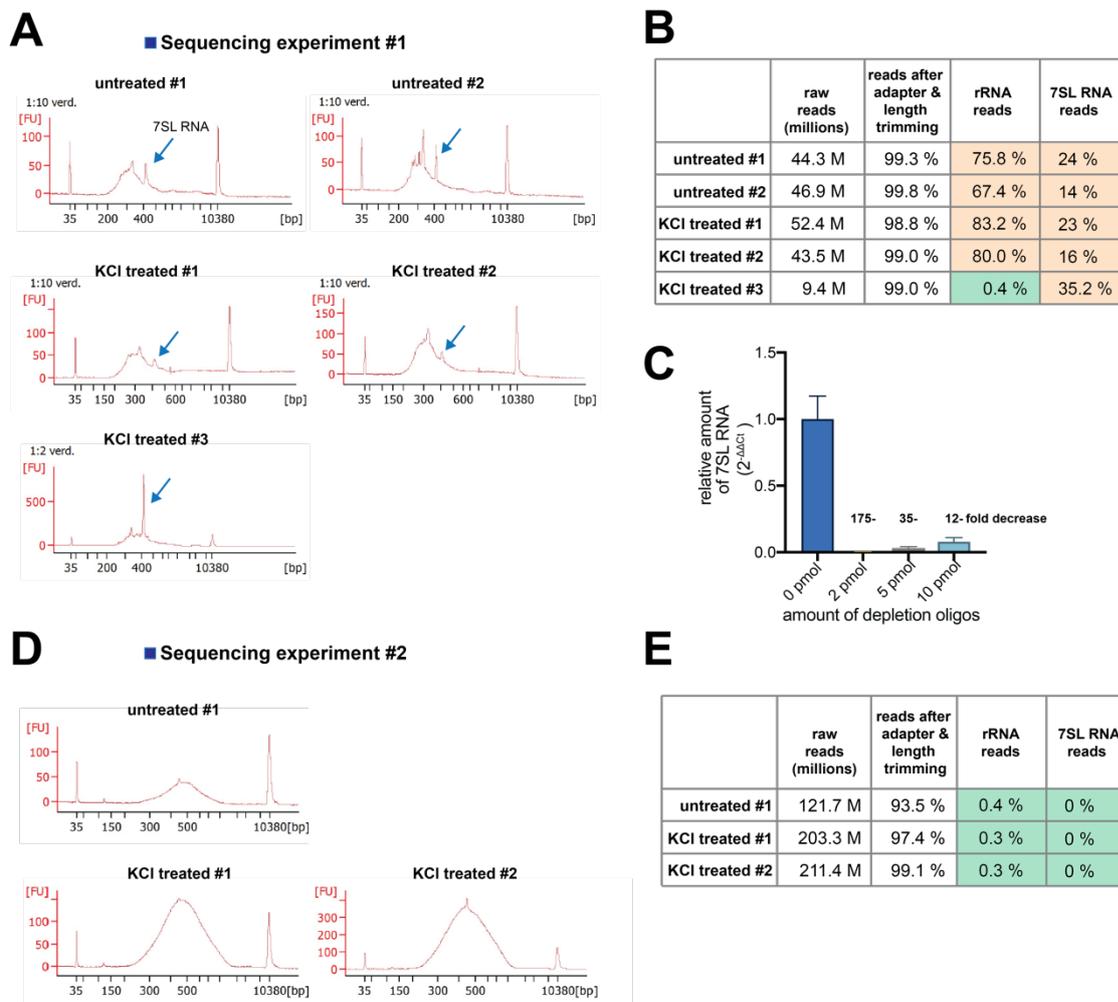
To uncover the molecular mode of enhancer RNA's action, I first sought to determine the exact sequence of the activity-induced enhancer RNA candidates from mouse cortical neurons. There are various library preparation methods for mapping of RNA 5'-ends by high throughput sequencing, like CAGE-seq (Takahashi et al., 2012), 5'-GRO-seq (Lam et al., 2013), PRO-cap (Mahat et al., 2016) and Exo-seq (Afik et al., 2017). The 5'-GRO-seq and PRO-cap belong to the global run-on sequencing methods that can only be performed on living cells and are usually used to capture the exact positions of paused and transcribing Pol II. In our case, the cultivation and handling of mouse cortical neurons was carried out in the laboratory of Dr. Taekyung Kim (UTSW Medical Center, Dallas, USA) and could not be transferred to Bayreuth. To be able to carry out the entire experiment in Bayreuth, I turned to Exo-seq, a robust and easily adaptable protocol that is sufficient to address our questions. The outline of the library construction workflow is schematically shown in Figure 3-1. The protocol relies on the treatment of fragmented total RNA with a 5'-phosphate dependent exonuclease that removes all uncapped fragments, thereby enriching for the capped 5'-ends.



**Figure 3-1 | Schematic representation of the Exo-seq workflow.** The main steps of the Exo-seq library preparation protocol are depicted above.

The original Exo-seq protocol was established to determine mRNA 5'-ends and thus started with a poly (A) selection (Afik et al., 2017). To adapt the Exo-seq protocol for capturing enhancer RNAs, required ribosomal RNA depletion instead of poly (A) selection, as enhancer RNAs are not polyadenylated, yet 5'-capped (De Santa et al., 2010; Kim et al., 2010). For the construction of Exo-seq libraries, I started with total, DNase I treated, RNA from potassium chloride (KCl)-treated and untreated mouse cortical neurons. Purified total RNA from mouse cortical neurons was provided by Dr. Katie Schaukowitch and Dr. Seung-Kyoon Kim (former group members of Dr. Taekyung Kim at the Department of Neuroscience UTSW, Dallas, USA). They prepared and handled the primary neuronal cell culture. The KCl-treatment leads to neuronal depolarization and rapidly induces the transcription of activity-induced enhancer RNAs, which in turn facilitate the transcription of target genes (Kim et al., 2010; Schaukowitch et al., 2014).

I prepared the first Exo-seq libraries after rRNA depletion using the commercial Ribo-Zero rRNA Removal Kit (Illumina), as described in sections 5.13 and 5.14. Five libraries were then sequenced in total, two biological replicates of untreated (-#1 and -#2) and three biological replicates of KCl-treated (-#1, -#2 and -#3) neurons. Before sequencing, the libraries were examined by high-resolution electrophoresis on a Bioanalyzer, to assess their quality and fragment size distribution (measurements were carried out by the Core Unit Systems Medicine staff at the University of Würzburg). The obtained electropherograms were very spiky and displayed a peak at around 400 bp. This peak was especially prominent in the KCl-treated library #3 (Figure 3-2A). In contrast to the other four libraries, this sample was produced from less RNA starting material (3  $\mu$ g instead of 5  $\mu$ g). Analysis of the sequenced reads revealed that the rRNA depletion failed for all libraries, except for the KCl-treated #3. Up to 80 % reads from those libraries mapped to rRNA, while in library #3 only 0.4 % of reads originated from rRNA (Figure 3-2B). This explains why the peak at 400 bp was more prominent in library #3, as it is not obscured by rRNA reads. The reason for the successful rRNA depletion in library #3 was highly likely the lower amount of starting material. These results suggest that taking the upper limit of 5  $\mu$ g as starting material (according to the manufacturer's protocol of the Illumina kit) is not advisable for efficient rRNA depletion.



**Figure 3-2 | Characteristics of Exo-Seq libraries before and after optimization.** (A) Electropherograms of the prepared Exo-seq libraries, examined on a Bioanalyzer before the first sequencing experiment. In retrospect the prominent peak, marked with the blue arrow, at around 400 bp could be assigned to the non-coding 7SL RNA (signal recognition particle RNA). (B) Statistics of the mapped reads for each library. The number of raw reads is shown rounded in millions. The percentage of the rRNA reads is calculated in relation to the absolute number of reads after trimming. The percentage of the 7SL RNA reads is calculated in relation to the absolute number of reads after sorting out the rRNA reads. (C) Depletion efficiency of 7SL RNA from total RNA was determined by qPCR using the relative quantification ( $2^{-\Delta\Delta C_T}$  method) (Livak & Schmittgen, 2001); against *Gapdh* as reference. (D-E) Same as described in (A) only for libraries after optimized rRNA and 7SL RNA depletion.

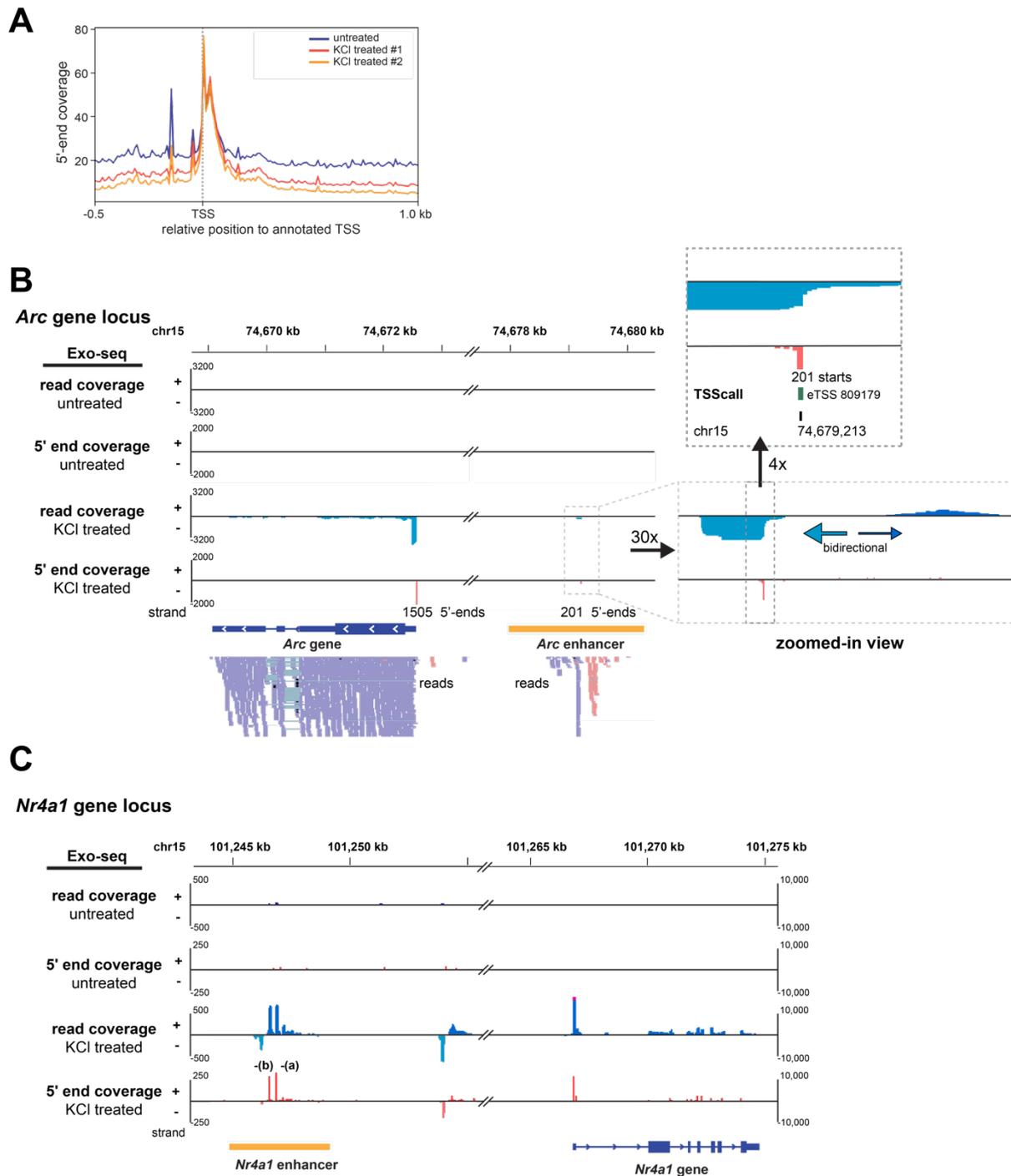
Furthermore, a considerable fraction of reads (up to 35% of all reads after rRNA sorting) in all libraries were originating from the signal recognition particle RNA (7SL RNA) (Figure 3-2B). 7SL RNA is a 300-nucleotides (nt) long non-coding RNA and is characterized by a pronounced double-stranded structure (Pool, 2005), which probably protects the RNA from degradation by the 5' exonuclease during the Exo-seq protocol (Figure 3-1). The prominent presence of such long non-coding RNAs in rRNA depleted libraries was previously described (Sultan et al., 2014) and is very unfavorable as it lowers the sequencing depth of the desired RNAs and

impedes the detection of low abundant RNAs, like eRNAs. Indeed, the read coverage at the enhancers was very low and the Exo-seq experiment was repeated with an optimized rRNA and 7SL RNA depletion protocol.

I thus set out to establish an improved Exo-seq protocol in which the 7SL RNA is depleted together with the rRNA using the same principle (section 5.13). The rRNA depletion procedure (provided by the utilized kit) relies on biotinylated antisense DNA probes against rRNA sequences. In a first step, the DNA probes are hybridized to the rRNA molecules, and in a second step, the formed RNA-DNA hybrids are specifically pulled out from the sample by streptavidin-coated magnetic beads (Kim et al., 2019; Kraus et al., 2019). I designed four antisense probes against the 7SL RNA and combined them into a depletion mix (1:1:1:1 ratio). The depletion efficiency of the antisense oligos was verified by qPCR (Figure 3-2C). 2 pmol of the depletion mix already caused a 175-fold decrease of 7SL RNA in the total RNA sample. For generating Exo-seq libraries, rRNA and 7SL RNA were co-depleted in a one batch reaction using the customized 7SL depletion oligos combined with another rRNA depletion Kit (Lexogen) (section 5.13). For the sake of a higher sequencing depth, we decided to sequence KCl-treated libraries in duplicate and only one untreated library. The electropherograms of the library samples and mapping of the sequencing reads confirmed the successful depletion of both rRNA and 7SL RNA (compare sequencing #1 Figure 3-2D,E with sequencing #2 Figure 3-2A,B).

### 3.1.2 Analysis of Exo-seq data

Metagene analysis of the Exo-seq reads 5'-ends attested their strong enrichment exactly at the position of transcription start sites (TSSs) of annotated mouse genes (RefSeqGenes) (Figure 3-3A). This result confirms that the Exo-seq data will be suitable for the precise determination of the eRNA 5'-ends. The additional, less pronounced peak directly downstream of the annotated TSS position points at the presence of alternative TSSs and/or some misannotation of TSS (Figure 3-3A) (Reyes & Huber, 2018; Xu et al., 2019). Figure 3-3B and C show Exo-seq coverage profiles at the gene and enhancer regions of two prominent, neuronal immediate early genes (IEGs), *Arc* and *Nr4a1*. The high Exo-seq signal at the gene and enhancer locus only observed for the KCl treated libraries emphasizes the strong induction of immediate early genes and their enhancers upon KCl stimulation.



**Figure 3-3 | Exo-seq allows for determining enhancer TSSs with single-nucleotide precision. (A)** A metaplot of the Exo-seq reads 5'-end coverage confirms the enrichment of Exo-seq signal at transcription start sites of annotated mouse genes (RefSeqGene). **(B-C)** Genome browser view (IGV browser) of Exo-seq data at the *Arc* (B) and *Nr4a1* (C) genomic loci. Exo-seq read coverage track and the coverage tracks for the 5'-ends only (5' end coverage; single-nucleotide resolution) in the untreated library (untreated #1) and the KCl treated library (KCl treated #2). Enhancer regions are labeled below the tracks (orange bars) and for the *Arc* enhancer a more detailed zoomed-in view is shown to the right, clearly showing the single-nucleotide resolution of the Exo-seq signal (201 5'-ends for *Arc* eRNA (minus/anti-sense strand) at the position 74,679,213 on chromosome 15; TSScall peak 809179). For *Nr4a1* eRNA the Exo-seq signal showed two prominent peaks on the plus/sense strand, indicating two alternative transcription start sites, in this work referred to as -(a) and -(b).

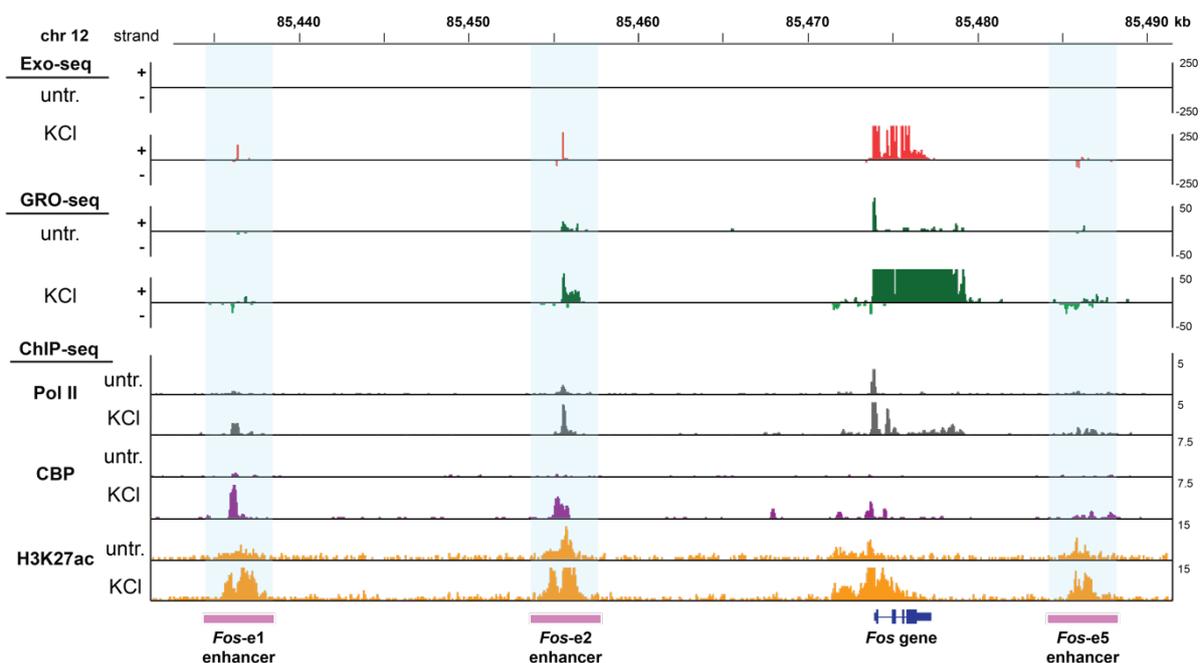
A zoomed-in view (Figure 3-3B shown to the right) or different y-axis scales (Figure 3-3C) are needed to visualize the coverage at gene and enhancer regions simultaneously, as the enhancer transcripts are much less abundant compared to the mRNA. The tracks further demonstrate the characteristic bi-directionality of eRNAs, with typically one preferentially expressed strand, as the (-)-strand (anti-sense strand) in the case of *Arc* eRNA (Figure 3-3B) or the (+)-strand (sense strand) as in the case of *Nr4a1* eRNA (Figure 3-3C). The two examples in Figure 3-3 illustrate that the Exo-seq data allows to identify the exact 5'-ends of neuronal enhancer RNAs.

The following computational analysis was mainly conducted by Andreas Pittroff under my supervision. To identify enhancer RNA 5'-ends genome-wide, initially, all extragenic 5'-ends were determined with single-nucleotide precision using the TSScall tool (Henriques et al., 2018). Next, the Exo-seq data were combined with GRO-seq data provided by our collaborator Dr. Taekyung Kim (Dr. Seung-Kyoon Kim carried out GRO-seq experiment). The GRO-seq method identifies nascent transcripts and the position of transcriptionally engaged Pol II (Danko et al., 2013) and thus substantiated the Exo-seq data that were based on mature RNA. Figure 3-4 shows an example of the overlap of Exo-seq signal with GRO-seq signal and further enhancer-specific ChIP-seq signals (publicly available data) for the *c-Fos* gene locus, another prominent neuronal IEG. Based on the GRO-seq data Dr. Seung-Kyoon Kim identified 1,226 intergenic eRNA transcription units *de novo* based on the overlap with intergenic regions (beyond a  $\pm 2$  kb window from gene TSSs and gene bodies) (Gorbovytska et al., 2021), that are enriched for the active enhancer mark H3K27ac (see section 1.4.2) (Malik et al., 2014). 252 of these enhancers were activity-induced, as defined by a >1.5-fold increase of eRNA GRO-seq signal (comparison of untreated and KCl treated GRO-Seq signal). Subsequently, we intersected the list of our TSSs with the GRO-seq based list of eRNAs provided by our collaborators. For 304 (281 for replicate #2) of the 1,226 enhancer units we found well-defined 5'-ends (>20 reads per eTSS). 86 (79 for replicate #2) of these eRNAs with precisely determined 5'-ends (79 for replicate 2) originated from activity-induced enhancers. Importantly, if multiple TSSs were called by TSScall, we kept the most prominent eRNA start site for each enhancer locus (Gorbovytska et al., 2021). In further confirming the validity of this analysis, we detected eRNA TSSs of prominent IEGs such as *Arc*, *Nr4a1*, *Junb*, *c-Fos* (enhancers e1, e2 and e5) and *Fosb*, for some of which eRNA expression had been reported before (Kim et al., 2010). After closer inspection of the Exo-Seq reads at the selected eRNA

transcription start sites, all sites with pervasive transcription or convergent transcription were excluded and a final list of 33 high-quality eRNA candidates was compiled.

Twenty-four of these eRNAs originated from activity-induced enhancers, as determined by GRO-seq. Moreover, 7 of the 33 eRNA candidates, amongst them *Junb* and *Nr4a1*, showed distinct, likely alternative transcription start sites. For these eRNAs we therefore included two separate eRNAs (termed (a) and (b)) (see Figure 3-3C), which resulted in a total of 39 eRNAs in our test set (supplementary Table 7-22) (Gorbovytska et al., 2021).

#### *Fos* gene locus



**Figure 3-4 | Exo-seq signal at active enhancers overlaps with ChIP-seq enhancer marks.** Genome browser view of the *c-Fos* gene and three of its surrounding enhancer loci, displaying Exo-seq tracks from this study, GRO-seq, and publicly available ChIP seq tracks for Pol II, CBP and H3K27ac (GEO record GSE60192 and GSE21161). Tracks for both untreated (untr.) and KCl treated mouse cortical neurons are shown. The increase of CBP and H3K27ac signals, both marks for enhancers, as well as the increase of GRO-seq signal upon KCl treatment denote activity-induced enhancers. Yet unpublished GRO-seq data were provided by Dr. Seung-Kyoon Kim.

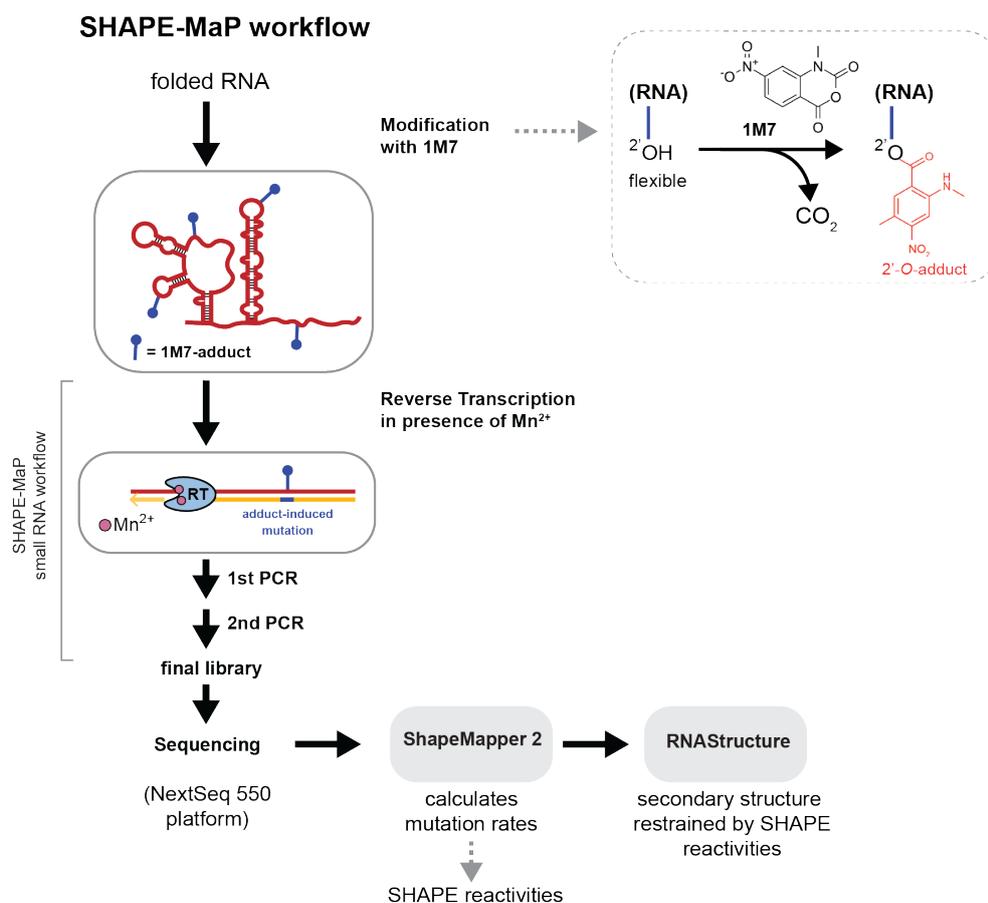
## 3.2 Mapping of eRNA secondary structure

### 3.2.1 Structure probing of eRNAs by *in vitro* SHAPE-MaP

Target gene induction in cortical neurons depends on eRNA production (Schaukowitch et al., 2014), yet there is no common sequence motif that could account for eRNAs function. Hence, we speculated that eRNAs might exhibit a common structural motif to exert their function. I thus set out to experimentally determine the secondary structures of all 39 eRNA candidates, for which we had identified the precise eTSS (Figures 3-3, 3-4, and supplementary Table 7-21) by using chemical structure probing. eRNAs are reported to have a median length of ~ 1kb (Schwalb et al., 2016). Yet we decided to focus the secondary structure analysis on the 5'-terminal 200 nucleotides of the eRNAs for the following two reasons. First, the few already experimentally determined lncRNA structures indicate that lncRNA domains comprise approximately 200 nucleotides (Qian et al., 2019). Second, eRNAs are stabilized at their 5'-ends by a 5'-cap structure, while their 3'-ends are prone to exosome-mediated decay as they are not polyadenylated (Andersson et al., 2014b). Thus the 5' terminal end of the eRNAs is most likely the functionally relevant part of the eRNA.

To map the secondary structures of all 39 eRNAs at once, I chose a high-throughput method for chemical structure probing, called SHAPE-MaP (Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling) (Siegfried et al., 2014; Smola, Rice, et al., 2015). The SHAPE-MaP experiment overview is outlined in Figure 3-5. For SHAPE-MaP, the RNA is treated with a small electrophilic reagent, typically 1-methyl-7-nitroisatoic anhydride (1M7), which selectively reacts with the ribose 2'-hydroxyl group of structurally flexible nucleotides (unpaired nucleotide e.g., in hairpin loops, internal loops, bulges). The modified RNA is reverse transcribed in the presence of  $Mn^{2+}$  ions. These conditions allow for the reverse transcriptase (RT) to read through the modified sites and induce a mutation at these positions. The final libraries are sequenced, and mutation rates are extracted. These mutation rates are then transformed into SHAPE-reactivities. Flexible, unpaired nucleotides display high SHAPE-reactivities while structurally constrained, paired nucleotides, as in double-stranded regions, display low SHAPE-reactivities. SHAPE-MaP has several advantages over other high-throughput structure probing techniques, like DMS-seq (Rouskin et al., 2014a), Mod-seq (Talkish et al., 2014), or SHAPE-seq (Lucks et al., 2011). First, 1M7 reacts with flexible

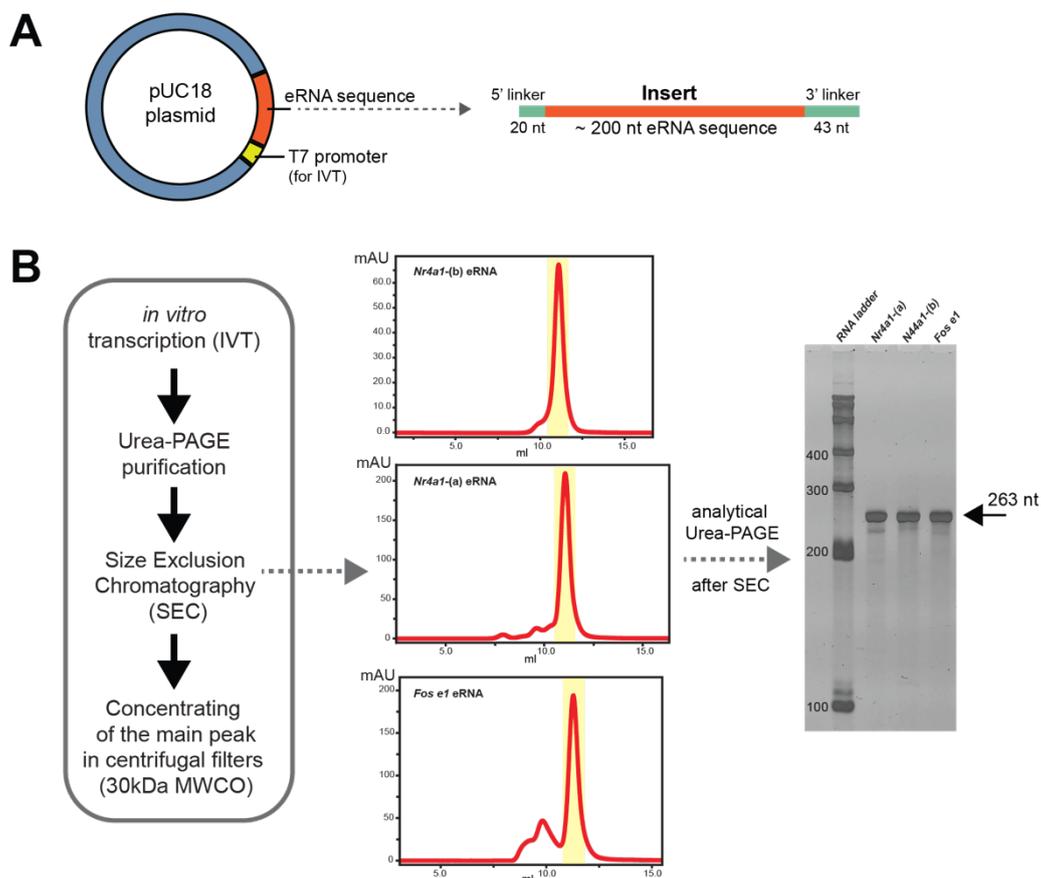
nucleotides, regardless of the associated base (A, C, G or U), thereby providing single-nucleotide resolution. Second, the SHAPE-MaP protocol omits adapter ligation steps, which are often inefficient and known to introduce bias to the experimental outcome. Third, other protocols include a reverse transcription step where the RT terminates at the sites of modification producing cDNA fragments of different sizes that also lead to a bias during the library construction. However, the SHAPE-MaP protocol produces uniformly sized dsDNA libraries.



**Figure 3-5 | Overview of the main SHAPE-MaP steps and subsequent analysis tools.** In the first step ribose 2'-OH residues in flexible RNA regions are modified with the chemical 1M7. The sites of modification (1M7-adducts) are recorded during reverse transcription in the presence of manganese ions (Mn<sup>2+</sup>) by allowing the read through of the reverse transcriptase (RT) causing the integration of a non-complementary nucleotide (adduct-induced mutation) this site. The rest of the library construction path used in this thesis follows the SHAPE-MaP small RNA workflow (Smola, Rice, et al., 2015). After sequencing of the final libraries on a next generation sequencing platform (e.g. NextSeq 550 in this study), the ShapeMapper 2 software (Busan & Weeks, 2017) is utilized, that calculates SHAPE-reactivities based on mutation rates. The obtained SHAPE-reactivities are then used as restrains to build the secondary structure of the RNA using the RNAstructure prediction tool (Reuter & Mathews, 2010).

### 3.2.2 Production of eRNAs for *in vitro* SHAPE-MaP

To probe the secondary structure of the 5'-terminal 200 nucleotides of the selected 39 eRNAs, the sequences were cloned from mouse cortical neurons cDNA into pUC18 plasmids (Figure 3-6A). The eRNA insert was flanked by a 5' and 3' linker sequence, that allowed to process all eRNAs with a universal reverse transcription primer and PCR primers during the SHAPE-MaP library construction. Furthermore, these linker prevent the loss of structure information at the actual 5' and 3' end of the eRNA during the PCR (section 5.16.1), as all 1M7-adduct induced mutations at the RNA ends would be obscured by the PCR primers. The linker sequences were derived from previously published ones (Lucks et al., 2011; Merino et al., 2005; Wilkinson et al., 2006) and ought to fold independently into stable hairpin loops without interfering with the fold of the eRNA insert. A T7 RNA polymerase promoter was used to generate the eRNAs by *in vitro* transcription (IVT) (section 5.9) (Figure 3-6A,B). The produced eRNAs were purified by urea-PAGE and size exclusion chromatography to obtain a homogeneous, monodisperse sample for each eRNA.

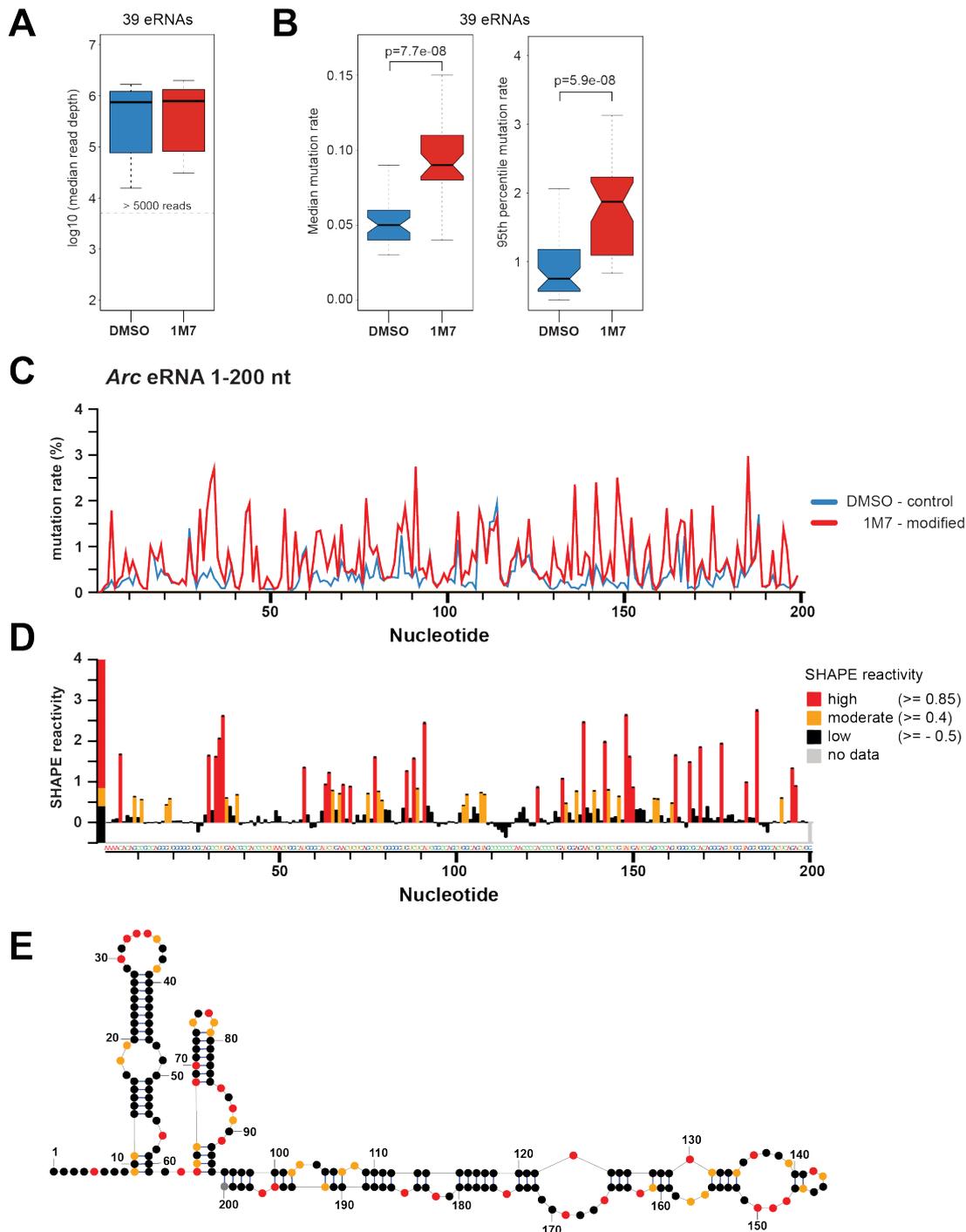


**Figure 3-6 | Preparation of eRNAs for SHAPE-MaP.** (A) eRNA (1-200 nt) fragments were cloned into pUC18 vectors with a preceding T7 promoter, required for *in vitro* production of the eRNAs. The flanking linker sequences were needed for the production of the SHAPE-MaP libraries. (B) shows the eRNA production and

purification workflow (left panel) along with example size exclusion chromatograms for three different eRNAs (middle panel) and the corresponding analytical 6% Urea-gel stained with Sybr Gold (right panel). 5 pmol of RNA were loaded per lane and the RiboRuler low range RNA ladder was used as a standard.

### 3.2.3 Analysis of the SHAPE-MaP data

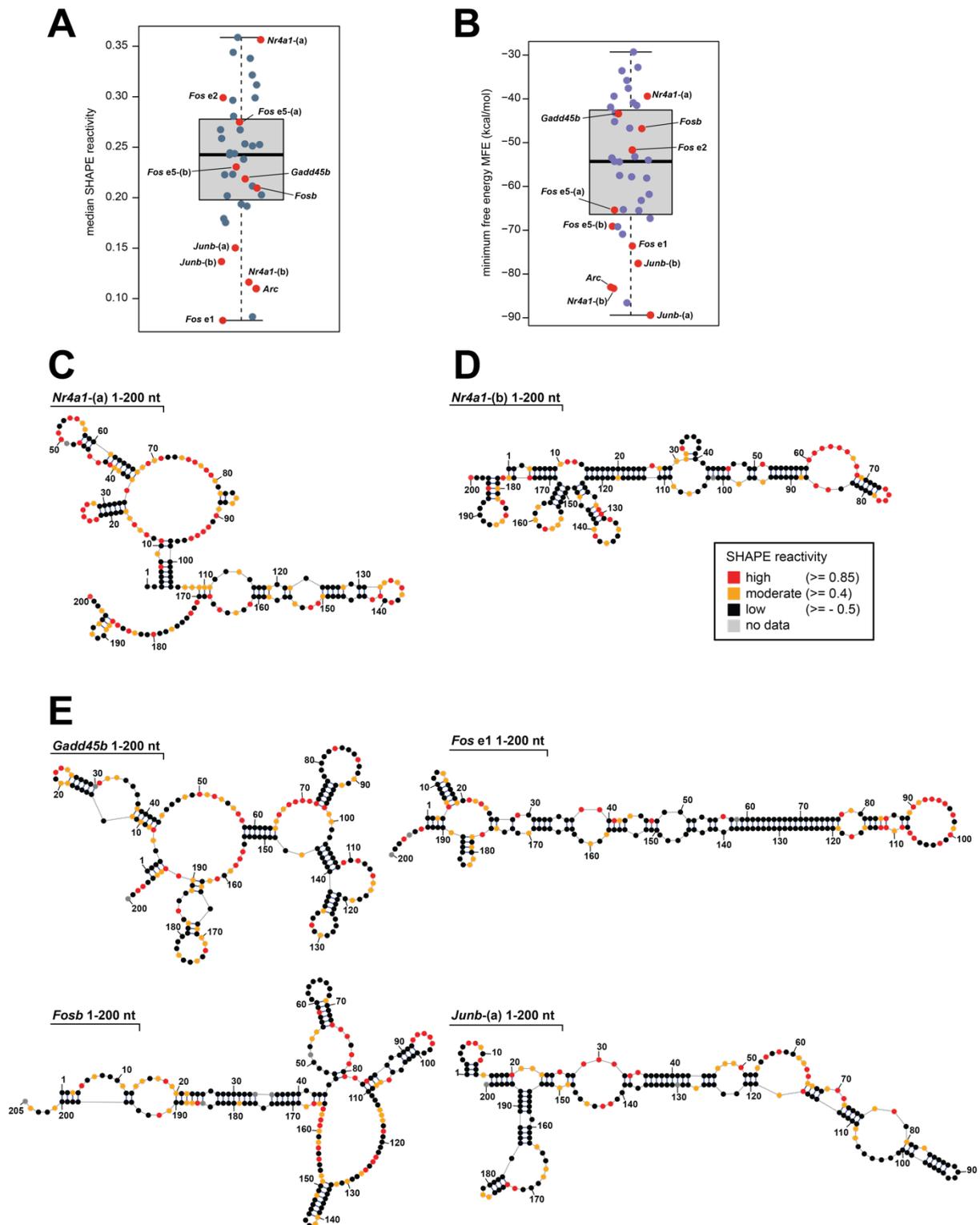
I prepared SHAPE-MaP libraries from 1M7-modified and DMSO-treated (control) eRNAs (section 5.16.1). After sequencing the SHAPE-MaP libraries, I calculated the mutation rates and SHAPE reactivities using ShapeMapper 2 (Busan & Weeks, 2017). The obtained SHAPE reactivities were applied as restrains for RNA secondary structure prediction using the RNAstructure tool (Reuter & Mathews, 2010) (Figures 3-5; Figure 3-7A-D). At the initially performed SHAPE-MaP experiment, the 39 folded eRNAs were pooled in an equimolar fashion at the beginning of the library preparation prior to the 1M7-treatment. However, with this approach I could reliably map the structure for only 17 out of the 39 eRNAs, as the read depth for the rest eRNAs was not sufficient (< 5000 reads). In general, the number of reads was very unevenly distributed between all eRNAs. Thus, I repeated the experiment for the remaining eRNAs. I prepared individual libraries for each eRNA and pooled them equimolarly just before sequencing. Using this approach significantly improved the distribution of sequenced reads. Figure 3-7A shows the distribution of the median read depth for all 39 eRNAs. About 50 % of the 39 eRNAs have a read depth of about 1 M, which is above the threshold of 5000 reads (Siegfried et al., 2014). Furthermore, the mutation rates of the 1M7 sample are significantly above the background mutation rates of the DMSO sample (Figure 3-7B). Both the large read depth and the mutation rates over the background allow for accurate modeling of the secondary structure. The read depth and the mutation rates for the DMSO-control and 1M7-treated eRNA are listed in the supplementary Table 7-24. Figure 3-7C and D show examples for the mutation rate and the SHAPE reactivity profile obtained for the *Arc* eRNA using ShapeMapper 2. The SHAPE reactivity restrained structure computed with RNAstructure is displayed in Figure 3-7E. The fold of *Arc* eRNA 1-200 nt appeared to be very structured, comprising many double-stranded regions and hairpins. Lastly, to compare all mapped eRNA structures, I used the tool locaRNA that combines sequence and structure alignment of RNAs to calculate a consensus structure. However, the comparison resulted in no common structural motif, neither for the entire set nor for subsets of the eRNAs.



**Figure 3-7 | SHAPE-MaP results for *Arc* eRNA.** (A) The distribution of median read depth for the mapped 39 eRNAs in the DMSO-control and 1M7-modified condition is sufficient (> 5000 reads) for accurate structure mapping. (B) The obtained mutation rates for the 1M7 treatment lie significantly above the background mutation rates of the DMSO-control, as determined by the nonparametric Mann-Whitney-Wilcoxon test (for paired samples). (C) and (D) shows the mutation rate and the resulting SHAPE reactivity profile for *Arc* eRNA 1-200, as computed by ShapeMapper 2 (Busan & Weeks, 2017). High ( $\geq 0.85$ , red) and moderate ( $\geq 0.4$ , orange) SHAPE reactivities mark flexible, unpaired nucleotides. Low SHAPE reactivities ( $< 0.4$ , black) denote less-flexible nucleotides that are predominantly located in double-stranded regions. (E) shows the MaxExpect structure (structure with most probable base pairs) generated with RNAstructure using the SHAPE reactivities from (D) as constraints. Nucleotides are shown as circles and are conditionally colored according to their SHAPE reactivities. (C)-(E) are adapted from Gorbovytska et al., 2021.

Hence, I decided to take a more general look at the acquired structure data. Instead of having a specific structural motif, enhancer RNAs might exhibit a generally high "structuredness" that could help to exert their function. *Arc* eRNA 1-200 nt displays a very structured fold comprising many double-stranded regions and several hairpins (Figure 3-7E). This raised the presumption that "structuredness" could indeed be a general feature of eRNAs. To estimate the overall "structuredness" of all 39 eRNAs, I compared their median SHAPE reactivity (Figure 3-8A) and the minimum free energy (MFE) of the lowest MFE structure (Figure 3-8B) (Zuker & Stiegler, 1981). Low SHAPE reactivity and low MFE would imply a high structuredness of eRNA. The analysis revealed that both parameters, the median SHAPE reactivity and the MFE, show a relatively broad distribution, without any particular clustering at low SHAPE reactivity or minimum free energy values. Low median SHAPE reactivities of 0.1 and less were previously reported for the highly structured tRNAs or abundant non-coding RNAs in *E. coli*. In contrast, the median SHAPE reactivity for coding RNAs was shown to extend from 0.15 to 0.35 (Mustoe et al., 2018). The median SHAPE reactivity of our mapped eRNAs ranges from 0.08 to 0.36 for *Fos* e1 and *Nr4a1*-(a) eRNA, respectively (Figure 3-8A). The range for the median SHAPE reactivity matches the distribution for coding RNAs rather than for tRNAs. The obtained MFE values for each eRNA (Figure 3-8B) largely correlate with the median SHAPE reactivity (Figure 3-8A) and range from -90 to -30 kcal/mol (Figure 3-8B) with a median MFE of -54 kcal/mol. When we normalize the MFEs to the length of eRNAs (200 nucleotides), by dividing by 200 and multiplying by 100, a widely used simple normalization method (Trotta, 2014; Zhang et al., 2006), we obtain adjusted MFEs (median MFE = -27 kcal/mol). The median adjusted MFE is comparable to the previously reported MFE for eRNA structures (median MFE = -32.5 kcal/mol) which was based on RNA fold predictions (Schwalb et al., 2016). This finding supports that eRNAs seem to be generally less structured than mRNAs, which were reported to have a median MFE of 47 kcal/mol.

Notably, five of the top six most structured eRNAs in our set, with a median SHAPE reactivity and minimum free energy below 0.15 and -70 kcal/mol (normalized: -35 kcal/mol), respectively, belong to immediate early genes (*Arc*, *Nr4a1* variant (b), *Junb* variant (a) and (b) and *Fos* enhancer 1 (e1) eRNAs) (Figure 3-8A,B). The SHAPE-MaP based secondary structures of these eRNAs are characterized by a high proportion of double-stranded regions and less prominent single-stranded regions (Figure 3-7E, Figure 3-8D, Figure 3-8E right panel, and the supplementary Figure 7-3.)



**Figure 3-8 | eRNAs display a broad spectrum of structures.** (A) Distribution of the median SHAPE reactivity for the 39 eRNAs subjected to SHAPE-MaP. (B) Distribution of the minimum free energy of the lowest energy (MFE) structures. Data points of immediate early gene eRNAs are highlighted in red. The SHAPE reactivity constrained secondary structures of *Nr4a1* variant (a) and (b) are shown in (C) and (D), respectively. (E) shows secondary structures of further immediate early gene eRNAs (left panel shows less structured eRNAs; right panel shows highly structured eRNAs as defined by median SHAPE reactivity and MFE (see A,B)). The structures shown in (C)–(E) are the MaxExpect structures predicted by the RNAstructure tool. Bases are shown in circles and are colored conditionally upon SHAPE reactivities.

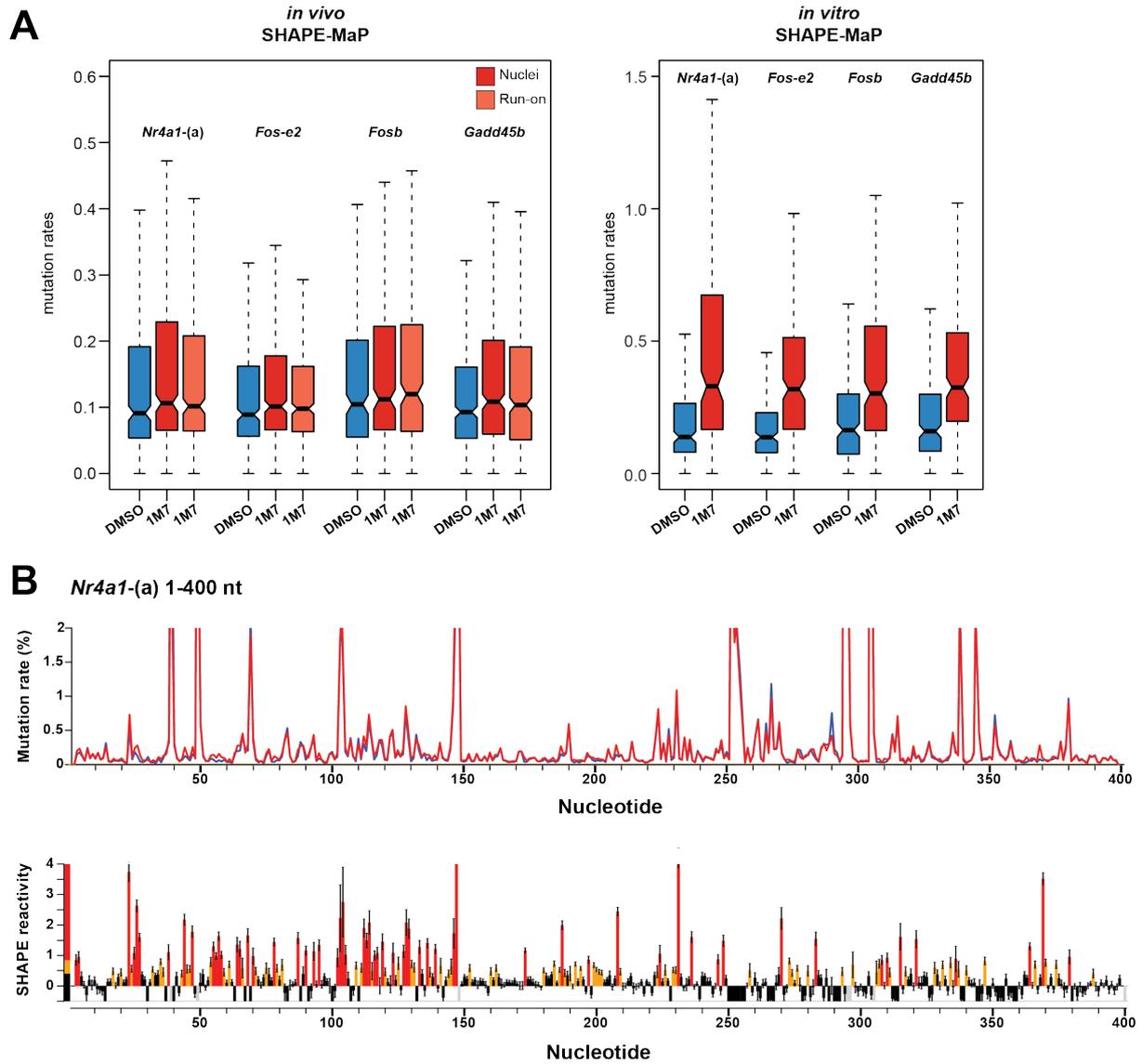
However, the additional eRNAs of *Fos* (-e2 and -e5) (Figure 3-4) and other prominent neuronal IEGs such as *Gadd45b* and *Fosb* (Kuroda et al., 2008; Ma et al., 2009) exhibit relatively high median SHAPE reactivities and MFEs ( $> 0.2$  and  $> -70$  kcal/mol). Thus, these eRNAs display a less structured fold, as reflected by longer, extended single-stranded and few double-stranded regions (Figure 3-8E left panel and supplementary Figure 7-3). The least structured IEG eRNA is *Nr4a1* variant (a) (Figure 3-8C) with an MFE of  $-39$  kcal/mol (normalized:  $19.5$  kcal/mol). In summary, the experimental SHAPE-MaP data on the 5'-terminal 200 nucleotides of eRNA demonstrate that eRNAs populate a wide range of structural space without sharing a common structural motif. Though the most structured eRNAs are coming from enhancers of immediate early genes, a high degree of "structuredness" does not appear to be a general feature of eRNAs, neither for IEG only nor all eRNAs.

### 3.2.4 *In vivo* SHAPE-MaP

I also aspired to perform *in vivo* SHAPE-MaP to probe the secondary structure of enhancer RNAs within cells. It is known that the *in vivo* structure of RNAs can deviate from its secondary structure under *in vitro* conditions. Genome-wide studies report that, generally, RNAs seem to adopt a less structured fold in the cell than under cell-free conditions (Mustoe et al., 2018; Rouskin et al., 2014b; Spitale et al., 2015). The complex cellular environment with the presence of protein binding partners and the effect of molecular crowding naturally impacts the RNA structure. *In vivo* SHAPE-MaP, also referred to as in-cell SHAPE-MaP, was previously successfully performed on several cell types, such as mouse stem cells and *E. coli* (Mustoe et al., 2018; Smola, Calabrese, et al., 2015; Smola & Weeks, 2018). However, the analyzed RNAs were mostly cytoplasmic and highly abundant mRNAs or rRNAs. To react with the lowly abundant eRNAs, the modifying agent 1M7 must reach the nucleus.

The 1M7 (and DMSO)-treatment of stimulated mouse cortical neurons was carried out by Dr. Seung-Kyoon Kim at the UTSW Medical Center (Dallas, USA). For this, 1M7 was directly added to mouse cortical neurons during KCl stimulation. Then nuclei were isolated to enrich the final RNA sample for the nuclear, low abundant eRNAs (*Nuclei* sample). Half of the isolated nuclei were used for nuclear run-on in the presence of additional 1M7 in order to achieve

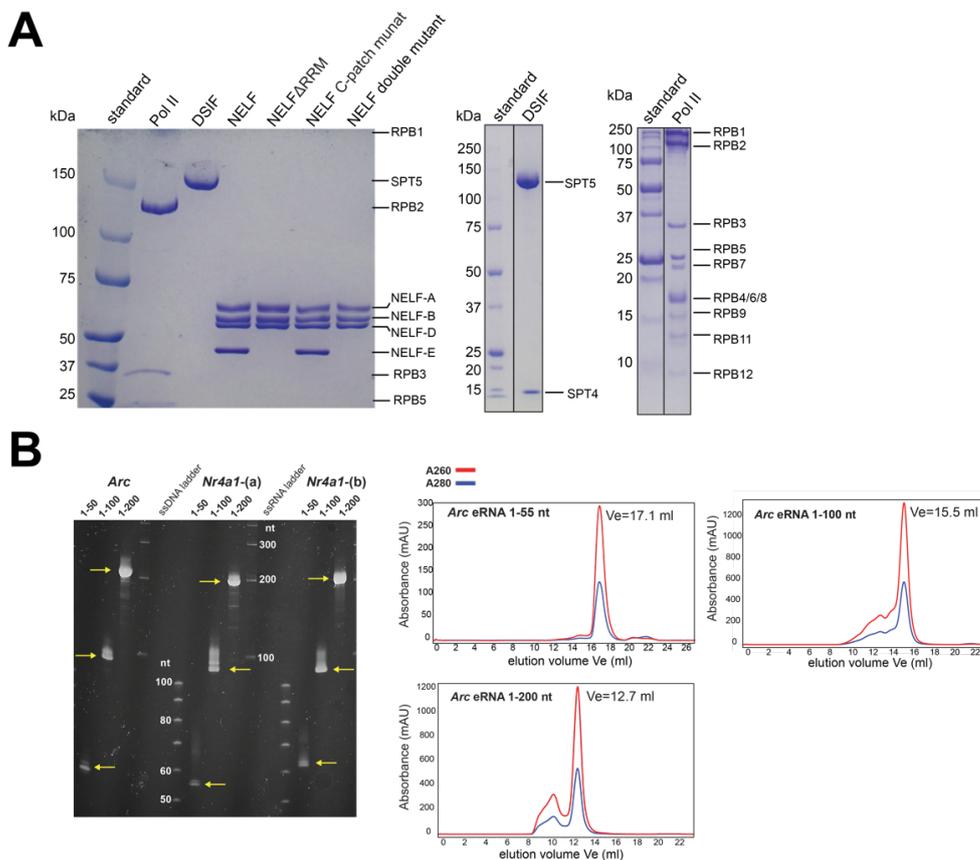
higher modification rates on nascent eRNA transcripts (*Run-on* sample). The library construction on the *Nuclei* and *Run-on* samples was performed by me using the small RNA workflow (Smola, Rice, et al., 2015) of SHAPE-MaP relying on eRNA-specific reverse transcription primers and PCR primers. For the pilot experiment, I selected four eRNAs (*Gadd45b*, *Fosb*, *Fos-e2*, *Nr4a1-(a)*) and aimed to produce two overlapping, ~250 nt long, fragments (1-250 nt and 150-400 nt) to cover the 5' terminal 400 nt of the eRNA. Furthermore, to sort out PCR duplicates after sequencing that can bias the results, I used PCR primers comprising a randomized sequence. After removing the duplicates, the obtained read depth was sufficient (supplementary Table 7-25) for accurate structure modeling. Yet, the mutation rates of the 1M7-treated sample were hardly exceeding the background mutation rates of the DMSO-control (Figure 3-9A left panel and Figure 3-9B upper panel) and exhibited a narrower distribution compared to the broader distribution of the 1M7 mutation rates for the *in vitro* SHAPE-MaP experiment (Figure 3-9A right panel). These data indicate that we were unable to achieve a sufficient modification of eRNAs *in vivo*. Furthermore, the additional treatment with 1M7 in the *Run-on* samples (2-times 1M7-treated) did not improve the mutation rates compared to the *Nuclei* sample (once 1M7-treated) (Figure 3-9A). As the mutation rates of the 1M7 samples were still slightly above the DMSO rates, the ShapeMapper 2 tool was able to calculate the SHAPE reactivities (Figure 3-9B bottom panel). However, the lack of distinctly high mutation rates over background renders these SHAPE reactivities unreliable.



**Figure 3-9 | *In vivo* SHAPE-MaP yields low mutation rates over background.** (A) The left panel shows the distribution of mutation rates for each tested eRNA 1-400 nt (*Nr4a1-(a)*, *Fos-e2*, *Fosb*, *Gadd45b*) under the three experimental conditions (DMSO-control, 1M7-modified-Nuclei, 1M7-modified-Run-on) (*Fosb* includes only mutation rates for the 1-250 nt fragment, as there were no reads for the second 150-400 nt fragment). The right panel shows the distribution of mutation rates for the same eRNAs 1-200 nt fragments for the *in vitro* performed SHAPE-MaP experiment. For the *in vivo* experiment the mutation rates for the 1M7 treated samples hardly outpace the background mutation rates of the DMSO-control, while in the *in-vitro* experiment the mutation rates of the 1M7 sample were clearly shifted upwards and showed a broader distribution than the DMSO-control. (B) The upper panel shows the mutation rate profile of *Nr4a1-(a)* 1-400 nt and the bottom panel the corresponding SHAPE reactivity profile.

### 3.3 Functional impact of eRNAs on the PEC *in vitro*

The structure probing of the eRNAs did not disclose any common structural motif. However, many regulatory RNA functions do not rely on particular sequence or structure motifs but promiscuous binding with low specificity. Hence, I turned to functional assays to test the impact of eRNAs on the paused elongation complex (PEC) *in vitro*. Two types of assays were established: an electrophoretic mobility shift assay (EMSA) and a transcription assay. The EMSA experiments were designed by me and established by me and Filiz Kuybu, who carried out all the final EMSA experiments shown below. The transcription assay was adapted from a previous publication (Vos et al., 2018) and initial experiments performed by Lisa-Marie Schneider. The prerequisite for both assays was the preparation of the PEC components (Pol II, DSIF and NELF) and the synthesis of eRNAs (Figure 3-10A).



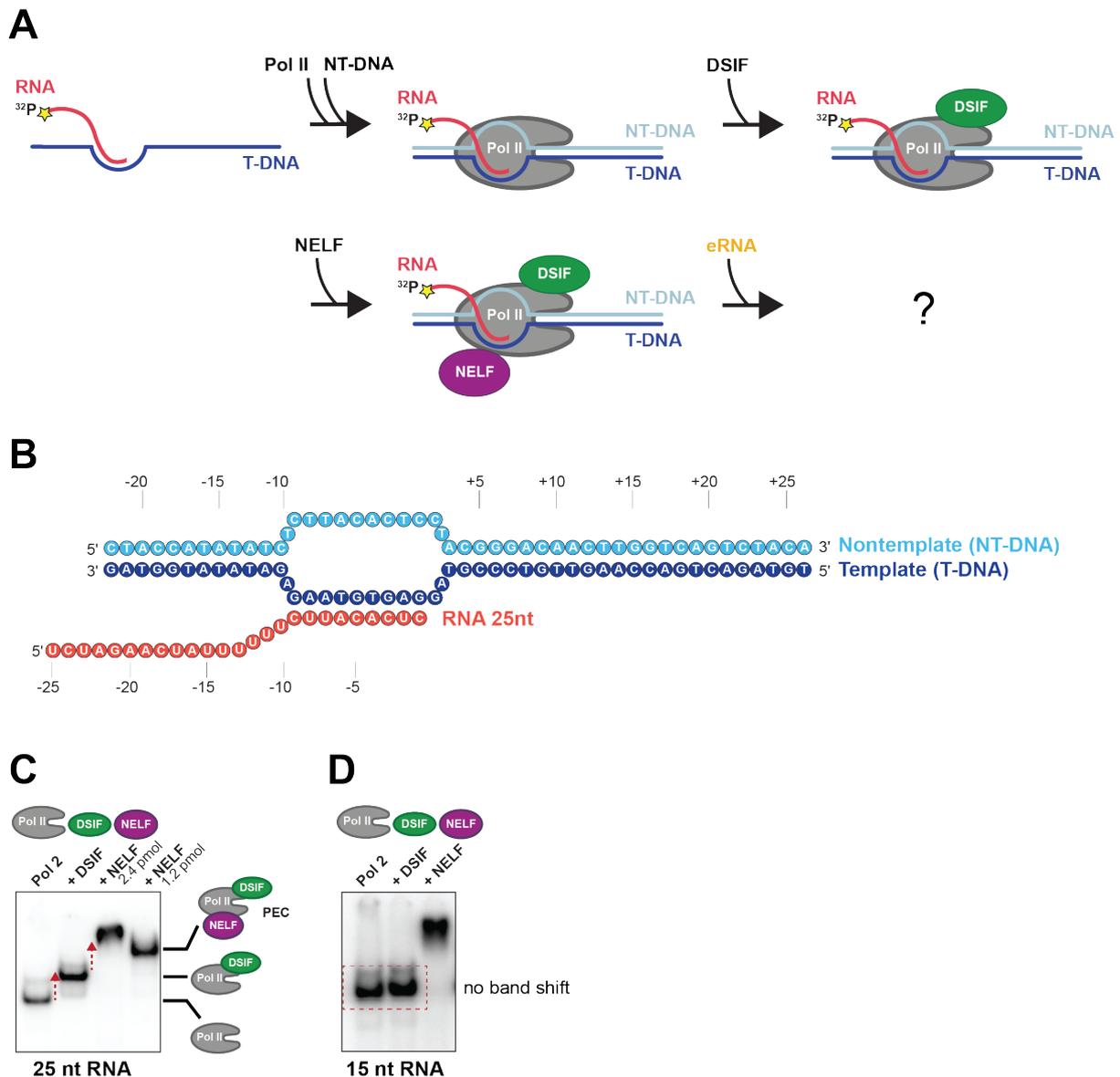
**Figure 3-10 | Purified protein and eRNA components required for *in vitro* assays.** (A) Shows an SDS-PAGE (left panel) and Bis-Tris PAGE gels (middle and right panel) of purified NELF variants, DSIF and Pol II. Middle gel (purified DSIF) was carried out by Dr. Felix Klatt, right gel (purified Pol II) was prepared by Robin Weinmann (master's thesis). (B) Shows to the left side an analytical urea PAGE gel of purified *Arc*, *Nr4a1*-(a) and -(b) fragments (1-50, 1-100 and 1-200; marked with arrows) stained with Sybr Gold. ssDNA (20/100 IDT ladder) and ssRNA ladder (RiboRuler low range, ThermoScientific). On the right side representative gel filtration chromatograms for the three *Arc* fragments are shown. Gel and chromatograms were recorded by Filiz Kuybu. (B) is adapted from Gorbovytska et al., 2021.

To this end, endogenous Pol II was isolated by me from porcine (*Sus scrofa*) thymus, based on a published protocol, which was slightly adapted and modified by Lisa-Marie Schneider and myself (see section 5.18). DSIF and NELF were overexpressed and purified from insect cell culture either by me, Dr. Felix Klatt, Silke Spudeit (technical assistant) or Lisa-Marie Schneider. The eRNAs were produced by *in vitro* transcription and purified using the same workflow as described for SHAPE-MaP (Figure 3-10B), mainly by Filiz Kuybu and myself.

### 3.3.1 Electrophoretic mobility shift assays

We turned to radioactive EMSAs as this technique offers high sensitivity for detecting protein-RNA interactions and allows to examine large macromolecular complexes, such as the PEC (which has a size of 0.9 MDa). The PEC was assembled in a stepwise manner on a synthetic transcription bubble comprising a 5'-<sup>32</sup>P-labelled "nascent" RNA with a length of 25 nucleotides (Figure 3-11A,B), similar to previously described protocols (Bernecky et al., 2017; Vos et al., 2018). Eventually, a specific eRNA was added with increasing amounts and the formed complexes were resolved on a native TBE-gel (EMSA gel) (Figure 3-11C). Early experiments in this study were performed with a shorter, 15 nucleotides long, "nascent" RNA as it was applied in a former publication in the context of transcription assays using the PEC (Vos et al., 2018). However, EMSAs with the 15 nt RNA did not produce a band shift upon the addition of DSIF (Figure 3-11D). Our finding was consistent with previous studies reporting that DSIF binding to Pol II requires a nascent RNA with a length of > 22 nucleotides (Missra & Gilmour, 2010). Formation of the complex on the 25 nt long RNA provided the desired band shifts to distinguish between Pol II only, Pol II-DSIF, and the Pol II-DSIF-NELF (PEC) complex (Figure 3-11C).

Initially, we used DSIF and NELF in 2-fold molar excess over Pol II to guarantee that all Pol II molecules form a PEC. However, we observed that the addition of this over-stoichiometric amount of NELF to the Pol II-DSIF complex leads to a supershift on the gel (Figure 3-11C; compare lane 3 and lane 4) indicating that a second NELF molecule might bind to the complex. It is unclear whether the behavior seen *in vitro* has any biological relevance or not. To better reflect the *in vivo* situation, where only one NELF molecule is found in the PEC, the following EMSAs performed with near stoichiometric amounts of Pol II and NELF.



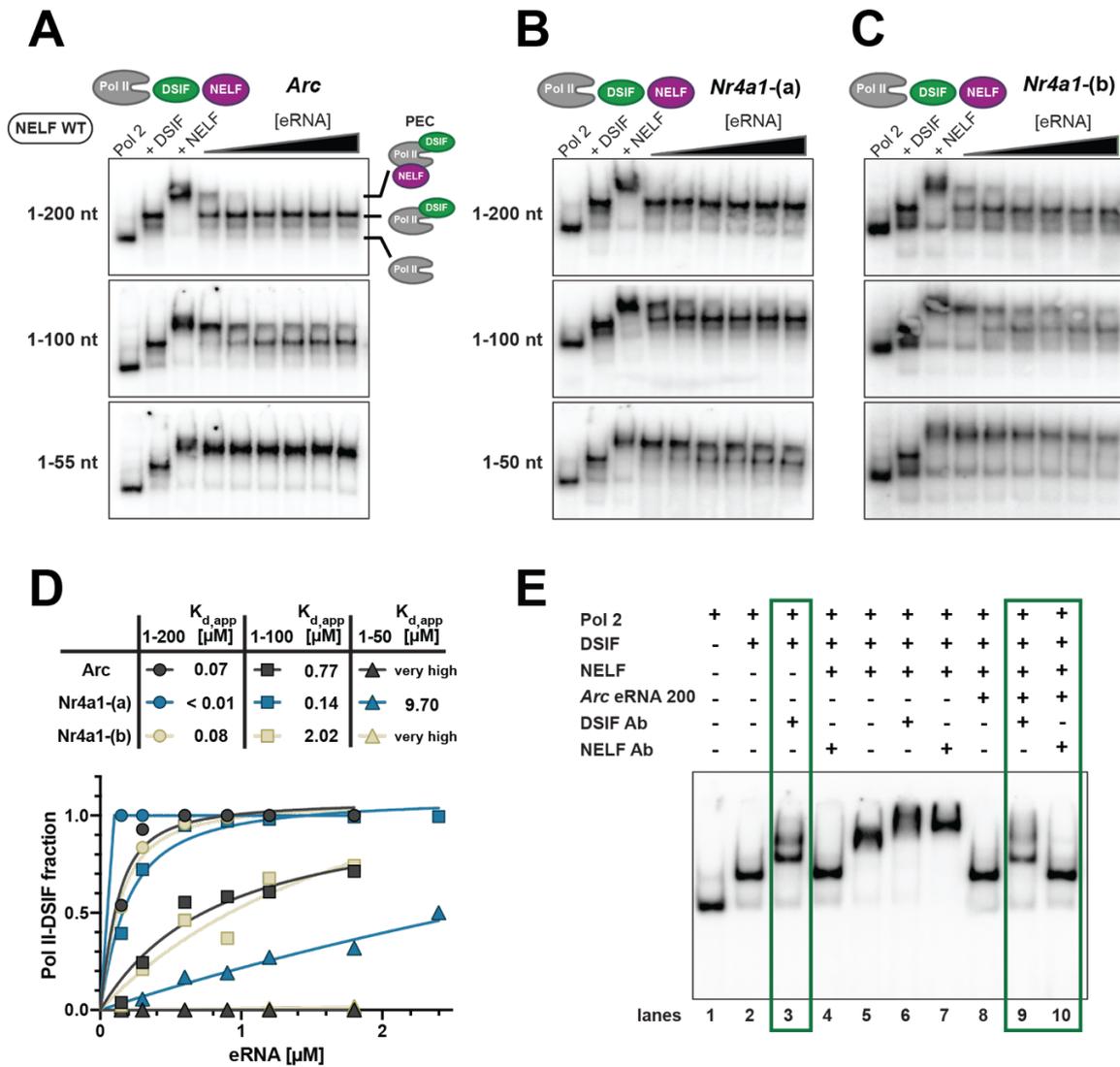
**Figure 3-11 | Workflow of the radioactive electrophoretic mobility shift assays (EMSAs).** (A) The scheme depicts the stepwise assembly of the paused elongation complex (Pol II-DSIF-NELF) on a nucleic acid scaffold (synthetic transcription bubble) containing a 5'-<sup>32</sup>P labelled "nascent" RNA; with the following amounts: RNA-DNA hybrid – 0.8 pmol; non-template DNA (NT-DNA) – 1.6 pmol; Pol II – 1.2 pmol; DSIF – 2.4 pmol; NELF – 1.2 pmol. (B) Nucleic acid scaffold used for the EMSA. The scaffold consists of a DNA template strand (49 nt), a fully complementary DNA non-template strand (49 nt) and a nascent RNA strand (25 nt). The 3' end of the RNA is complementary to the template DNA, forming a 9 bp long RNA-DNA hybrid. Annealing of RNA to the template DNA (T-DNA) is the first step in the workflow (A). The NT-DNA is added after the binding of Pol II to the hybrid, to prevent displacement of the RNA strand by the NT-DNA strand. (C) EMSA gel demonstrating the band shifts upon subsequent addition of DSIF and NELF. The band shift upon DSIF addition occurs only when the PEC assembly is performed with a 25 nt long RNA. The third lane shows the band shift, when NELF is added hyperstoichiometrically (2.4 pmol) to Pol II-DSIF in comparison to lane 4 (1.2 pmol NELF). The supershift in lane 3 indicates that two NELF molecules could bind to the complex, if NELF is added hyperstoichiometrically. (D) When using a 15 nt long RNA in the nucleic acid scaffold, no band shift was observed after addition of DSIF. The EMSAs shown were carried out by Filiz Kuybu. The subfigures (B) and (D) are adapted from Gorbovytska et al., 2021.

### 3.3.2 eRNAs can detach NELF from the PEC

Following the PEC assembly, 5'-terminal 200 nucleotide fragments of three selected eRNAs, *Arc*, *Nr4a1-(a)*, or *Nr4a1-(b)*, were added with increasing amounts and samples analyzed on native TBE-gel (Figure 3-12A-C top gels). These eRNAs were chosen as they belong to prominent neuronal immediate early genes, and *Arc* eRNA was previously shown to interact with NELF *in vivo* (Schaukowitch et al., 2014). Further, *Arc* and *Nr4a1-(b)* were shown to exhibit highly structured folds whereas *Nr4a1-(a)* was displaying a highly flexible structure (Figures 3-7E and 3-8C,D). Thereby, the selection covers both extremes of high and low structuredness. Intriguingly, we could observe the dissociation of NELF from the PEC upon adding approximately stoichiometric amounts of eRNAs, regardless of the type of eRNA. In contrast to NELF, DSIF remained bound to Pol II, even in the presence of a high excess of eRNA. This experimental outcome is in agreement with today's knowledge of the pause release *in vitro* and *in vivo*, where only NELF is leaving the complex, while DSIF is turned into a positive transcription factor and stays bound to Pol II throughout the elongation process (Fujinaga et al., 2004; Peterlin & Price, 2006; Yamada et al., 2006). Using a supershift assay, we could confirm that the complex band appearing after the addition of an eRNA at the same height as the Pol II-DSIF band (compare lanes 2 and 8 in Figure 3-12E) indeed contains only DSIF and no NELF (lane 9 and 10).

To narrow down the eRNA region that is responsible for the effect, we tested shorter 1-100 nt and 1-50 nt fragments with the same EMSA setup (middle and bottom gels in Figure 3-12A-C). Interestingly, with the shorter fragments, we observed a substantially diminished effect on the release of NELF compared to the effect of longer 1-200 fragments. The shortest fragments (1-50) exhibited no dissociative potential on NELF within the used concentration range (highest concentration equals a 12-fold molar excess over NELF). Though the 1-100 nt fragments retained some potential to dissociate NELF, the effect was less pronounced. The results indicated that the capability of eRNAs to dissociate NELF is profoundly dependent on their length. For an eased and more quantitative comparison of the EMSA gels, pseudo dissociation constant ( $K_d$ ) values were calculated.  $K_d$  values describe the position of the thermodynamic equilibrium of a complex formation between two components (e.g., ligand and receptor). In our case, we can indirectly measure the dissociation of NELF from the PEC as a function of the eRNA concentration, by evaluating the appearance of the Pol II-DSIF

complex in the EMSA gels (see section 5.2.3). The fraction of Pol II-DSIF plotted against the eRNA concentration allows to determine kind of  $K_d$  values, describing the ability of the eRNA to dissociate NELF from the PEC and reflecting the affinity of the eRNAs for the PEC bound NELF (Figure 3-12D).

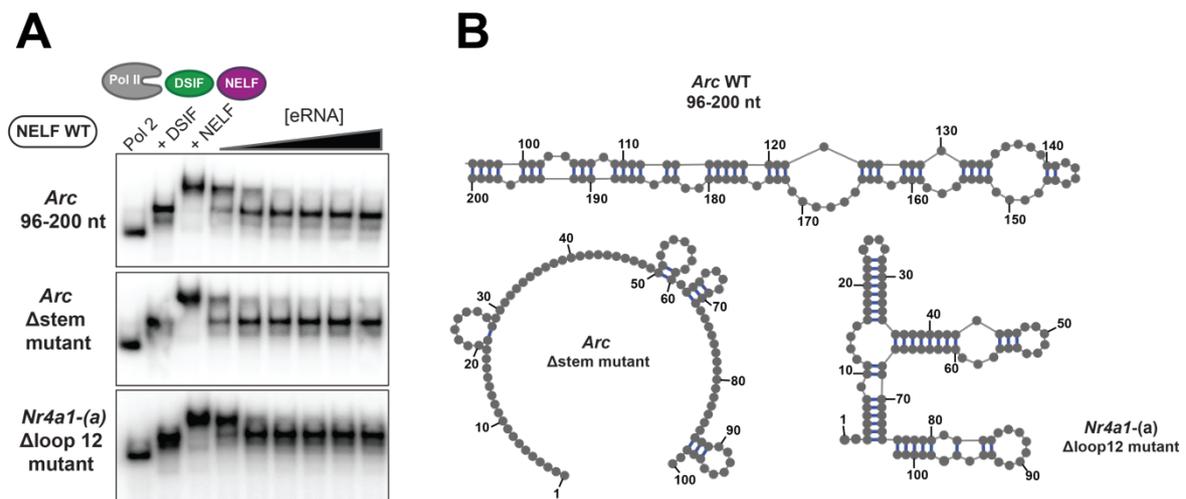


**Figure 3-12 | eRNAs are able to dissociate NELF from the paused elongation complex in a length-dependent manner.** (A) EMSA gels, demonstrating the release of NELF from the assembled paused elongation complex (PEC), triggered by three differently sized *Arc* eRNA variants: the top gel shows the effect of *Arc* eRNA (1-200 nt), the middle gel displays the effect of *Arc* eRNA (1-100) and the bottom gel examines the effect of *Arc* eRNA (1-55) on the PEC. (B) and (C) shows the same experimental setup as in (A), yet for *Nr4a1-(a)* and *Nr4a1-(b)* eRNA. The formation of the PEC was carried out outlined in Figure 3-12A. The concentration series of the eRNAs applied to the PEC (0.1 μM) was (0.15, 0.3, 0.6, 0.9, 1.2, 1.8, 2.4 μM). (D) shows the plot of the Pol II-DSIF fraction against the eRNA concentration corresponding to the EMSAs performed in (A-C). Curves were fit using a single-site binding model as described in Vos et al., 2016. The estimated apparent  $K_d$  values are listed above the plot. (E) Supershift assay with an anti-NELF and anti-DSIF antibody (Ab), validating the complex constitution of each band shift. Lane 9, and 10 confirm the presence of DSIF and the absence of NELF in the down shifted band after addition of an eRNA (here *Arc* eRNA 1-200 nt fragment was used). The EMSA experiments shown in (A), (B), (C), and (E) were carried out by Filiz Kuybu. The subfigures are a reprint from Gorbovytska et al., 2021.

As enhancer RNAs were shown to have an average length of ~1 kb (Schwalb et al., 2016), they certainly fulfill the observed strict length requirement *in vivo*. At the same time, although the tested enhancer RNAs exhibit quite different sequences and structures, there was no pronounced difference in their effects if comparing fragments of the same length. This argues for a rather promiscuous interaction between the eRNA and NELF, triggering its release. Furthermore, the length-dependency of the effect suggests the existence of multiple distant binding sites on the PEC. Our observations are reminiscent of the studies on the Polycomb repressive complex 2 (PRC2), in which PRC2 was shown to bind non-coding RNAs in a promiscuous and size-dependent manner (see discussion section) (Davidovich et al., 2013). Despite the apparent length-dependency of the effect, we noticed that the shorter fragments of *Nr4a1*-(a) (1-100 and 1-50) both exhibited a stronger effect than the cognate fragments of the other two eRNAs, *Arc* and *Nr4a1*-(b). Interestingly, while *Arc* and *Nr4a1*-(b) eRNA display a more double-stranded structure (Figure 3-7E and Figure 3-8A,D), *Nr4a1*-(a) is characterized by a more flexible structure with prominent single-stranded regions (Figure 3-8C). This suggests that RNA flexibility might facilitate the dissociative effect of eRNAs on NELF.

### 3.3.3 The eRNA secondary structure is not essential for triggering NELF release

Our initial EMSA results hint at the possibility that eRNA's secondary structure could play a role in their potency to dissociate NELF. To address this question, we designed and tested several eRNA mutants. As we knew that the effect of eRNAs highly depends on their length, we made sure to compare only equally-sized eRNA mutants. However, we did not detect a considerable difference in the effect of a nearly entirely double-stranded wild type (WT) fragment of *Arc* (96-200 nt) and the *Arc*  $\Delta$ stem mutant, which is free of any prominent secondary structure (Figure 3-13A,B). It should be noted that the *Arc*  $\Delta$ stem mutant differed a lot from the wild type *Arc* 1-100 sequence. To dissolve the stem structure of the 5' terminal hairpin loop (1-55 nt; see Figure 3-7E), many guanosines (Gs) were exchanged to adenosines (As), and some regions were deleted. Further, a mutant of *Nr4a1*-(a) ( $\Delta$ loop 12) was designed that was lacking prominent single-stranded regions, in contrast to its cognate WT eRNA. Again, no relevant difference of the effect between the wildtype and the mutant eRNA was observed (compare Figure 3-12B middle and 3-13A bottom panel).



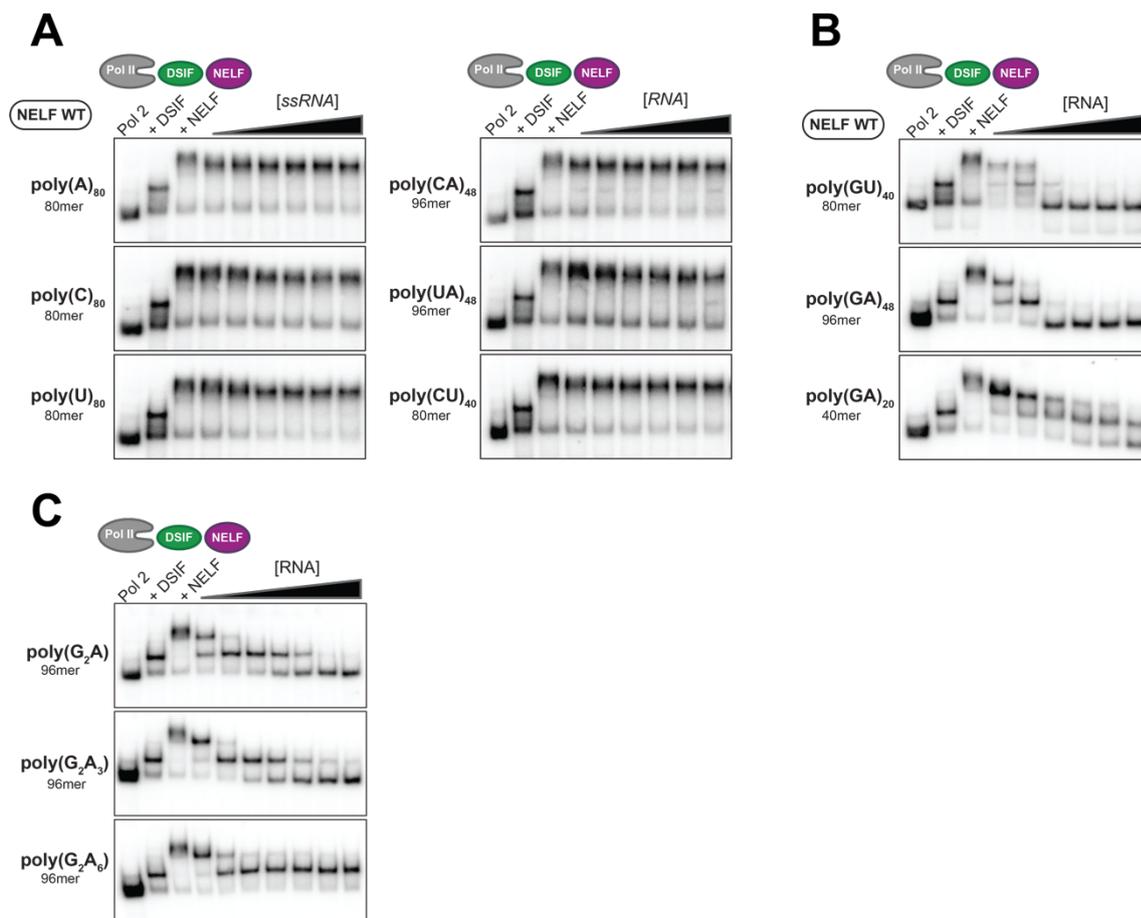
**Figure 3-13 | EMSAs with eRNAs mutants. (A)** EMSAs performed with the highly structured WT *Arc* eRNA (96-200) fragment (length = 104 nt), a single-stranded *Arc* eRNA  $\Delta$ stem mutant (length = 100 nt) and a *Nr4a1-(a)*  $\Delta$ loop 12 mutant (length = 102 nt), which is a mutant of *Nr4a1-(a)* lacking all prominent single stranded regions. EMSAs were carried out as described for (B-D). The EMSA experiment was carried out by Filiz Kuybu. The mutants were designed and cloned by me.

Taken together, these data suggest that there is no significant contribution of the secondary structure to the strength of the dissociative effect of the eRNA. This finding is consistent with our eRNA structure mapping results (section 3.2.3), demonstrating that eRNAs fold into a wide range of structures, from highly structured eRNAs like *Arc*, *Nr4a1*-variant (b), *Junb* and *Fos-e1*, to highly flexible eRNAs such as *Gadd45b*, *Fosb* or *Nr4a1-(a)* (Figure 3-8). However, we cannot exclude the possibility that some elements of the secondary structure, that influence the tertiary structure of the eRNA might impact the NELF-dissociating ability of the eRNAs.

### 3.3.4 Unpaired guanosines play a critical role in the dissociative effect of eRNAs on NELF

Our EMSA results revealed that the length of eRNAs plays a critical role in dissociating NELF, while there seems to be no contribution to the effect by their sequence or fold. To test whether nucleotide content might play a role and to double-check the impact of the secondary structure, we utilized 80-96 nt long, low complexity RNAs with repetitive sequences (e.g., poly (A), poly (C), poly (U), or poly (CA)) and measured their effect on NELF dissociation. Strikingly, all of the tested single-stranded RNAs that were lacking guanosines (Gs) were not able to dissociate NELF from the PEC (Figure 3-14A). In contrast, poly (GU) and

poly (GA) RNA showed a strong dissociative effect (Figure 3-14B). In particular, G-containing RNAs first dissociated NELF, however, at higher concentrations, they even led to the dissociation of DSIF. The effect of the G-containing RNAs was also length-dependent, as a shorter poly (GA) variant (40 nt), exhibited a weaker effect compared to the long variant (compare Figure 3-14B middle and bottom panel). Our EMSA results suggest that guanosines are critical to driving the dissociation of NELF. Yet, an excess of guanosines seems to trigger the undesired dissociation of DSIF. We thus asked which "density" of guanosines would be sufficient to provoke the release of NELF without affecting the binding of DSIF. To this end, we tested 96-mer poly (G<sub>2</sub>A) RNA, poly (G<sub>2</sub>A<sub>3</sub>) RNA, and poly (G<sub>2</sub>A<sub>6</sub>) RNA (Figure 3-14C). Remarkably, though all RNAs efficiently triggered the release of NELF, in contrast to the 96-mer poly (GA) RNA, their ability to dissociate DSIF was considerably diminished. The poly (G<sub>2</sub>A<sub>6</sub>) RNA even did not show any dissociative effect on DSIF, but only on NELF.



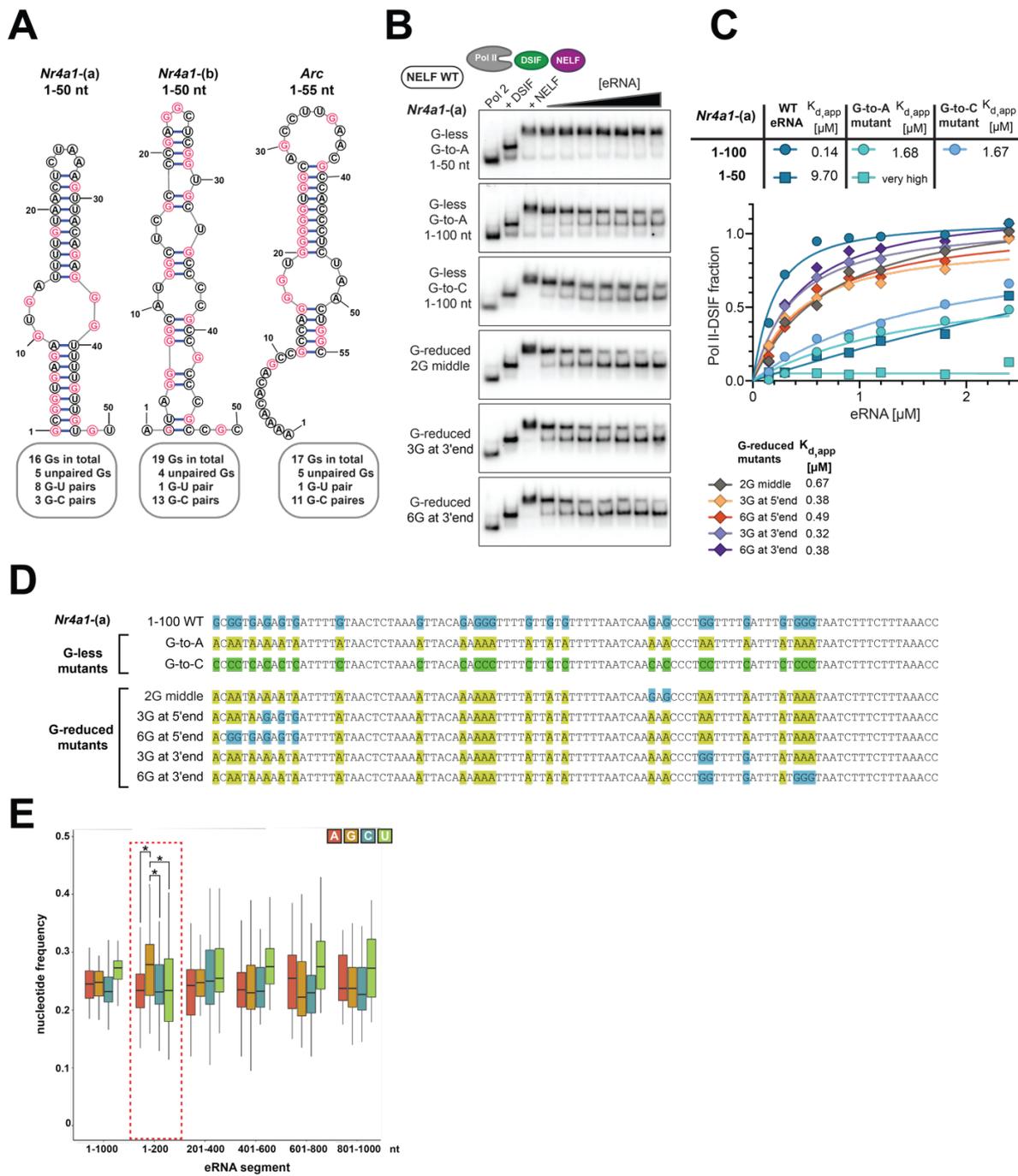
**Figure 3-14 | Guanosines play a critical role in the dissociation of NELF.** EMSAs performed with single-stranded, repetitive RNAs (80-96 nt) lacking (A) or comprising (B) guanosines (poly (GU) and poly (GA)). For the bottom gel in B a short 40 nt poly (GA) RNA was utilized. Only G-containing RNAs were potent to dissociate NELF but also DSIF from the PEC. (C) EMSAs performed with 96 nt long poly (G<sub>2</sub>A), (G<sub>2</sub>A<sub>3</sub>) and (G<sub>2</sub>A<sub>6</sub>) RNAs show that a reduced

"density" or content of Gs as in poly (G<sub>2</sub>A<sub>6</sub>) prevents the undesired dissociation of DSIF, while preserving a strong dissociative effect on NELF. All EMSAs shown in this figure were carried out by Filiz Kuybu under the same conditions as described in Figure 3-12. The subfigures are reprinted from Gorbovytska et al., 2021.

Taken together, our results suggest that unpaired guanosines contribute to the ability of an RNA to release NELF from the PEC. Strikingly, a high content of unpaired guanosines triggers in addition the dissociation of DSIF, while RNAs with a moderate G-content selectively displace only NELF. Based on the observation that the presence of guanosines is critical for NELF release, we could now rationalize our previous EMSA results (Figure 3-12A-C). *Nr4a1*-(a) 1-50 was the only short fragment exhibiting a mild dissociative effect on NELF. The sequence of *Nr4a1*-(a) 1-50 does not comprise more Gs, however, most of the guanosines are not paired to cytosines, as is the case for *Arc* and *Nr4a1*-(b) eRNAs (Figure 3-15). Instead, most of the guanosines in *Nr4a1*-(a) 1-50 are predicted to be unpaired or paired with uridines (Figure 3-15A). However, guanosines in G-U pairs are not comparable with guanosines in G-C pairs. G-U pairs have unique chemical features, such as the exocyclic amino group of guanine extending into the minor groove of the RNA and causing a strong negative electrostatic potential in the major groove (Varani & McClain, 2000). Furthermore, consecutive G-U base pairs inside RNA helices, as found in *Nr4a1*-(a) 1-50, can be destabilizing, whereas those at the ends of helices are thermodynamically stabilizing (Gu et al., 2015). Either way, the absence of stable double-stranded RNA structures, in which all Gs are paired with Cs, seems to favor the dissociation of NELF.

To verify that Gs are required for the NELF-dissociating effect of *Nr4a1*-(a), we tested a G-less mutant of *Nr4a1*-(a) 1-50 and 1-100, in which all guanosines were substituted with either adenosines (G-to-A) or cytosines (G-to-C) (Figure 3-15B-D). Indeed, the 1-50 nt G-less mutant did not trigger any NELF release. The 1-100 G-less mutants, G-to-A and G-to-C, both showed a substantially reduced ability to dissociate NELF, as reflected by the K<sub>d</sub> value that is about 10x higher compared to the *Nr4a1*-(a) WT fragment (Figure 3-15C). Notably, the 1-100 nt G-less mutants retained some potential to dissociate NELF and were not entirely dysfunctional. This observation indicates that guanosines strongly facilitate the effect of eRNAs, yet they are probably not obligatory to achieve a dissociative effect if the eRNA is long enough. Interestingly, the restoration (G-reduced mutants) of two guanosines in the middle (2G middle) or three guanosines at the 5'- or at the 3'-end of the G-less G-to-A mutant (3G at

5'-end or 3G at 3'-end) re-enhanced eRNA-driven NELF detachment 2.5 to 5-fold, respectively ( $K_d = 0.38 \mu\text{M}$  for 3G at the 5'-end and  $K_d = 0.32 \mu\text{M}$  for 3G at the 3'-end) (Figure 3-15B-D).



**Figure 3-15 | G-less mutants of *Nr4a1*-(a) show a reduced dissociative effect on NELF. (A)** Secondary structures of the *Nr4a1*-(a), -(b) 1-50 and *Arc* 1-55 nt fragment predicted with RNAstructure. The predicted structure of *Arc* eRNA (1-55) is consistent with the fold determined by SHAPE-MaP for the *Arc* 1-200 fragment (Figure 3-7E). All Gs are highlighted in pink. Despite a similar G-content between the fragments, the majority of guanosines in *Nr4a1*-(b) and *Arc* are paired with cytosines, while in *Nr4a1*-(a) Gs are located in loop regions or G-U pairs. **(B)** EMSAs were performed with G-less *Nr4a1*-(a) 1-50, 1-100 mutants, in which all Gs were substituted with As (G-to-A mutant) or Cs (G-to-C mutant) respectively, and with *Nr4a1*-(a) 1-100 G-reduced mutants, in which two, three or six of the original Gs were restored in the sequence. The corresponding sequences of *Nr4a1*-(a) 1-100

wild type (WT) and mutants tested in (B) are displayed in (D) together with two additionally tested mutants (EMSA not shown). The restored Gs in the G-reduced mutants are highlighted in blue. (C) shows the plot of the Pol II-DSIF fraction against the eRNA concentration corresponding to the EMSAs performed with the G-less and G-reduced mutants, and WT *Nr4a1*-(a) fragments (from Figure 3-12B) for comparison. Curves were fit as described in Figure 3-12B. Apparent  $K_d$  values estimated from the fit are listed. (E) Nucleotide frequency plot for all 39 experimentally verified eRNAs (see supplemental Table 7-21 for sequences). eRNA sequences were extended to 1 kb and divided into bins of 200 nt. Guanosines are only in the 5'-terminal 200 nt significantly overrepresented in comparison to all other three nucleotides (with p-values (A/G) = 0.025; (C/G) = 0.029; (U/G) = 0.029) as determined by a pairwise t-test (the plot shown was generated by Andreas Pittroff). The shown EMSAs were performed by Filiz Kuybu using the same conditions as described for Figure 3-12. The subfigures (A)-(C) and (E) are reprinted and partially adapted from Gorbovytska et al., 2021.

However, the restoration of additional three guanosines at the RNA's 5'- or 3'-end (6 guanosines in total) did not boost NELF dissociation any further (Figure 3-15C). The apparent  $K_d$  for NELF release by any of the tested G-reduced mutants is higher compared to the WT *Nr4a1*-(a) (1-100) eRNA (0.14  $\mu$ M, Figure 3-15C-D). Hence, our data argue that highly potent eRNAs require several widely spaced, unpaired guanosines in their sequence, whereas a single cluster of guanosines is less effective in dissociating NELF (Gorbovytska et al., 2021). In support of our *in vitro* data, we found that Gs are significantly overrepresented (p-value < 0.05) within the first 200 nucleotides of our 39 selected mouse eRNAs (Figure 3-15E). Remarkably, 5' UTRs of genes are highly structured and hence characterized by a high GC-content, meaning that both Gs and Cs are equally overrepresented (Leppek et al., 2018). In contrast, our eRNA set displays an elevated nucleotide frequency solely for the guanosines (28%). The frequency of cytosines and the other two nucleotides (A and U) constitutes just 24% (Figure 3-15E). This statistic implies an excess of unpaired Gs at the 5' terminal part of the eRNAs that could contribute to a more efficient release of NELF.

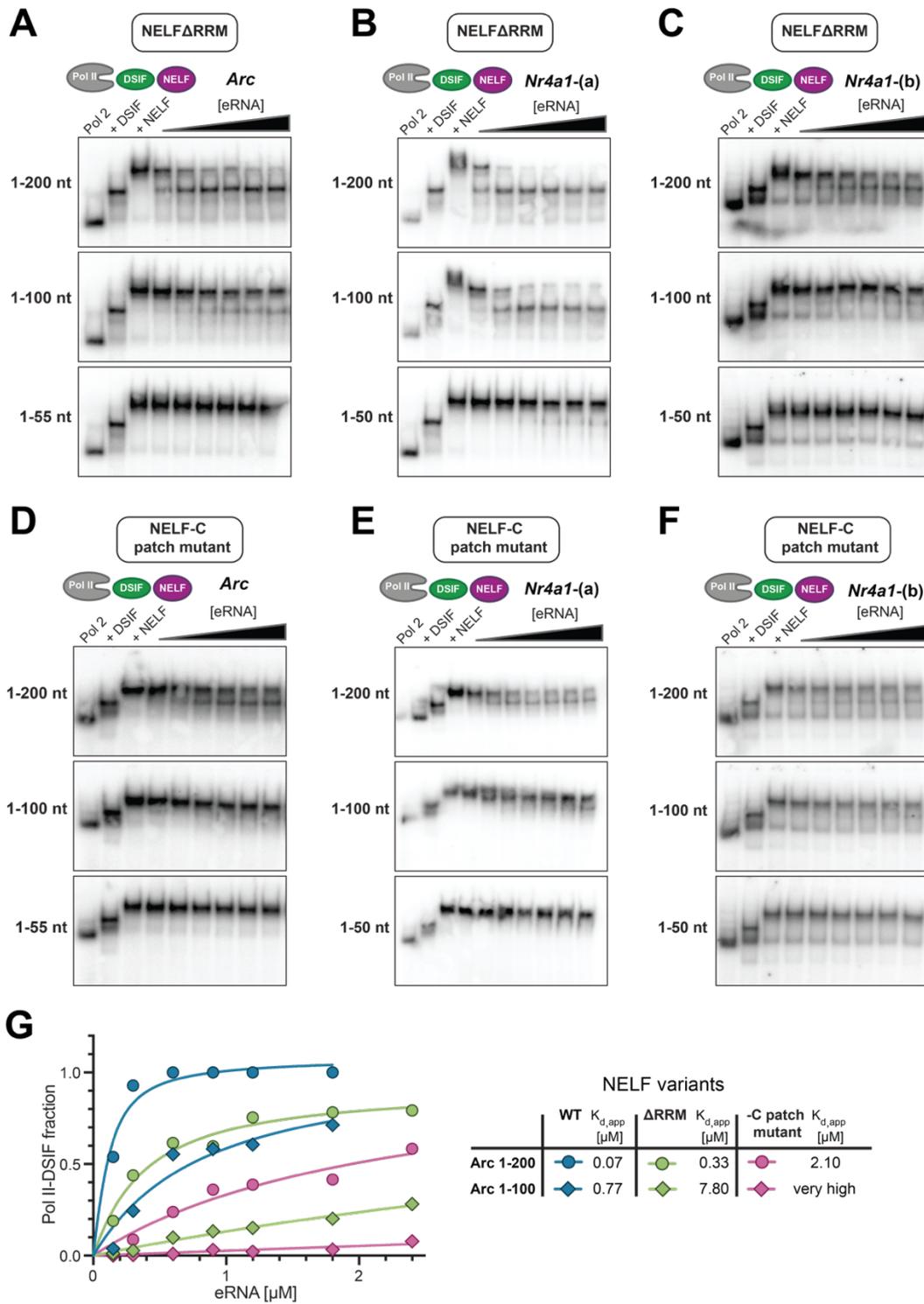
### 3.3.5 eRNAs bind to a positive patch on NELF-C and to the NELF-E RRM domain

So far, our findings allowed us to deduce that eRNA length and the presence of guanosines are critical for the eRNA's function to dissociate NELF from the paused elongation complex *in vitro*. We next sought to locate the part or parts of the paused elongation complex that are causal for its susceptibility to the dissociative action of eRNAs. As the enhancer RNAs trigger the dissociation of NELF, they are likely directly contact NELF. NELF is a complex consisting of four proteins, NELF-A, -B, -C, and -E (Yamaguchi et al., 1999). The subunit NELF-E comprises an RRM (RNA recognition motif) domain, which was shown to bind single-stranded RNAs and

loop regions in structured RNAs, such as the HIV TAR element, *in vitro* (Pagano et al., 2014; Jampani Nageswara Rao et al., 2008; Yamaguchi et al., 2002). Furthermore, it was previously shown that the RRM domain is essential for eRNA function *in vivo* (Schaukowitch et al., 2014). Beyond its canonical RNA binding domain, NELF was found to feature additional nucleic acid binding sites on the NELF-AC lobe and NELF-B (Vos et al., 2016). The NELF-AC lobe displays four positively charged patches (that comprise lysines and arginines) on its surface that are likely mediating the interaction with nucleic acids. While the RRM domain is thought to be involved in binding the nascent RNA during promoter-proximal pausing (Yamaguchi et al., 2002), the biological role of the additional nucleic acid binding interfaces has not been established. Given the background of the potential RNA binding sites on NELF, we purified and tested first a NELF variant (NELF $\Delta$ RRM) lacking the RRM domain in our standard EMSA setup. Second, we purified and tested a NELF variant (NELF-C patch mutant), in which all residues of the positively charged patches on the NELF-C subunit were mutated to methionine to dispose of the positive charges. Third, we purified and tested a double mutant (NELF double mutant), a combination of the two mutants above. The plasmids for these variants were a kind gift from Prof. Dr. Patrick Cramer (Max-Planck-Institute for Biophysical Chemistry, Göttingen) and have already been used in a previous publication (Vos et al., 2016). Both patch mutant-containing variants were expressed and purified by Felix Klatt. The NELF $\Delta$ RRM variant was expressed and purified either by Lisa-Marie Schneider or me. It is of note that the wild type NELF (NELF WT), as well as the three variants, all contained the NELF-D isoform that is lacking the first nine amino acids (aa) of NELF-C (1-9 aa) (see section 1.2). However, to be consistent with the previous publication (Vos et al., 2016), the NELF-D residues were numbered according to the NELF-C nomenclature, which is relevant for the next results section 3.4.

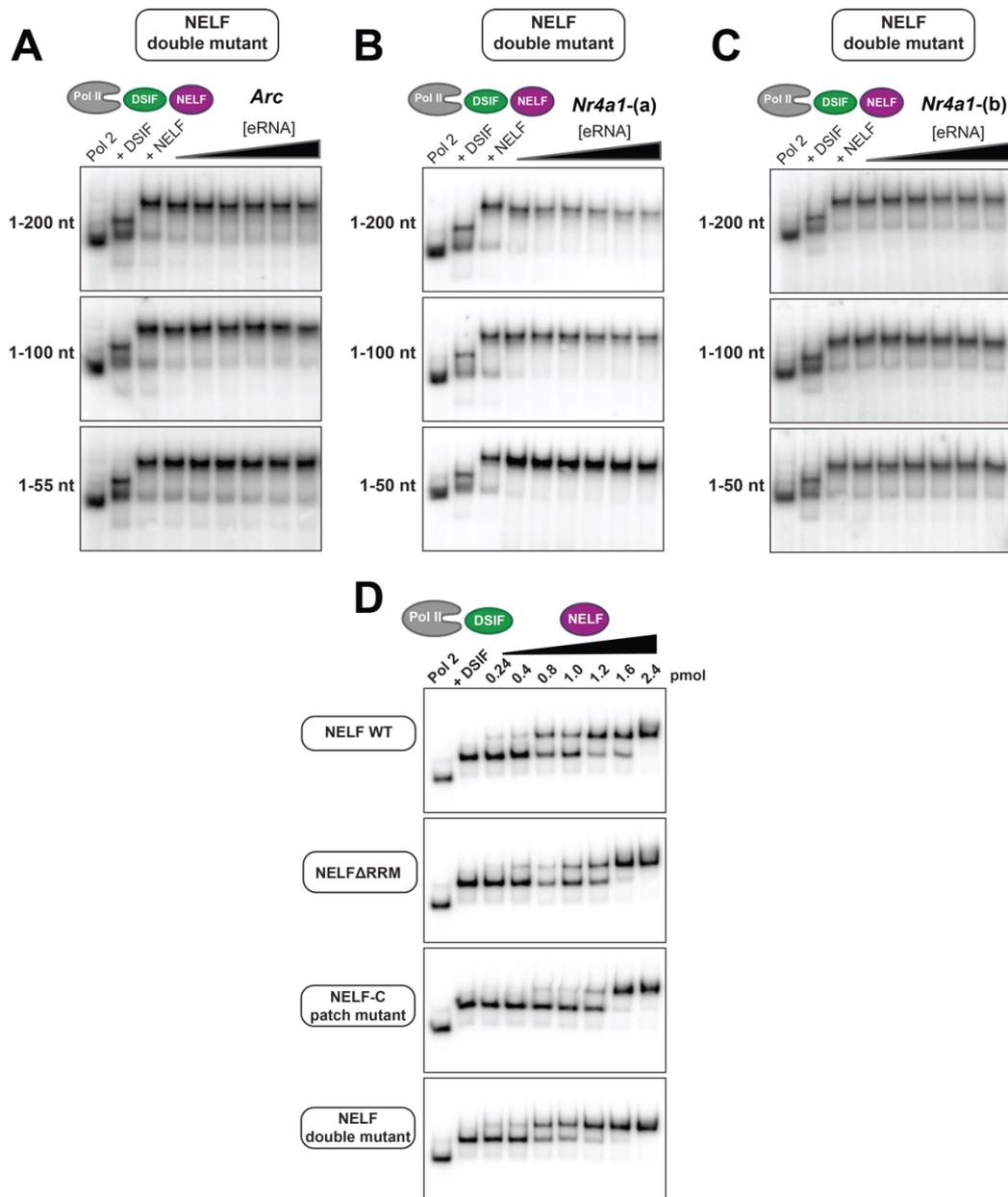
In examining the effect of eRNAs on the NELF $\Delta$ RRM mutant, we observed that all eRNA fragments triggered significantly less dissociation of the mutant NELF from the PEC (Figure 3-16A-C) than in the case WT NELF (Figure 3-12A-C). The estimated apparent  $K_d$  values for the 1-200 and 1-100 fragment of *Arc* (0.33  $\mu$ M and 7.8  $\mu$ M) were about 5- to 10-times higher than the  $K_d$  for the same fragments in the assays with NELF WT (0.07  $\mu$ M and 0.77  $\mu$ M) (Figure 3-16G). These data evidence that the NELF-E RRM domain is involved in the process of eRNA-induced NELF dissociation. However, deletion of the RRM domain did not entirely abolish the release of NELF, which was especially visible for the *Nr4a1*-(a) 1-200 and 1-100

fragments (Figure 3-16B). It implied that the eRNA-driven release of NELF probably depends not solely on the RRM domain. This is in line with the observed dependency of NELF dissociation on eRNA length. Interestingly, when we tested the NELF-C patch mutant, we observed an even more substantial reduction of NELF dissociation than for the NELF $\Delta$ RRM mutant (Figure 3-16D-F). More specifically, *Arc* and both *Nr4a1* eRNA (1-100) fragments hardly triggered NELF detachment, whereas the 1-200 nt fragments did, but to a much lower extent as compared to the NELF $\Delta$ RRM variant (Figure 3-16G).



**Figure 3-16 | Charged patches on NELF-AC and the NELF-E RRM domain are both involved in eRNA-driven NELF dissociation.** EMSAs were carried out with fragments of *Arc*, *Nr4a1-(a)* and *-(b)* eRNA, and a NELF mutant lacking the RRM-domain (NELF $\Delta$ RRM) (A-C) or a NELF mutant, in which the positively charged patches on NELF-C were removed (NELF-C patch mutant) (D-F). (G) shows the plot of the Pol II-DSIF fraction against the eRNA concentration corresponding to the EMSAs performed with *Arc* eRNA 1-200 and 1-100 (top and middle panel of A and D) and the EMSAs performed with the same *Arc* fragments and WT NELF (Figure 3-12A and D), to aid comparison. Curves were fit as described in Figure 3-12B and the estimated apparent  $K_d$  values are listed. The shown EMSAs were performed by Filiz Kuybu using the same conditions as described for Figure 3-12. The figure is adapted from Gorbovytska et al., 2021.

Last, when we used the NELF double mutant that combines two previous variants, we were unable to detect any dissociative effect of the eRNAs on NELF (Figure 3-17A-C). The double mutant has entirely lost its ability to dissociate from the PEC, even if eRNA (1-200) fragments were added in large molar excess.

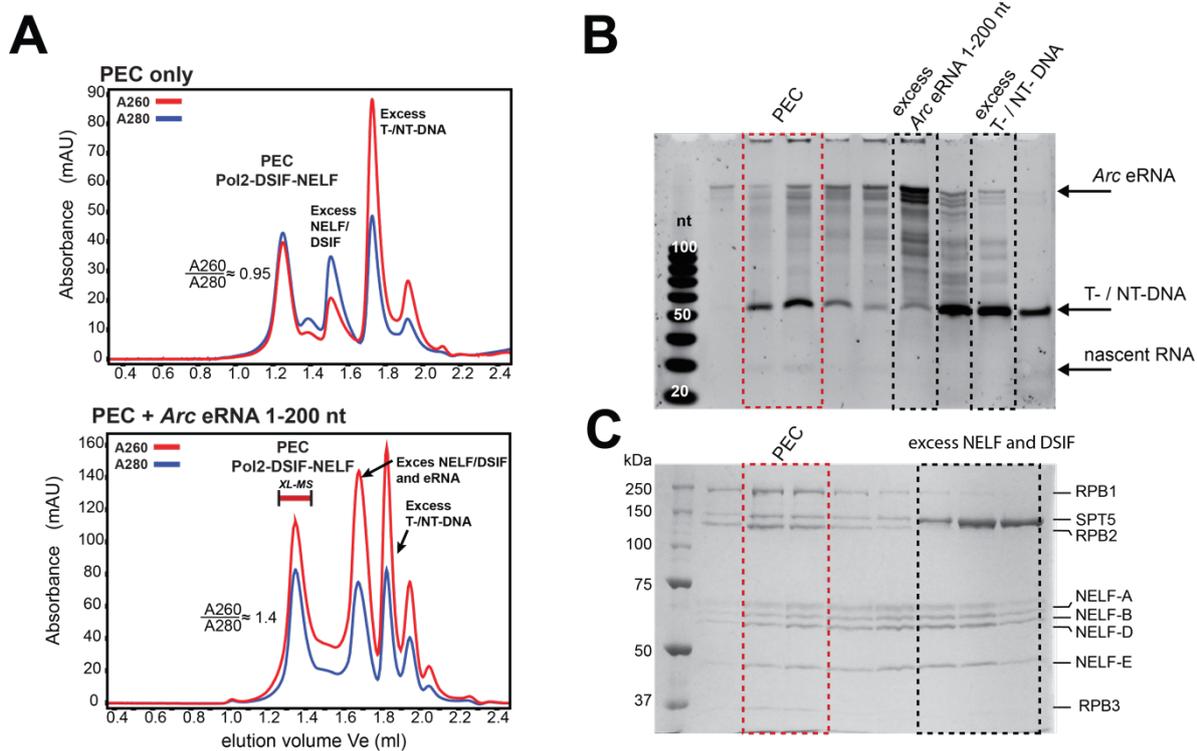


**Figure 3-17 | The NELF double mutant cannot be displaced from the PEC by eRNAs.** EMSAs were carried out with fragments of *Arc*, *Nr4a1-(a)* and *-(b)* eRNA, and the NELF double mutant lacking the RRM-domain and the positively charged patches (A-C). (D) EMSA titration experiment in which WT NELF and the three NELF mutants (NELF $\Delta$ RRM, NELF-C patch mutant, and the double mutant) were added to preformed Pol II-DSIF complexes. The gels demonstrate that WT and mutant NELF variants form the PEC comparably well. The shown EMSAs were performed by Filiz Kuybu. Adapted from Gorbovytska et al., 2021.

We conclude from the obtained results that both the NELF-E RRM domain and the positively charged surface patches on the NELF AC-lobe are essential to enable eRNA-induced NELF dissociation from the PEC (Figure 3-15, 3-16 and 3-17). Of note here is that all mutant PEC complexes could be assembled as efficiently as wild type complexes (Figure 3-17D), a finding that allows the aforementioned interpretation of our data (Gorbovytska et al., 2021).

### 3.4 Protein-RNA UV-crosslinking mass spectrometry

To verify the hypothesis that eRNAs bind simultaneously to the charged patches on NELF-AC and the NELF-E RRM domain to trigger NELF release from the PEC, I applied UV-induced protein-RNA crosslinking followed by mass spectrometry. To this end, I assembled preparative amounts of the PEC on the same nucleic acid scaffold as used for the EMSA experiments (Figure 3-11A-C), except that the nascent RNA was not radioactively labeled. After addition of *Arc* eRNA (1-200) or *Nr4a1*-(b) eRNA (1-100), the eRNA-bound PEC was separated from non-incorporated components by size-exclusion chromatography. The presence of bound eRNA in our PEC preparation was attested by both urea PAGE analysis and by an increased A260/A280 absorbance ratio as compared to the regular PEC assembly without eRNA (Figure 3-18A,B). Under the experimental conditions, only a fraction of the PEC is bound to an eRNA, as reflected by the relatively small increase of A260 absorbance after the addition of *Arc* eRNA (1-200), which has a size of 66 kDa (33 kDa for *Nr4a1*-(b) (1-100)). This low binding efficiency is likely due to our ability to isolate transient eRNA-PEC complexes that, at higher eRNA concentrations, would trigger the complete dissociation of the PEC, as demonstrated in the EMSAs (Figure 3-12). Fractions containing the eRNA-bound PEC were pooled, concentrated, and were sent to Dr. Alexander Leitner and Dr. Michael Götze (Institute of Molecular Systems Biology, ETH Zurich, Switzerland), who carried out the UV crosslinking and mass spectrometry experiment, comprising data acquisition and analysis. After UV crosslinking and before the mass spectrometry measurement, the complexes were treated with RNases and trypsin for limited digestion of RNA and proteins. We were able to retrieve 83 unique protein-RNA crosslink positions on the *Arc* eRNA (1-200)-bound PEC and 237 unique crosslink positions for the *Nr4a1*-(b) eRNA (1-100) variant after filtering with an xQuest score >25 for reliable crosslinks (supplementary Tables 7-26 and 7-27) (Rinner et al., 2008).

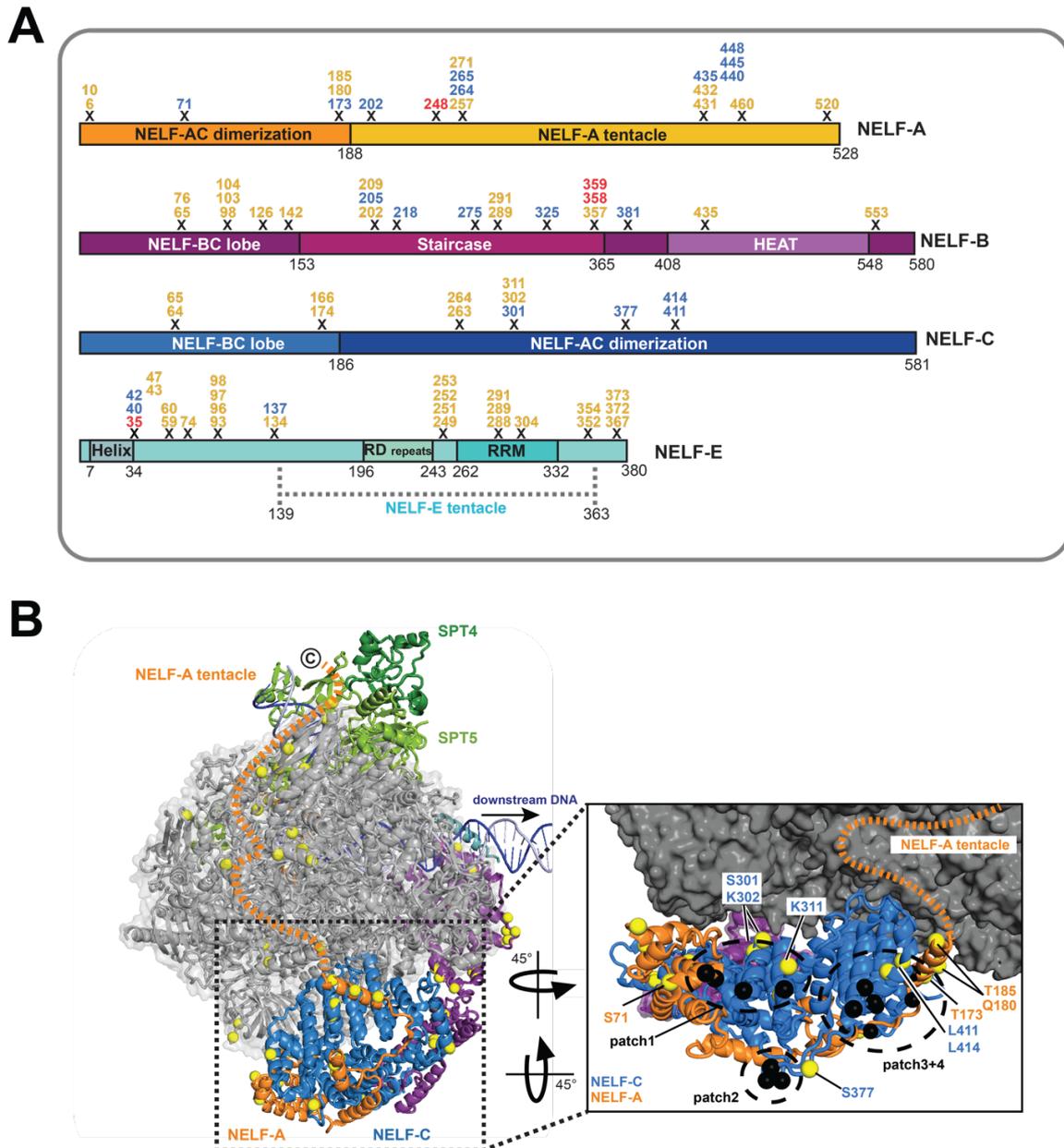


**Figure 3-18 | Gelfiltration of eRNA-bound PEC complexes used for Protein-RNA crosslinking.** (A) Elution profiles of the PEC assembly without eRNA (PEC only; upper chromatogram) and with *Arc* eRNA (1-200) (bottom chromatogram) on an analytical Superose 6 3.2/300 Increase column. Fractions from a preparative Superose 6 column corresponding to PEC peak (marked XL-MS) were used for UV crosslinking coupled to MS/MS experiments to determine eRNA binding sites on the PEC. An increased ratio of A260/A280 absorbance observed in the PEC+eRNA chromatogram compared to the PEC only control indicates eRNA binding to the PEC. (B) Analytical urea-PAGE gel of the size exclusion chromatography fractions of the experiment in (A) using *Arc* eRNA (1-200). Fractions containing the PEC display bands of the T-DNA, NT-DNA, nascent RNA and *Arc* eRNA. (C) SDS-PAGE (8% Tris-glycine gel stained with Coomassie brilliant blue) of the size exclusion chromatography fractions. Fractions containing the PEC display protein bands of Pol II, DSIF and NELF. Adapted from Gorbovytska et al., 2021.

Unfortunately, the experiment did not allow us to reliably detect >2 nt attached to any protein residue, which precluded the unambiguous assignment of RNA sequence to the protein surface. Nevertheless, the vast majority of protein-RNA crosslinks, which were characterized by high scores, were detected mainly for uridines and residues of the RPB1 dock domain of Pol II, the RPB2 wall, switch 3 loop region and the SPT5 KOWx-4 and KOW4-5 linker region (Figure 3-19A,B) (Gorbovytska et al., 2021). These regions are known to contact nascent RNA within the Pol II-DSIF and the PEC structures, as displayed in Figure 3-19A. Thus, we can unambiguously attribute those crosslinks to a stretch of uridines on the nascent RNA (position -10 through -17; Figure 3-19A). Thereby we can confirm the integrity of the PEC preparation (Bernecky et al., 2017; Vos et al., 2018; P. K. Zuber, 2018).



experiment, using only the PEC without an eRNA, would need to be performed to make a definite statement about this.



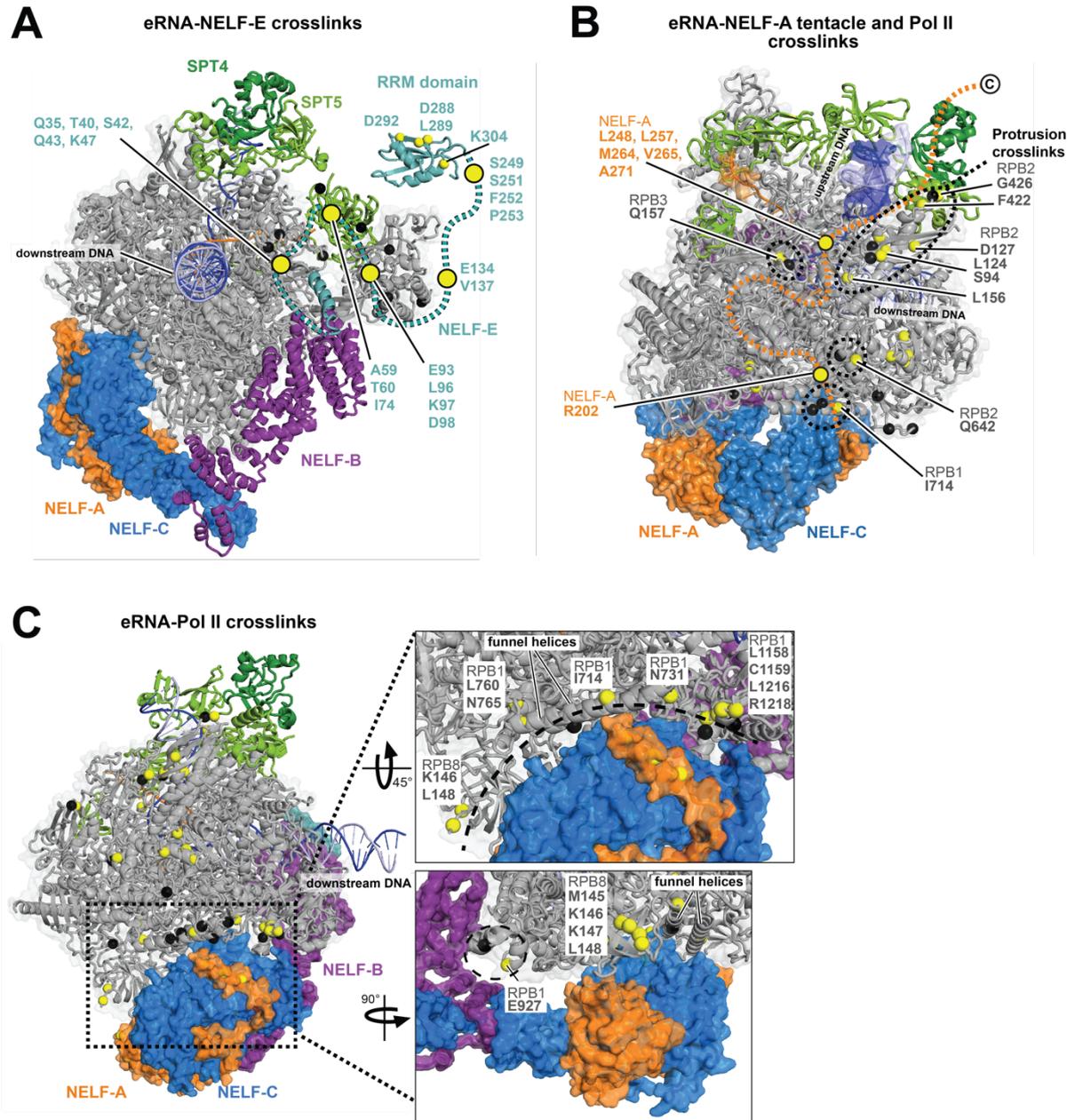
**Figure 3-20 | RNA crosslinks to NELF.** (A) RNA crosslinks to NELF (with xQuest score > 25) are shown on the linear domain organization of individual NELF subunits (adapted from Vos et al., 2018). The used crosslink color code is as described in Figure 3-19B. (C) Visualization of RNA crosslinks on NELF-AC, NELF-B and on the Pol II surface of the PEC (PDB code 6GML). Crosslinked residues (from B) are displayed as yellow spheres. The close-up view on the right, details RNA crosslinks on the NELF-AC lobe (yellow spheres). Positively charged patches (1, 2, 3+4) are encircled and the corresponding residues on NELF-C are marked as black spheres, according to Vos et al., 2016. Adapted from Gorbovytska et al., 2021.

When focusing on the NELF-AC lobe, we found numerous polar residues on NELF-A (S71, T173, Q180, T185) and NELF-C (S301, K302, K311, S377, L411, L414) to be crosslinked to RNA. All these residues are located in vicinity of the positively charged patches 1, 2 and 4 on the

surface of the NELF-AC dimer (detailed view in Figure 3-20B). These patches were earlier demonstrated to bind RNA in the context of purified NELF (Vos et al., 2016), and we showed them to be critical for eRNA-driven NELF dissociation from the PEC in our EMSA experiments (Figure 3-16D-F and Figure 3-17A-C). Furthermore, we observed several RNA crosslinks to the unstructured N-terminal domain of NELF-E (Q35, T40, S42, Q43, K47, A59, T60, I74, E93, L96, K97, D98, E134, V137), as well as crosslinks to the NELF-E tentacle just before its RRM domain (S249 – P253) and within the RRM domain (D288, L289, D292, K304) (Figure 3-21A and Figure 3-20A). Interestingly, we noticed that many of these RNA-protein crosslinks lie close to protein-protein crosslinks that had been previously reported between NELF-E and Pol II or DSIF (Vos et al., 2018). When further analyzing the RNA crosslinks to NELF, we found many crosslinks located within the unstructured NELF-A tentacle (residues R202, L248, L257, V265, A271, L431, L432, R435, A440, F445, A448, L460). The NELF-A tentacle is a highly flexible and to date unresolved region of NELF-A that was shown to be required for binding to Pol II and pause stabilization (Figure 3-21B and Figure 3-20A) (Narita et al., 2003; Vos et al., 2018). Like in the case for NELF-E, a closer inspection of the listed crosslink positions revealed a spatial correlation with NELF-A tentacle residues (K200, K207, K215, K219, K255 and K276), for which protein-protein crosslinking data had established the location of the NELF-A tentacle along the Pol II surface in the PEC context (Vos et al., 2018) (Figure 3-21B). In addition to the RNA crosslinks with NELF subunits, we detected multiple RNA crosslinks to subunits of Pol II, that are surface exposed and cannot be attributed to the nascent RNA. However, majority of these crosslinks could be rationalized, *e.g.*, crosslinks to RPB1 (I714), the protrusion domain of RPB2 (S94, L124, D127, L156, F422, G426), and RPB3 (D141, Q157) (Figure 3-19B and Figure 3-21B). These residues also coincide with the path of previously published protein-protein crosslinks, which mapped the NELF-A tentacle course on the surface of the PEC (Figure 3-21B) (Vos et al., 2018). Lastly, our data contained further RNA crosslinks to RPB1 (I714, N731, L760, N765, E927, L1158, C1159, L1216, R1218) and RPB8 (M145, K146, K147, L148, F150) that map to the interface of the NELF-AC lobe and Pol II (Figure 3-21C) (Vos et al., 2018).

Taken together, we detected RNA crosslinks not only to the charged NELF-AC patches and NELF-E RRM domain, but also to both the flexible NELF-A and NELF-E tentacle as well as the interface of Pol II and NELF (Figure 3-20A, Figure 3-21A,B; supplementary Tables 7-26 and 7-27). These crosslink data support and rationalize our previous EMSA results, which strongly

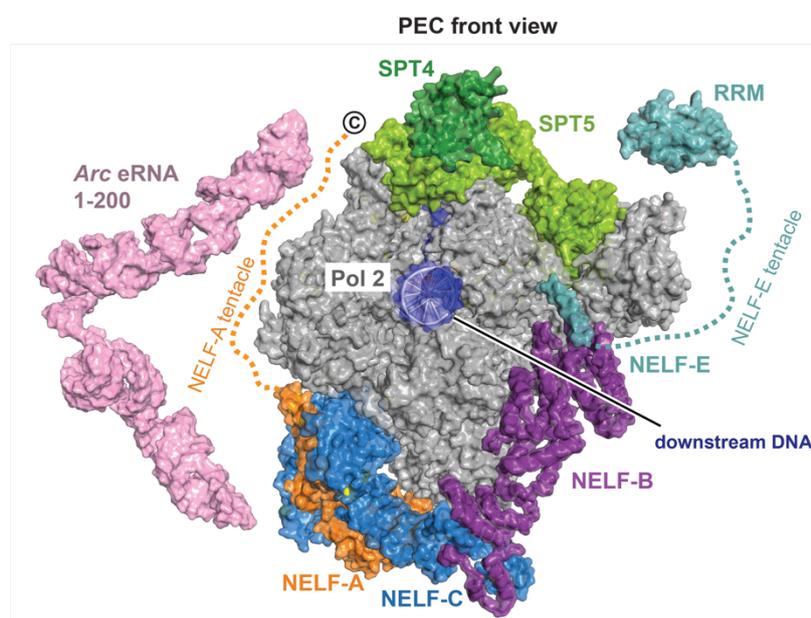
indicated that a single, long eRNA molecule has to bind to several RNA binding sites on NELF simultaneously to trigger the release of NELF from the PEC (Gorbovytska et al., 2021).



**Figure 3-21 | eRNA crosslinks to the NELF-E and NELF-A tentacle as well as to the interface Pol II-NELF-AC interface. (A)** RNA crosslinks to the NELF-E RRM domain (PDB code 2JX2) and to the structurally unresolved parts of NELF-E (shown as a dashed line) are shown in context of the entire PEC (PDB code 6GML). RNA crosslinked residues on NELF-E are highlighted in yellow. Residues on Pol II and DSIF previously reported to form lysine-lysine crosslinks with NELF-E along its unresolved parts are shown as black spheres (Vos et al., 2018). **(B)** RNA crosslinks to the NELF-A tentacle and to the surface of Pol II. The position of the unresolved NELF-A tentacle is indicated as a dashed line. RNA-crosslinked residues on the NELF-A tentacle and on Pol II are shown as yellow spheres. Previously identified protein-protein crosslinks between the NELF-A tentacle and Pol II are shown as black spheres (PDB code 6GML) (Vos et al., 2018). **(C)** RNA-crosslinked residues on Pol II (shown as yellow spheres) which are located at the interface between Pol II and the NELF-AC lobe. Previously reported residues

on Pol II that form lysine-lysine crosslinks to the NELF-AC lobe are shown as black spheres (PDB code 6GML) (Vos et al., 2018). Adapted from Gorbovytska et al., 2021.

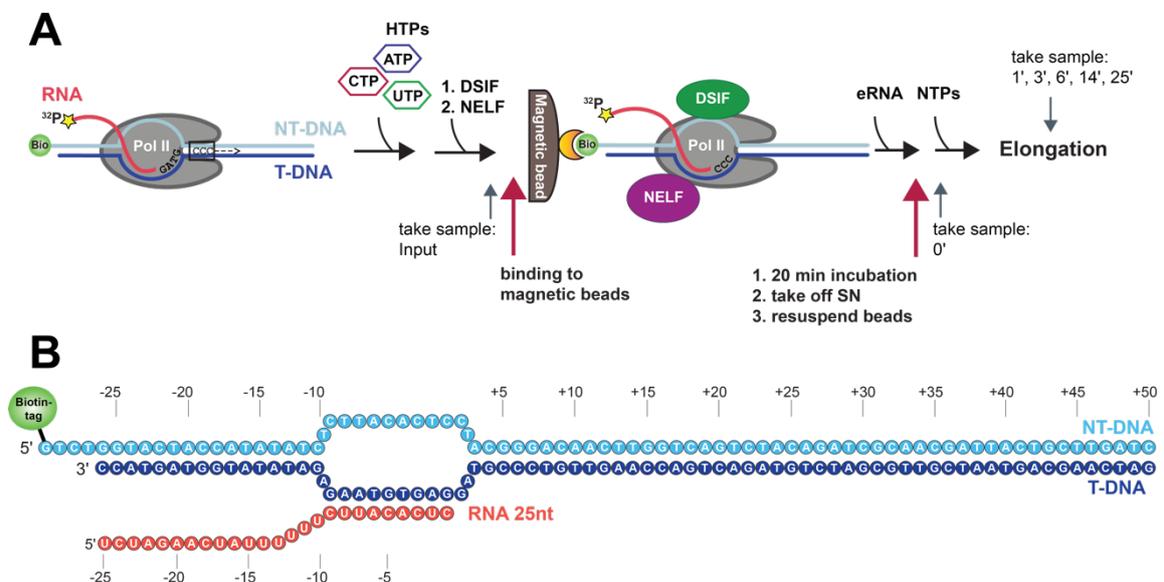
Even more, emanating from our crosslink data, we envision that eRNA binding to NELF could disrupt critical contacts between NELF and Pol II, thereby triggering its dissociation. In support of the crucial nature of eRNA length for facilitating NELF release, a 3D model of *Arc* eRNA (1-200) was computed using the Rosetta-based FARFAR2 tool (Lyskov et al., 2013; Watkins et al., 2020) and constrained by the experimentally determined SHAPE-MaP secondary structure for *Arc*. Remarkably, *Arc* eRNA 1-200 folds into a structure with dimensions that are similar to the dimensions of the large PEC complex (Figure 3-22). Thus, eRNAs with a length of 200 nt and beyond exhibit dimensions that allow for simultaneous eRNA binding to NELF-AC and NELF-E. Last, it is of note that we also observed RNA crosslinks to the NELF-BE lobe (Figure 3-20A; Tables 7-26 and 7-27). This finding is consistent with the previous study, which reported besides NELF-AC, also nucleic acid binding to NELF-B *in vitro* and *in vivo* (Vos et al., 2016). However, our EMSA results using the NELF double mutant, demonstrated that the NELF-BE lobe is not critical for the eRNA-triggered NELF release *in vitro* (Figure 3-17A-C). Though, it cannot be ruled out that interactions between eRNAs and NELF-B could contribute to the general attraction of eRNAs to the PEC (Gorbovytska et al., 2021).



**Figure 3-22 | 3D Structure model of *Arc* eRNA 1-200 next to the PEC.** Front view of the PEC (PDB codes 6GML and 2JX2). Alongside the PEC, a 3D model of *Arc* eRNA (1-200) is shown. The *Arc* eRNA (1-200) model was generated using SHAPE restrained Rosetta modelling with FARFAR2 (Watkins et al., 2020). The model supports that eRNAs need to be 200 nt long in order to efficiently trigger NELF release from the PEC (see Figure 3-12). Reprinted from Gorbovytska et al., 2021.

### 3.5 *In vitro* Transcription assays

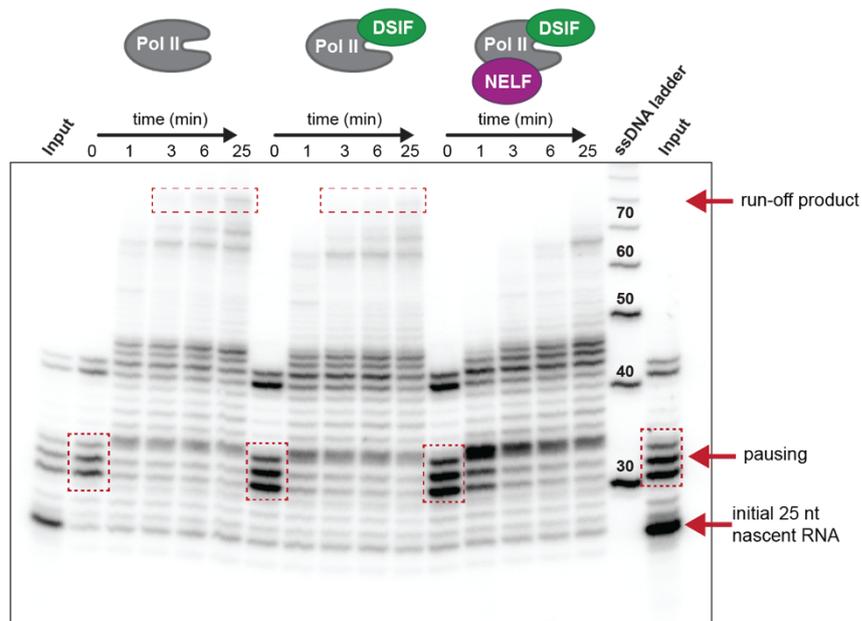
The release of NELF from the paused elongation complex is a hallmark of the transition from promoter-proximal pausing to productive elongation. The EMSA results already demonstrated that enhancer RNAs are able to dissociate NELF from the PEC *in vitro* (Figure 3-23). Next, I tested whether it is possible to assess the impact of NELF release in a functional assay. To address this question, I established an *in vitro* transcription assay based on previously published assay setups and pilot experiments carried out by Lisa-Marie Schneider (Dengl & Cramer, 2009; Missra & Gilmour, 2010; Vos et al., 2018) (Figure 3-23). For this purpose, the PEC has been assembled analogously to the EMSA experiments, except that longer DNA strands were utilized, bearing a short (4 nt) G-less cassette downstream of the RNA-DNA hybrid region. Furthermore, the non-template strand carried a biotin-tag that allowed to purify the assembled PEC from unbound protein and nucleic acid components. The transcription is pausing at the end of the G-less cassette when no GTP is supplied. The addition of DSIF and NELF stabilizes the pause, and transcription can be resumed afterwards by the addition of GTP/NTPs (Figure 3-23A,B). The addition of eRNAs was expected to facilitate the pause release.



**Figure 3-23 | Schematic, stepwise representation of the *in vitro* transcription assay.** (A) Pol II is bound to a transcription bubble consisting of a template DNA (T-DNA), non-template DNA (NT-DNA) and a radioactively 5' end labeled nascent RNA. Addition of HTPs (mix of ATP, CTP, UTP) lets Pol II transcribe four nucleotides until it pauses due to the lack of GTP. Addition of DSIF and NELF stabilizes the paused Pol II by forming the PEC. The fully assembled PEC is purified from unbound proteins and nucleic acids over magnetic Strep-beads via the biotin-tag at the 5' end of the NT-DNA strand. The bead-bound PEC is incubated with an eRNA and transcription is resumed after the removal of the eRNA containing supernatant (SN) by addition of NTPs. Samples are taken

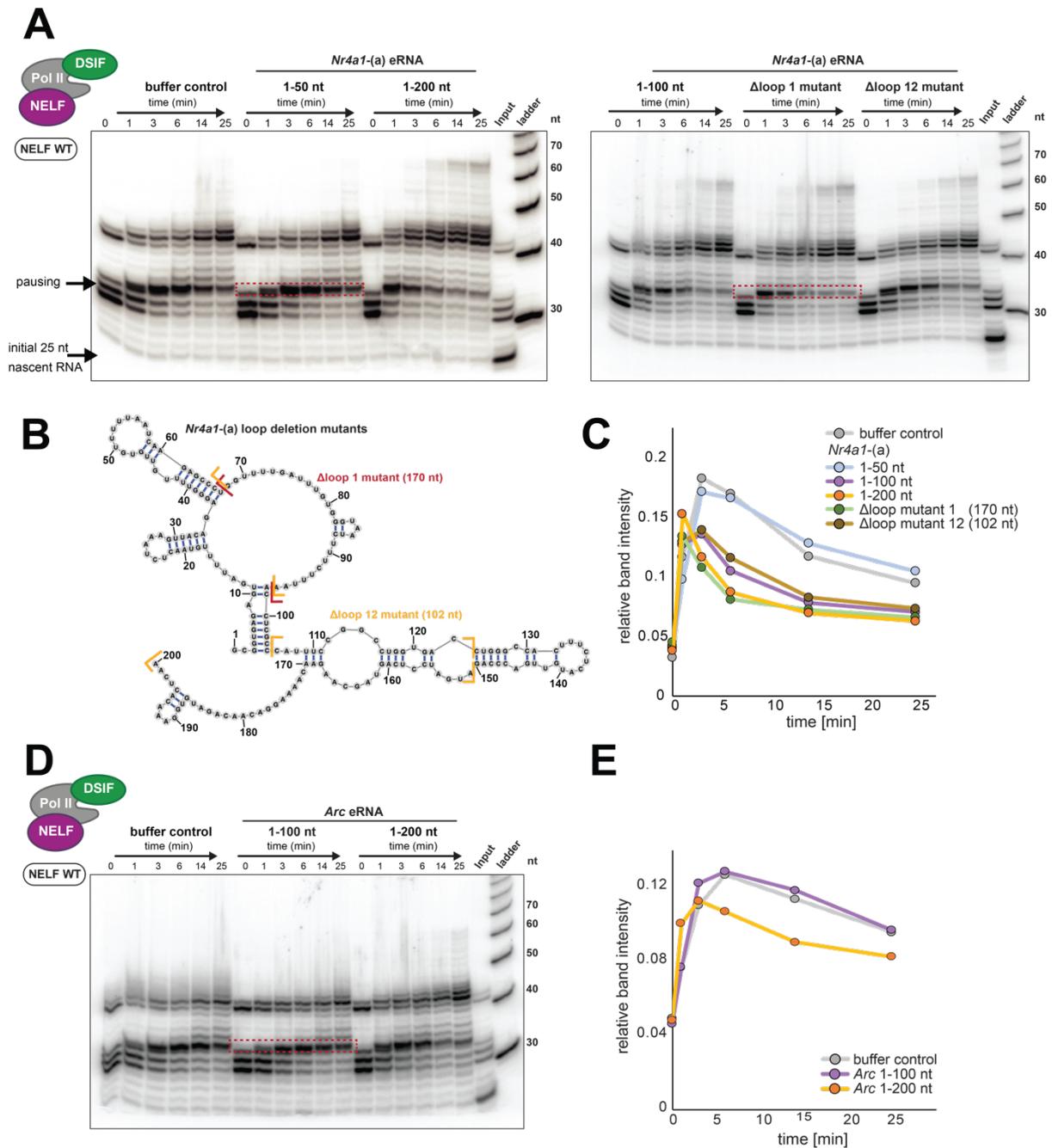
after different time points to trace the progress of transcription. **(B)** Nucleic acid scaffold used for the transcription assay described in (A). The scaffold consists of a T-DNA strand (76 nt), a fully complementary NT-DNA strand (80 nt) and a nascent RNA (25 nt). The NT-DNA has a 4 nt overhang at the 5' end, which is modified with a biotin tag. (A) is reprinted from Gorbovytska et al., 2021.

Before examining the effect of eRNAs, I verified the experimental setup and that the formation of the PEC leads to an observable pause stabilization as expected (Figure 3-24). Transcription carried out with Pol II only and the Pol II-DSIF complex lead to a rapid elongation of the pause product bands and generated run-off transcripts already after 3 minutes. In contrast to this, in the presence of both DSIF and NELF, the pause product bands were elongated more slowly, and no run-off product was generated. The observation that the transcription is only slowed down by the pausing factors and not entirely halted is consistent with previously published results of *in vitro* elongation assays (Misra & Gilmour, 2010).



**Figure 3-24 | Presence of DSIF and NELF slows down Pol II transcription.** The transcription assay was performed as outlined in Figure 3-23 and samples were resolved on a 15% urea gel together with a ssDNA ladder (20/100 nt ssDNA; IDT). Transcription was performed without DSIF and NELF, in the presence of DSIF only or in the presence of DSIF and NELF together. No run-off transcript is formed in the presence of both pausing factors, DSIF and NELF and the release from the pause site is reduced, as the pausing bands (triple band boxed in red) disappear much slower with increasing transcription time. Adapted from Gorbovytska et al., 2021.

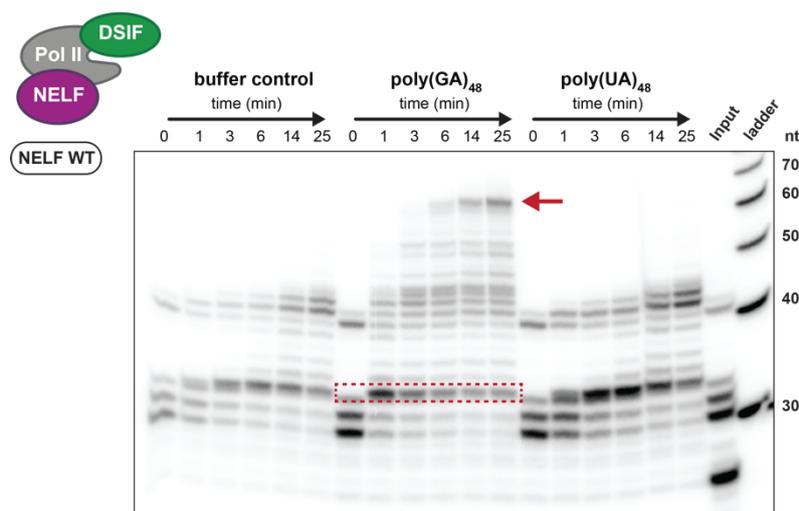
When long *Nr4a1*-(a) or *Arc* eRNA fragments were added to the PEC, Pol II resumed transcription more rapidly (Figure 3-25A,D). Of note, the short *Nr4a1*-(a) (1-50) fragment did not trigger any visible pause release similar to the buffer control (left gel Figure 3-25A), while the *Arc* and *Nr4a1*-(a) 1-200 (left gel Figure 3-25A,D) fragments evoked a considerable release from the paused state.



**Figure 3-25 | eRNAs fasten the pause release of Pol II in the transcription assay. (A)** Transcription assay using wild type NELF (WT) and *Nr4a1*-(a) eRNA fragments (left gel: 1-50 and 1-200 nt; right gel: 1-100,  $\Delta$ loop 1 mutant (170 nt),  $\Delta$ loop 12 mutant (102 nt)). Samples were taken just before NTP addition (0 min) and at different time points after NTP addition (1, 3, 6, 14 and 25 min). The "input" sample contains the PEC sample before its affinity purification using the streptavidin-coated beads (for outline see Figure 3-23A). Thus, it allows for the visualization of unbound nascent RNA. Elongation products were separated on a 15% Urea gel. **(B)** secondary structure of *Nr4a1*-(a) (1-200) is shown and the deleted regions in the *Nr4a1*-(a)  $\Delta$ loop 1 (red) or the  $\Delta$ loop 12 mutant (orange) are marked with brackets. **(C)** To follow the rate of pause release of the assay shown in (A), the intensity of the first transcript elongation band past the pause (boxed in red) was quantified and plotted against the time. **(D)** Shows the urea gel for a transcription assay as described in (A) only for *Arc* eRNA 1-100 and 1-200 fragment. **(E)** Quantification of the gel shown in (D), as described in (C). Adapted from Gorbovytska et al., 2021.

To monitor the pause release dynamics between the different eRNA fragments, the normalized intensity (see section 5.24) of the transient RNA product band (highlighted with a red box in Figure 3-25A) above the upper pause band was plotted against the transcription time (Figure 3-25 C,E). The addition of 1-200 eRNA fragments led to a rapid increase and subsequent rapid decrease of this band, reflecting fast pause release kinetics. For the buffer control and the *Nr4a1*-(a) 1-50 fragment, the dynamic intensity change of the same band was much more slowed down. Pause release caused by the *Nr4a1*-(a) 1-100 fragment was slightly reduced, as compared to the 1-200 fragment. Furthermore, the tested *Nr4a1*-(a) eRNA deletion mutants, which were lacking single-stranded regions from the 1-200 fragment (right gel Figure 3-25A,B), displayed an extent of pause release rather corresponding to their length than to their structure. The effect of the  $\Delta$ loop 1 mutant (170 nt long) was similar to the effect of the wild type (WT) 1-200 fragment, and the effect of the  $\Delta$ loop 12 mutant (102 nt long) had a similar magnitude as the WT 1-100 fragment. These results are in striking analogy to the obtained EMSA results (Figure 3-12A-C), as first the effect was critically dependent on eRNA length and less on the "structuredness" of the tested eRNAs (Figure 3-13A).

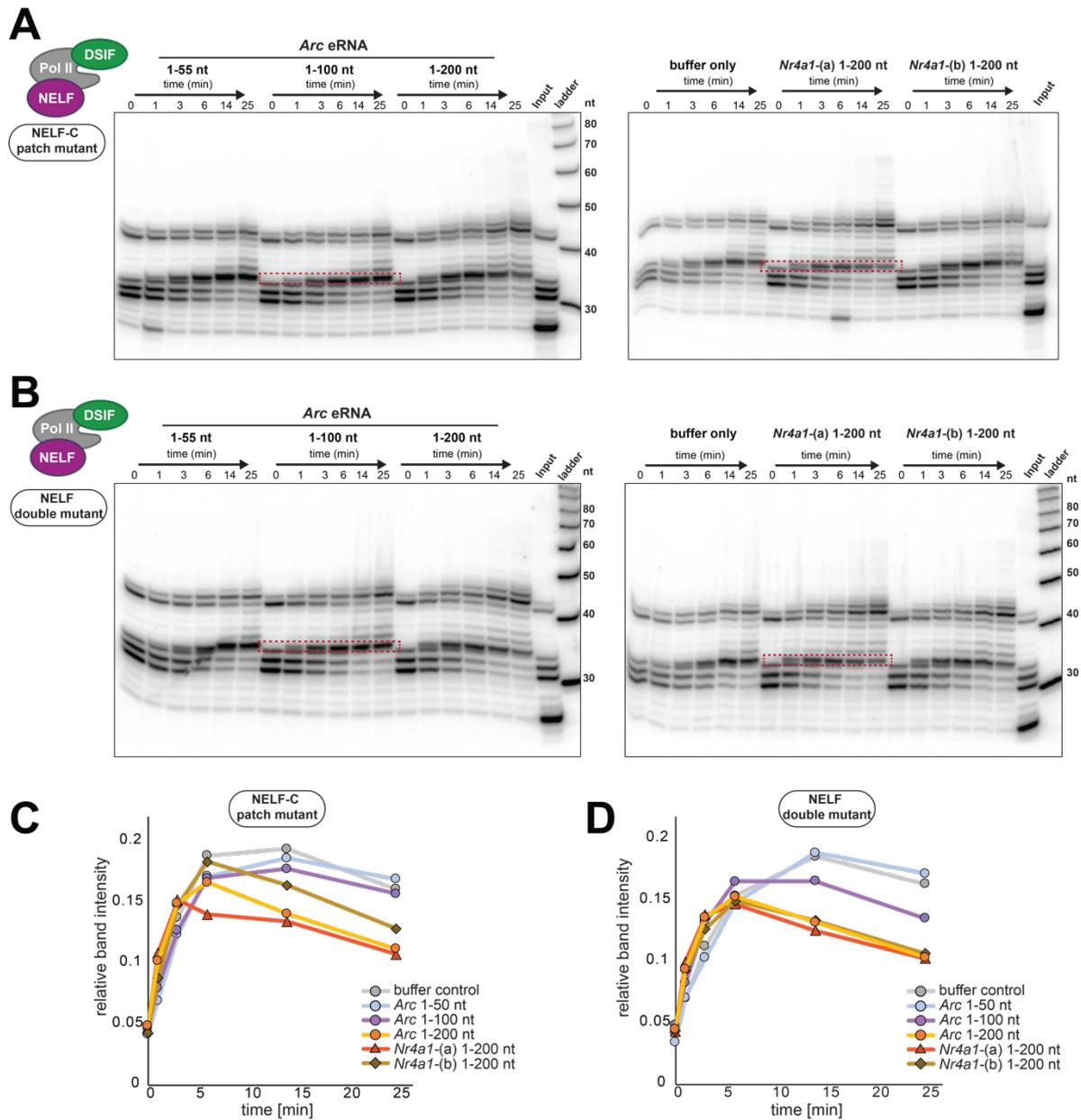
To further corroborate the agreement of the transcription data with our EMSA results, I performed a transcription assay with poly (GA)<sub>48</sub> and poly (UA)<sub>48</sub> RNAs. While poly(GA)<sub>48</sub> facilitated pause release dramatically, poly (UA)<sub>48</sub> had no effect on pause release (Figure 3-26). This result further supports the guanosine dependence of the NELF release, which was observed with the EMSA experiments (Figure 3-14B).



**Figure 3-26 | Transcription assay with repetitive G-containing and G-less RNA.** Urea gel of a transcription assay performed with the 96-mer poly (GA)<sub>48</sub>, and poly (UA)<sub>48</sub> RNA, as described in Figure 3-25A. poly (GA)<sub>48</sub> strongly induces the pause release as reflected by the immediate decrease of the transient band (boxed in red) above the pausing bands and the rapid appearance of the run-off transcript (red arrow).

Next, I measured eRNA-induced Pol II pause release with a PEC variant that comprised the NELF-C patch mutant and utilizing *Arc*, *Nr4a1*-(a) and -(b) eRNAs. As shown in Figure 3-27A,C *Arc* eRNA (1-50) or (1-100) hardly induced any pause release and the longer eRNA variants (1-200) exhibited substantially diminished rates of release. This is, again, in good agreement with our EMSA data (Figure 3-16D-F). It is to note, that the NELF-C patch mutant apparently leads to an overall increase in Pol II pause stabilization. This becomes evident when comparing the WT NELF (Figure 3-25C) and the patch mutant dynamics of pause release (Figure 3-27C). In the case of WT NELF, the curves show a rapid increase during the first 3 minutes of transcription and subsequently a rapid decrease. When using NELF-C patch, the curves rise much slower during the initial 6-14 minutes and show hardly any subsequent decline. Finally, when testing the NELF double mutant (both lacking the NELF-E RRM domain and comprising the NELF-C patch mutant) Pol II pause release rates similar to those for the NELF-C patch mutant were obtained (Figure 3-27B,D).

Taken together, the results of the transcription assay confirm our prior findings from the EMSA experiments and further establish that, indeed, eRNAs of sufficient length are able to release Pol II from its paused state by facilitating NELF release. Notably, the eRNA-induced increase in Pol II pause release efficiency was much less dramatic compared to the highly efficient release of NELF that we observed upon eRNA addition in the EMSA assays (Figure 3-12). This is likely the case because the sole dissociation of NELF may not be sufficient to jump-start transcription *in vitro*, even more so as NELF was reported to induce a tilted conformation of the DNA-RNA hybrid in the active site of Pol II (Vos et al., 2018). Furthermore, the established transcription assay system does not recapitulate P-TEFb phosphorylation (Cheng & Price, 2007; Yamada et al., 2006), which could additionally impede an efficient resumption of transcription elongation *in vitro*.



**Figure 3-27 | Reduced pause release with NELF-C patch and double mutant.** Transcription assay, using the NELF-C patch mutant (A) or NELF double mutant (B) and *Arc* eRNA (1-55, 1-100 and 1-200), *Nr4a1*-(a) and *Nr4a1*-(b) 1-200, performed as described in Figure 3-25. (C-D) Quantification of the assays that are shown in (A-B), performed as described in Figure 3-25C. Adapted from Gorbovytska et al., 2021.

## 4 Discussion

Enhancer RNA (eRNA) transcription was shown to be rapidly induced by external stimuli. eRNAs, in turn, facilitate the expression of their target genes by deploying various mechanisms and interacting with different transcription factors at different steps of the transcription process (Sartorelli & Lauberth, 2020). This thesis ought to reveal whether eRNAs play a role in the release of Pol II from promoter-proximal pausing, as previously hypothesized for neuronal eRNAs (Schaukowitch et al., 2014). In more detail, the previous study suggested that eRNAs could act as a decoy for the pausing factor NELF and thereby facilitate the transition of paused Pol II into productive elongation.

### 4.1 eRNAs function is not associated with a common structural motif

eRNAs do not feature any sequence motif that could explain their function. Therefore, we asked whether eRNAs could share a common structural motif, making them capable of exerting their function in promoter-proximal pause release. Consequently, we determined the precise 5'-ends of neuronal enhancer RNAs by Exo-seq and then mapped the secondary structure of the 5' terminal sequences (1-200 nt fragments) of a selected set of 39 eRNAs by chemical structure probing using the SHAPE-MaP technique. The results of this approach revealed that the eRNAs, at least in our set, do not possess a specific structure motif. As our set of eRNAs include those of prominent neuronal IEGs, like *Arc*, *Nr4a1*, *Fos*, *Fosb*, and *Junb*, we assume that the set is representative for neuronal eRNAs. Our results further indicate that eRNAs are not characterized by a particularly high or low degree of "structuredness" (Figure 3-8). For accurate subsumption of the overall "structuredness" for our eRNA set, we lack SHAPE-MaP data on a typical set of mouse mRNAs or other RNA classes. The SHAPE-MaP experiment would need to be performed under the same experimental conditions as for eRNAs. A comparison between experimental, SHAPE-MaP based structure data of eRNAs with secondary structure data based on prediction (see section 3.2.3) can only serve as a benchmark but might bias the outcome.

However, the lack of any specific sequence or structure motifs is in agreement with previous studies on eRNA function, that reported on rather general than sequence-specific interaction between eRNAs and the targeted transcription factors/co-activators like YY1 and

CBP (Bose et al., 2017; Sigova et al., 2015). Such broad, sequence-unspecific interactions were also shown for RNA binding to the Polycomb repressive complex 2 (PRC2) (read next section) and the CCCTC-binding factor (CTCF) (Davidovich et al., 2013; Saldaña-Meyer et al., 2019). Moreover, "non-specific" binding between RNA and proteins is not a rare case. Roughly half of all protein-RNA interactions fall into this category, as specificity is even disadvantageous for many cellular processes (Jankowsky & Harris, 2015). In the case of eRNA binding to NELF or other transcription factors (see above), the specificity of interactions can result from the *cis*-acting nature of enhancers and promoters. As described in the introduction (section 1.4), active enhancer and promoter regions reside in close proximity due to the formation of chromatin loops, sometimes referred to as chromatin hubs, which activate transcription (de Laat & Grosveld, 2003). These chromatin loops or hubs create local hot spots for transcription factors and eRNAs, which get trapped in those hubs. The spatial delimitation of the eRNAs and TFs, associated with one enhancer and target promoter pair, finally dictates the specificity of their interactions.

## 4.2 eRNAs can detach NELF from the paused elongation complex *in vitro*

Within this study, we successfully established two *in vitro* assays to investigate the impact of eRNAs on the paused elongation complex (PEC). Using our EMSA setup (Figure 3-11), we could clearly show that eRNAs are potent to dissociate NELF from the PEC without affecting the binding of DSIF (Figure 3-12). This result is consistent with the general knowledge about the promoter-proximal pause release during which NELF dissociates while DSIF stays bound to Pol II. Further, the data mechanistically substantiate the underlying study that suggested enhancer RNAs to be involved in the release of Pol II from promoter-proximal pausing by decoying NELF from paused Pol II (Schaukowitch et al., 2014). We further observed that the dissociative effect of the eRNAs is strongly dependent on their size. The longer the RNA (> 200 nucleotides), the more efficient it detaches NELF from the PEC (Figure 3-12A-C compare top and bottom panels). The second RNA feature that we found to be important to trigger the efficient release of NELF is unpaired guanosines distributed along the entire enhancer RNA sequence (Figures 3-14B,C and 3-15). Using eRNA mutants, in which the secondary structure was altered to be more (by deleting single-stranded regions) or less double-stranded (by removing stem structures) did not significantly affect their potency to trigger NELF release.

This finding points out that the secondary structure of eRNAs does not play a critical role in their action (Figure 3-13). Furthermore, this finding matches our secondary structure probing result that attested no general structure similarities among eRNAs.

Both the size of the eRNA and the presence of unpaired guanosines are rather loose requirements for the function of eRNA in dissociating NELF. Interestingly, both criteria are highly reminiscent of formerly reported interactions between RNA and the Polycomb repressive complex 2 (PRC2) (Davidovich et al., 2013). PRC2 is a complex involved in the epigenetic silencing of genes and comprises a histone methyltransferase activity that trimethylates H3K27 histone residues (Margueron & Reinberg, 2011). Initially, PRC2 was shown to exhibit promiscuous and size-dependent RNA binding, with a lower affinity for shorter RNAs (Davidovich et al., 2013, 2015). Subsequently, it was reported that PRC2 binds preferentially to guanines (Gs) in single-stranded regions and with an even higher affinity to G-quadruplex structures (G4) (Wang et al., 2017). A recent study of RNA-PRC2 binding confirmed the preferential binding to G-rich sequences in CLIP data and showed that G4 structures are enriched at the PRC2-bound RNA transcripts (Rosenberg et al., 2021). In our study, poly(GA) and poly(GU) RNAs that are G-rich yet should not form any G4 structures (require guanosine repeats with at least two consecutive Gs (X. Wang et al., 2017)) appeared to be highly potent in dissociating NELF but subsequently also DSIF from Pol II. The dissociation of DSIF is not consistent with the known outcome of the promoter-proximal pause release, where only NELF dissociates from paused Pol II while DSIF should stay bound. Remarkably, when the density of Gs was reduced, as in the poly(G<sub>2</sub>A<sub>6</sub>) RNA, NELF was still efficiently released without a subsequent dissociation of DSIF (Figure 3-14C). Of note is that G-rich RNAs (poly(GA), poly(G<sub>2</sub>A), poly(G<sub>2</sub>A<sub>3</sub>)) that triggered the undesired DSIF release exhibited unnaturally high G-contents (50%, 67%, 41%). In contrast, the poly(G<sub>2</sub>A<sub>6</sub>) RNA that did not trigger dissociation of DSIF has a G-content of 27% that matches more the general G-content of mouse RNA transcripts (mouse cDNA: median: 25%; upper quartile: 27%; 5' UTR: median: 30%; upper quartile: 35%) (Sadovskaya et al., 2020). Whether G-quadruplexes may also play a role in the release mechanism of NELF is still an open question. According to its sequence, poly(G<sub>2</sub>A<sub>6</sub>) RNA should be capable of forming G-quadruplexes. However, our experiments so far did not directly address the role of G4 structures and their effect on NELF release compared to unpaired Gs.

The *in vitro* pause release transcription assay, established in this thesis, corroborated the EMSA results. The addition of eRNAs caused a faster release of Pol II from the pause site in an eRNA size-dependent and guanosine-dependent manner (Figures 3-23 through 3-27). However, the kinetics were not comparable to the transcription in the absence of the pausing factors NELF and DSIF, indicating that the simple dissociation of NELF after the pause does not suffice to restore productive elongation. It is possible that the experimental setup of the transcription assay, which is performed on beads, precludes a complete dissociation of NELF due to sterical hindrance of eRNA binding. Moreover, even when assuming that NELF is efficiently dissociated from the PEC, there is still a rationale for why Pol II could be hindered from resuming productive elongation. First, phosphorylation of DSIF and Pol II CTD by P-TEFb happens during the canonical pause release mechanism. Even if NELF is dissociated from the PEC by the action of eRNAs, phosphorylation of DSIF could be required for the efficient transition into productive elongation. The binding of NELF to Pol II-DSIF might induce a conformation in the Pol II-DSIF complex that inhibits transcription, as long as DSIF and Pol II are not phosphorylated. Second, pausing is known to be accompanied by backtracking of Pol II along the template DNA. Through backtracking, the 3'-end of the nascent RNA becomes disengaged with the active site of Pol II. Such backtracked complexes need the action of the factor TFIIIS to resume transcription. For this, TFIIIS stimulates the cleavage of the nascent RNA, generating a new 3' end, which is properly aligned with Pol II's active site (Adelman et al., 2005; Cheung & Cramer, 2011; Sheridan et al., 2019). To prove these assumptions, the transcription assays would need to be performed in the presence of P-TEFb and TFIIIS.

### **4.3 NELF release is triggered through multivalent interactions with an eRNA molecule**

As eRNAs trigger NELF dissociation, this likely happens through mutual interactions between eRNAs and NELF. This study reveals that two independent interfaces on NELF are mainly responsible for the eRNA-driven release of NELF. These two interfaces are the RRM domain and the positively charged patches on NELF-AC lobe. When we replaced WT NELF with a mutant lacking the RRM domain (NELF $\Delta$ RRM mutant), we observed a reduction of the dissociative effect of eRNAs on this mutant compared to WT (Figure 3-16A-C). However, when we employed a patch mutant NELF, in which the positively charged residues (arginine and

lysine) of the surface of NELF-C were substituted with neutral residues (glutamine and methionine), the reduction of the eRNA-triggered NELF dissociation was even more pronounced. Finally, when utilizing the NELF double mutant, that unites the  $\Delta$ RRM and the patch mutant, the dissociative effect of eRNAs was fully abolished. While the RRM domain was already implicated in binding to the nascent RNA in the paused elongation complex and eRNAs (Schaukowitch et al., 2014), the function of additional nucleic acid binding interfaces previously reported for NELF was obscure (Vos et al., 2016). The binding of nucleic acids to positively charged patches on NELF-AC was reported for the isolated NELF complex, without the context of Pol II (Vos et al., 2016). Remarkably, these patches persist surface exposed when NELF is bound to Pol II (see PEC structure; PDB: 6GML) alluding to their relevance for promoter-proximal pausing (Vos et al., 2018). Based on our results, we can definitely attribute a central role in the eRNA-dependent pause release mechanism to these patches. What is more, the striking length dependency of eRNA-driven NELF dissociation (Figure 3-12) and the significant distance between the NELF-AC lobe and the NELF-E RRM domain, both of which we find directly involved in promoting NELF dissociation upon eRNA binding (Figures 3-16 and 3-17), allow us to suggest that eRNAs simultaneously occupy several binding sites across the PEC to trigger NELF release (Figure 4-1).

Further, our protein-RNA crosslinking data show that eRNAs extensively contact the NELF-A tentacle, the staircase domain of NELF-B, and the NELF-E N-terminal region (Figures 3-20A and 3-21). Indeed, all these additional regions, beyond the patches on NELF-AC and the RRM-domain, seem to contribute to NELF's affinity for nucleic acids, as judged by the comparison of the previously published  $K_d$  values for different NELF variants (Table 4-1) (Vos et al., 2016).

**Table 4-1** | Published  $K_d$  values for nucleic acid binding of different NELF constructs (from Vos et al., 2016)

<b>NELF construct</b>	<b><math>K_{d,app}</math> (<math>\mu</math>M)</b>
NELF-A (6-188) + NELF-C (183-590)	6.87 $\pm$ 0.46
NELF-B	8.50 $\pm$ 1.59
NELF-B + NELF-E (1-137)	2.83 $\pm$ 1.00
NELF-ABC	0.074 $\pm$ 0.014
NELF-ABC patch mutated -C	0.290 $\pm$ 0.99
NELF $\Delta$ RRM	0.03 $\pm$ 0.007
NELF $\Delta$ RRM patch mutated -C	0.094 $\pm$ 0.020

Notably, the NELF-A tentacle was previously shown to be critical for the stable binding of NELF to Pol II-DSIF and pause stabilization (Narita et al., 2003; Vos et al., 2018). As we further found multiple protein-RNA crosslinks at the interface of Pol II with NELF-AC, NELF-A, and NELF-E, we assume that eRNAs might perturb the interactions between NELF and Pol II at multiple sites (Figures 3-21 and 3-22). This idea is consistent with the highly charged binding interface between NELF and Pol II (Vos et al., 2018). RNA is a highly polar and charged molecule and is therefore well-suited to interfere with and abrogate the binding of NELF and Pol II.

#### 4.4 Reported preferences of the RRM domain and NELF-AC lobe for guanosines

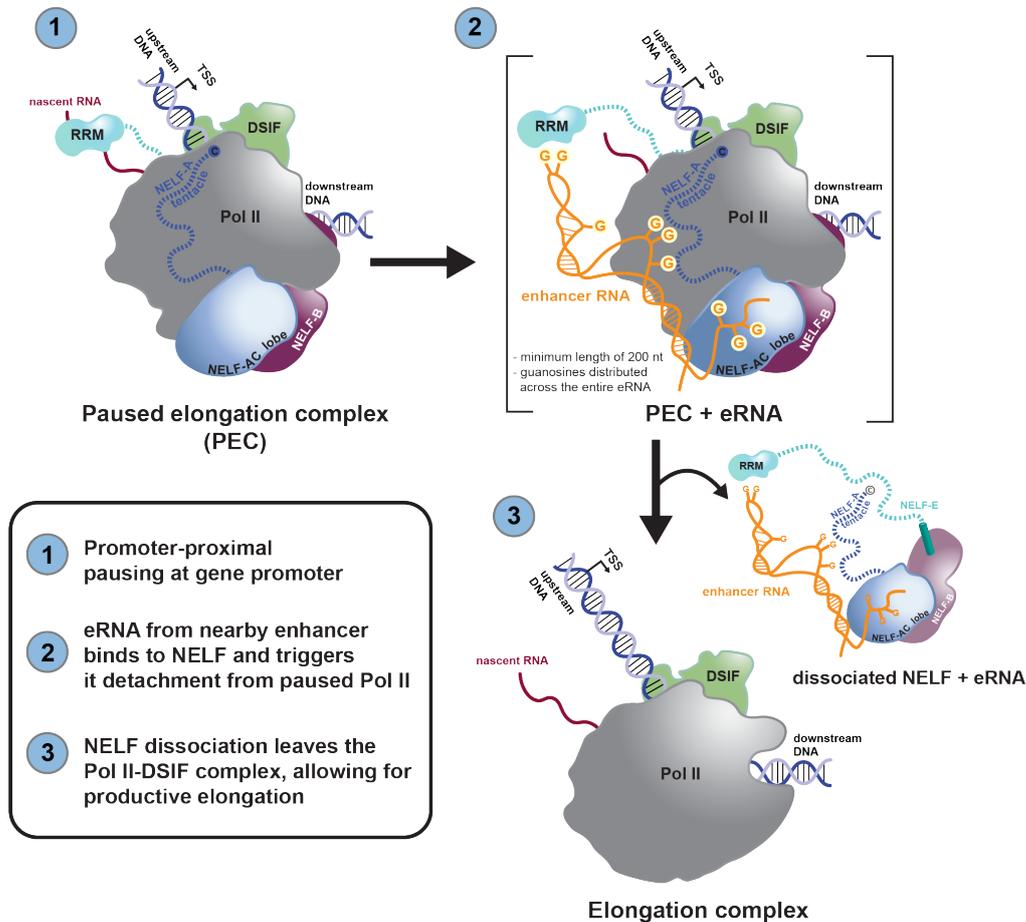
Our study demonstrated that unpaired guanosines play an important role in the NELF-dissociating capability of eRNAs (Figures 3-14B,C and 3-15). This finding connects well to previous reports on RNA binding of both the RRM-domain and the NELF-AC lobe. First, the NELF-AC subcomplex was shown to bind single-stranded RNA with a GC content of 60%, but not RNA with a GC content of 44% (Vos et al., 2016). A closer inspection of the utilized RNA sequences used in that study reveals that the RNA with the 44% GC-content contained no guanosines at all, while the RNA with the 60% GC content comprised 56% Gs (14 Gs in a 25-mer RNA). Second, preferential binding to guanosines was also reported for the *D. melanogaster* homolog of NELF-E and its RRM domain (Pagano et al., 2014). The study showed that *Drosophila* RRM-domain and full-length NELF-E are binding with very high affinities to the sequence CUGGAGA(U), which was termed the NELF-E binding element (NBE). For efficient binding to NELF-E, the NBE must be located in a single-stranded region, either a hairpin loop region or without any secondary structure environment. Mutating all guanosines in the NBE motif to adenosines abolished its binding to NELF-E and the RRM-domain, highlighting their preference for Gs (Pagano et al., 2014). However, in contrast to *Drosophila* NELF-E, the human homolog does not exhibit such a pronounced affinity for the NBE motif. A comparison between the fly and human NELF-E reveals that the *Drosophila* homolog lacks the RD repeat domain (Figure 1-4). This unstructured repetitive domain, whose function is unclear, might contribute to the lower sequence-specificity of human NELF-E, enabling it to bind a broader range of RNAs, as our data indicate.

## 4.5 A revised model for eRNA-triggered NELF dissociation

Taken together, this study provides a revised model for eRNA-driven NELF release and gives molecular insights into the underlying mechanism. The publication by Schaukowitch et al., which was the starting point for this study, proposed that activity-induced neuronal eRNAs decoy NELF from paused Pol II and thereby facilitate transcription of immediate early genes (IEGs) (Schaukowitch et al., 2014). The authors of the study showed that the RRM domain is critical for the expression of IEGs, which are known to be regulated by promoter-proximal pausing (Saha et al., 2011). Therefore, the authors suggested that the promoter-proximal pause at IEGs is stabilized by binding of the RRM domain to the nascent mRNA transcript. Furthermore, they hypothesized that upon depolarization of neurons, eRNAs are induced and will compete with the nascent mRNA for binding to the RRM domain. Thereby eRNAs would destabilize the PEC and cause the dissociation of NELF.

Using the well-defined experimental setup of an *in vitro* system (EMSA and *in vitro* transcription assays), we could validate that eRNAs indeed trigger the release of NELF from the paused elongation complex. Furthermore, our results clearly show that NELF features not only one but multiple RNA binding interfaces responsible for the eRNA-triggered NELF dissociation from the PEC. In addition to the previously suggested RRM domain, the positive patches on NELF-AC play a critical role in the release mechanism. Therefore, our data disagree with the hypothesis of Schaukowitch et al., who proposed that eRNAs might compete with the nascent RNA in the paused elongation complex for binding to the NELF-E RRM domain and thereby facilitate the release of the NELF complex. In contrast, our data argue that eRNAs extensively contact NELF interfaces outside of the RRM domain and that NELF release is triggered by allosteric contacts between the eRNA and NELF at interfaces outside the RRM domain. This is in agreement with the solved structure of the PEC (Vos et al., 2018), which proves that the RRM domain is not essential for the formation of the paused elongation complex and hence cannot be the sole target of the eRNA to drive the disassembly of the complex. Building on our findings, we envision that eRNAs are attracted in the first place to the PEC in a sequence motif-independent manner by the positively charged patches on NELF-AC, though the presence of unpaired guanosines likely facilitates the binding (see section 4.2). This initial binding event could be the seed for spawning further interactions between the

eRNA and the NELF-A tentacle, as well as the distant NELF-E RRM domain. Eventually, the sum of all eRNA-NELF interactions would then trigger NELF release (Figure 4-1).



**Figure 4-1 | Mechanistic model for the eRNA induced Pol II pause release.** We envision that a single enhancer RNA molecule (> 200 nt) interacts with Pol II-bound NELF at multiple sites (positive patches on NELF-AC, NELF-A tentacle, RRM domain). The interaction with the eRNA abrogates NELF binding to the PEC. NELF leaves the Pol II-DSIF complex, which can then resume transcription elongation. Unpaired guanosines within the eRNA strongly increase its affinity to NELF and facilitate the dissociation (Reprinted from Gorbovytska et al., 2021)

The lack of well-defined motifs in eRNAs and, at the same time, the critical role of guanosines for the interaction with NELF is reminiscent of the PRC2 interaction with RNAs (as discussed in section 4.2) (Davidovich et al., 2013, 2015; Wang et al., 2017). Intriguingly, G-tract-containing RNAs were shown to evict the PRC2 complex, responsible for gene silencing, from its nucleosome substrates, thereby leading to gene activation (Beltran et al., 2019; Wang et al., 2017). Hence, decoying protein factors with overall promiscuous RNA binding activity by high-affinity binding to G-rich RNAs might be a general concept in RNA-dependent regulation of gene expression. Furthermore, the lack of any strict sequence-structure motifs in eRNAs and the multivalency of the interactions between the eRNA and NELF fits nicely into the model of liquid-liquid phase separation (Alberti et al., 2019; Banani et al., 2017; Hyman et al., 2014).

The formation of membrane-less sub-cellular compartments by liquid-liquid phase separation functions to concentrate proteins and nucleic acids. Multivalent protein-protein and protein-RNA interactions, relying on intrinsically disordered regions (IDRs) of proteins and large RNA molecules, are major drivers for the assembly of the biomolecular condensates. These condensates can exert diverse functions and are involved in different cellular processes like RNA metabolism or stress response. Recently, transcriptional condensates (Cramer, 2019; Hnisz et al., 2017) and enhancer condensates containing eRNAs (Nair et al., 2019) were proposed as a general concept to regulate eukaryotic transcription.

What is more, the NELF-A and NELF-E tentacle constitute IDRs of NELF, which were shown to drive phase-separation of NELF during stress response (Rawat et al., 2021). These findings combined with ours hint at a model in which enhancer RNAs could contribute to the formation and/or could alter the properties of transcription condensates at promoter-proximal regions. Multivalent interactions within these condensates, as between an eRNA and multiple RNA binding interfaces on the pausing factor NELF, greatly enhance eRNA's capacity to regulate gene expression by abrogating promoter-proximal pausing. In this context, a recent *in vitro* study under cellular conditions demonstrated that eRNAs transcribed from super-enhancers affect condensate formation of the purified Mediator subunit MED1 (Henninger et al., 2021). While low eRNA concentrations promote the formation, high concentrations lead to the dissolution of the condensates.

## 4.6 Outlook

Building on the results of this thesis, it would be aspiring to get a closer view of the eRNA-triggered dissociation of NELF in terms of the molecular mechanism dynamics. Our data do not answer whether the interactions between the eRNA and the different binding interfaces on NELF are established sequentially starting at a specific site or whether the multiple interactions happen simultaneously. We speculated that the positively charged patches on NELF-AC could be the initial interface to attract the eRNAs. However, this does not necessarily have to be the case. So, it would be informative to study the dynamics of these interactions and the release of NELF in more detail. To this end, single-molecule fluorescence resonance energy transfer (smFRET) experiments could be employed, though designing and establishing this technique is highly complex and laborious (Broussard & Green, 2017; Roy et al., 2008).

FRET assays can also be used to compare WT NELF with the NELF mutants and elucidate whether the mutants per se prevent eRNA binding to NELF or only affect their ability to dissociate NELF.

Furthermore, it would be interesting to follow up on the idea of liquid-liquid phase separation. Appropriate experiments could be performed *in vitro* and also *in vivo*. One could test whether eRNAs and the paused elongation complex show a tendency to form droplets *in vitro* and whether they co-localize within the same condensates. It is tempting to speculate that eRNAs could form condensates with NELF after triggering its dissociation from the PEC. The remaining Pol II-DSIF complexes might be excluded from these droplets. In this context, it would also be interesting to see the impact of the eRNA's size and G-content on their ability to form putative condensates and alter their physical properties. The same could be checked for the phosphorylation status of Pol II CTD, DSIF and NELF.

More straightforward experiments immediately following from our results are to look closer into the guanosine dependency of the NELF dissociation. Our results did not exclude the involvement of higher-order poly(G) structures as G-quadruplexes. It would be informative to test whether the presence of G-quadruplexes, in analogy to PRC2 (Wang et al., 2017), affects the RNA's potency to dissociate NELF. Further, we determined two binding interfaces, the RRM domain and the patches on NELF-AC, to be critical for the eRNA-triggered dissociation of NELF from the PEC. However, our crosslink data and a previous report (Vos et al., 2016) suggest that NELF harbors even more interfaces that could potentially bind eRNAs. It would be useful to test whether the elimination of any other putative binding interface or combinations of them could abolish the dissociative function of eRNAs in the same way as the tested NELF double mutant ( $\Delta$ RRM/patch mutant) in this thesis. This would clarify the significance of each RNA binding interface on NELF and reveal whether the two interfaces we determined are indeed the only critical interfaces for eRNA-triggered release. It might still be the case that the deletion/mutation of any two RNA-binding interfaces on NELF impedes the action of eRNAs. Another open question was why the sole dissociation of NELF does not restore Pol II's full capacity to elongate the nascent RNA in our *in vitro* assay. To investigate this, one could test the impact of the factor P-TEFb and TFIIS on the *in vitro* assay outcome (see section 3.5). If the activity of TFIIS would further enhance the transcription, this would imply that efficient elongation was prevented by backtracking of Pol II, which can be rescued by TFIIS. If the activity of P-TEFb would improve the transcription, this could mean that

---

phosphorylation of DSIF is required to restore the ability of Pol II to elongate. Finally, protein-RNA crosslinking mass spectrometry using the PEC and an eRNA, as it was performed with wild type NELF, could be performed with the NELF mutants. This could give information on whether eRNAs would still interact with the PEC or whether the interactions are entirely abolished.

## 5 Materials and Methods

### 5.1 Chemicals, Materials and Equipment

All chemicals, materials and equipment used in this thesis are listed in the supplement section (section 7.1).

### 5.2 Bacterial strains and insect cells

Bacterial strains used for cloning and insect cells used for recombinant protein expression are listed below.

**Table 5-1.** Bacterial strains and insect cells

Strain / Cells	Species	Application	Reference
DH5 $\alpha$	<i>Escherichia coli</i>	Cloning	Prof. Dr. Stemmann (University of Bayreuth)
TOP10	<i>Escherichia coli</i>	Cloning	Prof. Dr. Stemmann (University of Bayreuth)
DH10EMBacY	<i>Escherichia coli</i>	Bacmid generation	Geneva Biotech
Sf21	<i>Spodoptera frugiperda</i>	MultiBac; virus generation	
High Five	<i>Trichoplusia ni</i>	MultiBac; Protein expression	

### 5.3 Transformation of plasmid DNA into competent cells

Luria-Bertani (LB)-medium:	10 g/L Tryptone, 5 g/L Yeast extract, 10 g/L NaCl
LB-Agar:	10 g/L Tryptone, 5 g/L Yeast Extract, 10 g/L NaCl, 15 g/L Agar
1000x antibiotic stock:	100 mg/mL ampicillin

Plasmid DNA was transformed by heat shock into chemically competent *E. coli* cells (DH5a or TOP10). To this end, an aliquot of *E. coli* cells was thawed on ice. Plasmid DNA (1  $\mu$ L or ca. 10  $\mu$ L of Gibson assembly (section 5.8) was added, and cells were incubated on ice for 10-15 min. The heat shock was performed in a water bath at 42 °C for 42 s, and the transformed cells were put back on ice for another 2 min. Subsequently, the cells were recovered at 37 °C and 900 rpm for 1 h in 1 mL of LB medium. 150-200  $\mu$ L of the cell suspension were plated out on LB-Agar

supplemented with antibiotics (100 µg/mL ampicillin for pUC18 plasmids) suitable to select positive transformants. Plates were incubated overnight at 37 °C.

## 5.4 cDNA synthesis

cDNA synthesis was used to produce DNA for cloning of eRNA sequences from mouse neurons. To this end, total RNA (ca. 3 µg per reaction) was reverse transcribed using the SuperScript III First-Strand Synthesis SuperMix (Invitrogen) with random hexamer primers, according to the manufacturer's instructions. The cDNA was amplified by PCR or stored at -80 °C.

## 5.5 Polymerase Chain Reaction

The polymerase chain reaction (PCR) was used to produce DNA inserts from cDNA (or plasmids) and DNA vectors for Gibson cloning or to produce DNA templates for *in vitro* production of eRNAs. The PCR mix and the thermocycler (peqSTAR, Peqlab) profile were set up based on the manufacturer's protocol for Phusion High-Fidelity DNA-Polymerase (NEB) (see example profiles below). Appropriate primer (supplementary section 7.2) combinations were used depending on the purpose.

Component	50 µL reaction	Step	Temp	Time
5x Phusion HF or GC buffer	10 µL	Initial denaturation	98 °C	1-2 min
dNTPs (10 mM)	1 µL	Cycle (35x):		
Primer forward (10 µM)	2.5 µL	Denaturation	98 °C	10 s
Primer reverse (10 µM)	2.5 µL	Primer annealing	60 °C	15-20 s
Template DNA	1-2 ng (plasmid) 300-500ng (cDNA)	Elongation	72 °C	20-25 s/kb
Phusion DNA Polymerase	0.5 µL	Final Elongation	72 °C	3-5 min
ddH <sub>2</sub> O	ad 50 µL	Hold	4 °C	

## 5.6 Colony PCR

Colony PCR was performed to screen *E. coli* colonies for positive clones after Gibson assembly. Taq DNA polymerase (homemade) was used instead of the high fidelity Phusion DNA polymerase. The PCR mix was prepared, and the thermocycler profile was set as described below, based on the manufacturer's instructions of Taq DNA polymerase (NEB). Typically, a mastermix was prepared and split into PCR tubes (each 25 µL). Single *E. coli* colonies were

picked from the LB-Agar plate using a sterile 10  $\mu\text{L}$  pipette tip and were dipped into one reaction mix. The tip was briefly stirred inside the mix before it was removed and stored in a 13-mL culture tube pre-filled with 0.5 mL LB medium, which was supplemented with a selective antibiotic (e.g. LB-amp; 100  $\mu\text{g}/\text{mL}$  ampicillin for pUC18 plasmids). Colony PCR was usually performed with pUC18 vector-specific primers, ER0033 and ER0034 (supplementary Table 7-18). Negative clones, lacking a DNA insert, produced a shorter amplicon than positive clones. The outcome of PCR was analyzed by agarose gel electrophoresis (section 5.7), and only positive clones were cultured overnight in 4-5 mL of LB-amp by shaking at 37  $^{\circ}\text{C}$  and 240 rpm for 12-14 h in an incubator (New Brunswick).

Component	25 $\mu\text{L}$ reaction	Step	Temp	Time
10x ThermoPol buffer	2.5 $\mu\text{L}$	Initial denaturation	98 $^{\circ}\text{C}$	5 min
dNTPs (10 mM)	0.5 $\mu\text{L}$	Cycle (25x):		
Primer forward (10 $\mu\text{M}$ )	0.5 $\mu\text{L}$	Denaturation	98 $^{\circ}\text{C}$	15 s
Primer reverse (10 $\mu\text{M}$ )	0.5 $\mu\text{L}$	Primer annealing	57 $^{\circ}\text{C}$	15 s
Taq DNA Polymerase	0.25 $\mu\text{L}$	Elongation	68 $^{\circ}\text{C}$	1 min/kb
ddH <sub>2</sub> O	ad 25 $\mu\text{L}$	Final Elongation	72 $^{\circ}\text{C}$	5 min
		Hold	4 $^{\circ}\text{C}$	

## 5.7 Agarose gel electrophoresis and gel extraction

50x TAE-Buffer (stock): 2 M Tris base, 1 M acetic acid, 50 mM EDTA (pH 8.0)

Agarose gel electrophoresis was used in a preparative (after Gibson assembly (section 5.8)) and analytical (e.g., after colony PCR (section 5.6)) manner to analyze and purify DNA fragments after a PCR. Samples were prepared by combining the DNA sample (PCR-mix) with 6x EZ-Vision One DNA loading dye (Amresco) to achieve a final concentration of 1x. The sample was loaded together with a suitable DNA ladder (100 bp DNA ladder or 1 kb ladder (NEB)) on a 1% – 2% agarose gel, depending on the size of the DNA fragments, and electrophoresed at 130 V in 1x TAE buffer until the desired resolution was achieved. Gels were then analyzed under UV-light. In the case of preparative gels, the target DNA bands were excised with a scalpel, and DNA was extracted from the gel using the QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions. The DNA concentration was determined by absorbance at 260 nm using a spectrophotometer (Eppendorf Biospectrometer basic).

## 5.8 Cloning of DNA by Gibson Assembly

5x Isothermal (ISO) reaction buffer: 500 mM Tris/HCl (pH 7.5), 50 mM MgCl<sub>2</sub>, 50 mM DTT, 1 mM each dNTPs, 5 mM NAD<sup>+</sup>, 25 % (w/v) PEG8000

Gibson Master-Mix: 1x ISO reaction buffer\*, 0.025 U/μL Phusion DNA Polymerase (NEB); 4 U/μL Taq Ligase (NEB); 0.004 U/μL T5 exonuclease (NEB) (stored at -20 °C in 7.5 μL aliquots)

\* 1x ISO reaction buffer was used instead 1.33x as stated in the publication (Gibson et al., 2009)

Cloning of eRNA fragments into a plasmid vector (mostly pUC18 vector) was carried out using the ligation independent Gibson assembly method as previously described with minor changes (Gibson et al., 2009). The protocol (one-step isothermal DNA assembly protocol) was adapted for our laboratory use by Dr. Felix Klatt. The insert DNA (starting from cDNA) and the plasmid vector for Gibson assembly was produced by PCR with insert-specific and vector-specific primers (supplementary Tables 7-18 and 7-19) and were purified on an agarose gel (section 5.7). The insert-specific primers were designed to have 23-25 bases long overhangs, overlapping with the 3'- and 5' ends of the linearized vector, as required for the Gibson assembly. After cleanup, the insert DNA (ca. 0.066 pmol) and the linearized target vector (ca. 0.033 pmol) were added at a molar ratio of 2:1 to a Gibson master-mix aliquot (7.5 μL) with a final reaction volume not exceeding 10-12 μL. The reaction mix was incubated at 50°C for 1 h and was then transformed into competent *E. coli* cells (section 5.3). Positive clones were determined by colony PCR on the next day and were cultured overnight (section 5.6). Plasmid DNA was extracted from the cells according to the manufacturer's protocol using the QIAprep Spin Miniprep Kit (QIAGEN). The sequence integrity was verified by Sanger sequencing. Cloning of eRNA fragments for SHAPE-MaP (section 5.16) was carried out by myself. Cloning of eRNA fragments from the ready SHAPE-MaP vectors into new pUC18 vectors to produce eRNAs without linker sequences for other applications, was carried out by Lisa-Marie Schneider or myself. *Arc* eRNA fragments were cloned from a plasmid (*Arc* eRNA-pBlueScript) that was a kind gift from Prof. Dr. Taekyung Kim (Pohang University of Science and Technology (POSTECH); previously University of Texas Southwestern (UTSW) Medical Center).

An example for the cloned constructs comprising the eRNA insert and flanking sequences of the pUC18 vector are shown in the supplement (section 7.4; Figures 7-1 and 7-2)

## 5.9 *In vitro* production of RNAs

eRNAs used for different applications (EMSA, Transcription assays and for SHAPE-MaP) were produced by T7 RNA polymerase mediated *in vitro* run-off transcription (adapted from (Brunelle & Green, 2013; Cazenave & Uhlenbeck, 1994). PCR generated DNA amplicons or purchased DNA oligos were used as templates. eRNAs for SHAPE-MaP were produced by me, eRNAs for EMSAs were mainly produced by Filiz Kuybu.

### 5.9.1 Production of DNA-Templates

5x Annealing buffer: 100 mM Na-HEPES pH 7.4, 500 mM NaCl, 15 mM MgCl<sub>2</sub>, 50% (v/v) glycerol

First, mouse eRNA sequences were cloned via Gibson assembly (section 5.8) into pUC18 plasmids comprising a T7 promoter sequence (5'-TAATACGACTCACTATAGG) directly upstream of the eRNA sequence (compare supplementary Figures 7-1 and 7-2). To generate linear DNA templates, eRNA sequences were then amplified from the cloned pUC18 plasmids using a universal forward primer overlapping the T7 promoter sequence (primer RP0007) and an appropriate eRNA-specific reverse primer (supplementary Table 7-1). A universal reverse primer (RP0008) was used in the template production for eRNAs used for SHAPE-MaP (section 5.20). 8 – 16 PCR reactions (50 µL each) were set up to produce sufficient amounts of DNA template for a ~1 mL scale *in vitro* transcription. Efficient transcription by T7 RNA polymerase requires two guanosines (marked in bold letters in the T7 promoter sequence) at the end of the promoter sequence. The transcription begins at the first G so that all *in vitro* produced RNAs begin with two Gs at their 5' end. If DNA oligos were used as templates, a template-strand and a fully complementary non-template strand (supplementary Tables 7-16 and 7-17) containing the T7 promoter sequences were annealed to produce a double-stranded DNA. The annealing reaction was set up and incubated in a thermocycler according to the protocol below.

Component	100µL reaction	Step	Temp	Time
Template strand oligo (100 µM)	20 µL	Denaturation	95 °C	5 min
Non-template strand oligo (100 µM)	20 µL	Annealing:	-1°C/min gradient	75 min (cycles)
5x Annealing buffer	20 µL	Hold:	20 °C	
ddH <sub>2</sub> O	40 µL			

### 5.9.2 *In vitro* Transcription (IVT)

10x IVT buffer:	400 mM Tris-HCl (pH 8.1), 50 mM DTT, 10 mM Spermidine, 0.1 % (v/v) Triton X-100
10x HEK buffer (stock):	250 mM K-HEPES (pH 8.0), 1 M KCl, 1 mM EDTA
DEPC-H <sub>2</sub> O:	Diethyl pyrocarbonate (DEPC)-treated water (1 mL DEPC per 1 L ddH <sub>2</sub> O, stirred overnight and autoclaved the next day)

The transcription reaction mix was prepared according to the table below. Depending on the desired yield of eRNA, either 200  $\mu$ L (per eRNA for SHAPE-MaP) or 0.8-1.2 mL (per eRNA for EMSA and transcription assay) reactions were set up and split in 200-250  $\mu$ L per one 1.5-mL reaction tube. Reaction mixtures were incubated at 37 °C for about 3-4 h.

Component	Final amount	200 $\mu$ L reaction
10x IVT buffer	1x	20 $\mu$ L
NTPs (each 20 mM)	4 mM	40 $\mu$ L
MgCl <sub>2</sub> (0.5 M)	20 mM	8 $\mu$ L
PEG8000 (40%)	8 %	40 $\mu$ L
DNA Template	400 nM or 2–4 $\mu$ g	X $\mu$ L
T7-RNA Pol (~5 mg/mL)		4 $\mu$ L
DEPC-H <sub>2</sub> O		ad 200 $\mu$ L

Afterwards, samples were supplemented with 10x DNase I buffer and treated with 2-3  $\mu$ L of DNase I (RNase free; Roche) per 200  $\mu$ L reaction for 30 min at 37 °C to degrade the DNA template. The reactions were stopped by addition of 0.5 M EDTA to a final concentration of 50 mM and were briefly centrifuged to pellet the magnesium pyrophosphate byproduct. The cleared supernatant was transferred into a fresh tube and the RNA was ethanol precipitated overnight (section 5.10), to reduce the volume to be loaded on a preparative Urea-gel (section 5.11). RNA pellets from a 1-mL scale reaction were usually dissolved in ca. 100-200  $\mu$ L of DEPC-H<sub>2</sub>O and the RNA was further purified by a preparative urea-PAGE (section 5.11). Gel purification was critical for the eRNAs, especially when they were used for the transcription assay. Other purification methods did not sufficiently remove free NTPs from the RNA samples or are less convenient for high amounts of RNAs.

Target RNA bands were detected by UV shadowing (using a UV-lamp and a thin-layer chromatography plate) and were excised from the gel. The RNA was eluted passively by the crush and soak method (Petrov et al., 2013). To this end, the excised gel slices were crushed through a 5-10 mL syringe into a 15-mL reaction tube and RNA was eluted with DEPC-H<sub>2</sub>O or 1x HEK buffer at 4 °C overnight on a rotator. Typically, gel slices from a 1-mL scale IVT reaction

were split into four tubes and RNA was eluted with 10 mL of solvent per tube. The next day, the eluate was cleared from remaining gel pieces by filtering through a 0.2 µm syringe filter and was concentrated to a volume of 0.5 – 1 mL using centrifugal filters (Amicon Ultra-4 /-15, Millipore) with a molecular weight cutoff (MWCO) depending on the molecular size of the produced fragment (1-50 nt – 3 kDa; 1-100 nt – 10 kDa; 1-200 nt – 10 kDa or 30 kDa). Subsequently, eRNAs were subjected to size exclusion chromatography in 1x HEK running buffer (the 1x HEK buffer for eRNAs used in SHAPE-MaP contained a higher concentration of KCl, 150 mM instead of 100 mM) on a Superdex 200 Increase 10/300 GL column to purify RNA monomers from aggregates. After elution from the column, the monodisperse peak fractions were pooled and concentrated to an appropriate volume and eRNA concentrations in the range of 10-20 µM.

### 5.10 Ethanol precipitation

Ethanol precipitation was used as a method to purify and concentrate nucleic acids in the presence of monovalent salts. RNA containing samples were supplemented with an appropriate volume of 3 M sodium acetate (NaOAc) (pH 4.8-5.2) to achieve a final concentration of 0.3 M. Following this, 2.5 volumes of ethanol (with respect to the volume of the nucleic acid solution) were added. When precipitating small amounts of RNA, samples were additionally supplemented with 1 µL of glycogen (20 µg/µL) that serves as a carrier to improve the precipitation. Samples were put on dry ice for at least 2-3 h or overnight at -80°C. Precipitated nucleic acids were pelleted by centrifugation at 4 °C and full-speed (15,000 rpm) in a table-top centrifuge for at least 30 min. The pellet was washed once with 75 % (v/v) ethanol (15,000 rpm, 4 °C, 10 min), air-dried for 5-10 min and dissolved in an appropriate volume of DEPC-H<sub>2</sub>O, 1x TE buffer or 1x HEK buffer depending on the further application.

### 5.11 Urea Polyacrylamide gel electrophoresis (PAGE):

6– 15% 1xTBE Urea-gels (for 25 mL):	6 – 15 mL Sequencing gel concentrate, 16.5 – 7.5 mL Sequencing gel diluent, 2.5 mL Sequencing gel buffer concentrate (10x), 200 µL 10 % APS, 10 µL TEMED
10x TBE (stock):	108 g/L Tris base, 55 g/L boric acid, 40 mL/L 0,5 M EDTA (pH 8.0)
20x TTE (stock):	216 g/L Tris base, 72 g/L taurine, 20 mL/L 0,5 M EDTA (pH 8.0)
2x RNA loading dye	95 % formamide (v/v), 18 mM EDTA, 0.025 % SDS, 0.025 % (w/v) Bromophenol blue (BPB), 0.025% (w/v) Xylene cyanol (XC)
1x SYBR Gold staining solution	3 µL SYBR Gold (10,000x stock) in 30 mL 1x TBE

Denaturing urea-PAGE was used in an analytical or preparative manner to separate and purify RNAs according to their size. 1x TBE-urea gels were prepared from ready-to-use solutions (Rotiphorese DNA sequencing system, Roth) according to the manufacturer's protocol (see recipe above for 25 mL gel mix). For the preparation of 0.5x TTE-urea gels, used for transcription assays (section 5.24), the Sequencing gel buffer concentrate was substituted by a custom-made 10xTTE gel buffer concentrate (20xTTE stock supplemented with 6 M urea). Different gel dimensions, electrophoresis apparatus systems, and running conditions were utilized depending on the purpose of Urea PAGE (see listing below).

<b>Electrophoresis apparatus</b>	<b>gel dimensions (width x</b>	<b>gel</b>	<b>Running buffer</b>	<b>runned at constant</b>	<b>Application</b>
Hoefler SE400/410	16 x 18 cm (SE400)	ca. 70 mL	1xTBE	power: 12-20 W	• preparative Urea-PAGE after in vitrhomatogrip-tion
	16 x 24 cm (SE410) thickness: 3 mm	ca.130 mL			
	16 x 18 cm (SE400) thickness: 0.4 mm	L	0.5x TTE	voltage: 21 mA	• Transcription assay (section 5.24)
Bio-Rad Mini-protean® Tetra system	8.6 x 6.7 cm thickness: 1.5 mm		1xTBE	voltage: 200 V	• analytical Urea-PAGE

Analytical gels were polymerized at least 1 h, preparative gels or gels for the transcription assays were polymerized at least overnight. Samples were prepared by combining RNA samples with an equal volume of 2x RNA loading dye and were boiled at 95 °C for 4 min. Samples were put on ice before loading and gels were pre-run in 1x TBE (or 0.5x TTE) for 30-45 min (no pre-run for small analytical gels). After the run, gels were treated differently according to the application (see in corresponding sections). Analytical gels were stained with SYBR Gold (Invitrogen) for 10-20 min, before they were inspected on an UV-light table and documented of the gel documentation system. Typically, the following RNA/ssDNA molecular weight markers were used: Low Molecular Weight Marker 10-100 nt (ssDNA) (Affymetrix), 20/100 Ladder (ssDNA) (IDT) and RiboRuler Low Range RNA Ladder (Thermo Scientific).

## 5.12 Mouse cortical neuron culture

Mouse primary neuron cell culture experiments were the prerequisite for the performed Exo-seq and SHAPE-MaP experiments and were performed by Dr. Katie Schaukowitch and Dr. Seung-Kyoon Kim (UTSW Medical Center, Dallas, USA).

In summary, our collaborators dissected mouse cortical neurons from mouse embryos and cultured them. For the stimulation of neurons, cells were made quiescent with tetrodotoxin (TTX) and were subsequently depolarized by 55 mM KCl for 30-60 min. Dr. Schaukowitch and Dr. Kim kindly provided us already extracted total RNA from untreated (only TTX) or KCl treated (TTX+KCl) cells.

## 5.13 rRNA and 7SL RNA depletion from total RNA

Prior to Exo-seq library preparation (section 5.14), rRNA and signal recognition particle RNA (7SL RNA) were depleted from total RNA using a commercial kit (RiboCop rRNA depletion kit V1.2 (human/mouse/rat), Lexogen) against rRNA and custom made oligos against 7SL RNA. The manufacturer's protocol for the kit was followed with changes described below. Briefly, four biotinylated DNA oligos (supplementary Table 7-9), complementary to the 7SL RNA sequence were utilized. They were pre-mixed with the probe mix (PM) from the kit, which also contains biotinylated DNA oligos but against the rRNA, in order to deplete both RNAs in a one batch reaction.

500 ng of total RNA from unstimulated (untreated) or stimulated (35' KCl treated) neuronal cells and 115  $\mu$ L of magnetic streptavidin beads (75  $\mu$ L depletion beads from the kit + 40  $\mu$ L Dynabeads MyOne Streptavidin C1, ThermoFisher Scientific) were used for one depletion reaction. All buffer volumes from the depletion kit were scaled up to match the relations of the manufacturer's protocol. The hybridization mix was prepared with total RNA (500 ng), 6.2  $\mu$ L hybridization solution (HS), 5  $\mu$ L probe mix (PM) (volume was not scaled up) and 3.5  $\mu$ L (3.5 pmol) anti 7SL RNA oligo mix (1  $\mu$ M mix of the four oligos in a 1:1:1:1 ratio) in a final volume of 54  $\mu$ L. After the first depletion step, the depleted sample was incubated a second time with magnetic beads to ensure that all depletion oligos were pulled out. For this, the supernatant containing the rRNA- and SRP-RNA depleted RNA was transferred into a fresh reaction tube and supplemented with 30  $\mu$ L of pre-conditioned magnetic streptavidin beads.

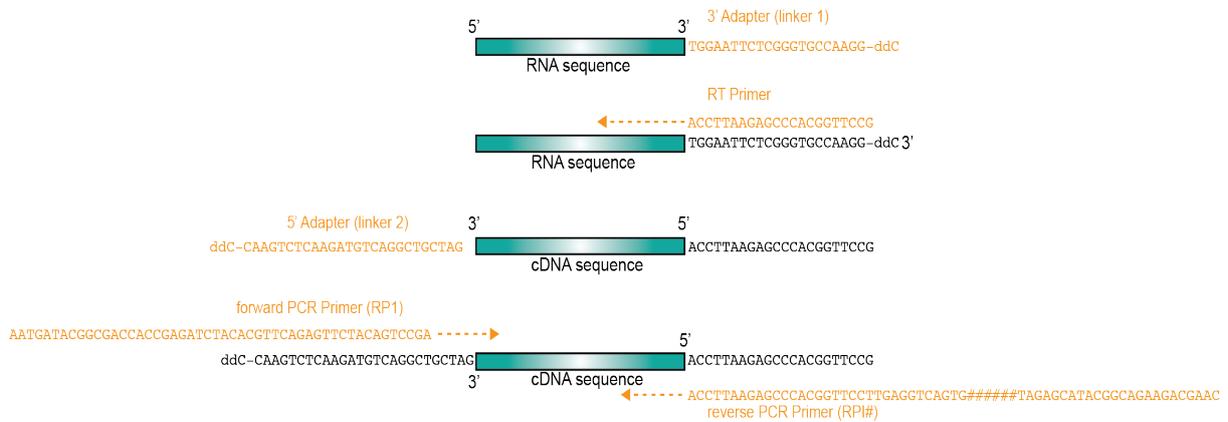
The mix was incubated at room temperature for 10 min and subsequently at 52 °C for 10 min in a thermomixer. The cleared supernatant was transferred into a fresh reaction tube and applied to spin-columns (RNA Clean & Concentrator-5 columns, Zymo-Research) for a cleanup according to the manufacturer's instructions. For one Exo-seq library sample, RNA from three depletion reactions (3x 500 ng total RNA) was pooled prior to column purification.

The necessity for the depletion of 7SL RNA arose through the results of the first Exo-seq sequencing run (Sequencing Run #1). In the initial experiment only rRNA was depleted, using another rRNA depletion kit (Ribo-Zero rRNA Removal Kit, Illumina). About one-third of the reads originated from the 7SL RNA. Another remark concerning the input amount of total RNA for the rRNA depletion when using commercial kits: it is recommended to use about 2-times less amount of input than the upper limit specification in the manufacturer's protocol to deplete the rRNA efficiently.

#### **5.14 Exo-seq library preparation**

Exo-seq libraries were prepared as previously described (Afik et al., 2017) with some modifications. The full protocol comprising each step and the amounts of components can be found in the supplement of the Exo-seq publication. In this study, rRNA and 7SL RNA depleted RNA produced from 1.5 µg of total RNA (section 5.13) was used as starting material instead of poly (A)-selected RNA. Furthermore, other adapters (linker) and PCR primers were used, based on the TruSeq Small RNA sample Kit (Illumina) (supplementary Table 7-10). The adapters feature a phosphate modification at the 5' end and a dideoxycytosine (ddC) at the 3' end that prevents self-ligation of adapter molecules. The reverse transcription step was carried out with SuperScript III (Invitrogen). The library was cleaned up and size selected in two consecutive rounds of binding to 1.2x and 0.8x SPRI beads (Agentcourt AMPure XP beads, Beckman Coulter) according to the manufacturer's instructions. The final libraries, two biological replicates of the KCl treated condition (KCl 1 and KCl 2) and one replicate of the untreated condition, were checked for integrity and size distribution on a Bioanalyzer before they were sequenced in a 75 nt, single-end, NextSeq High-Output mode on a NextSeq 500 platform. The final steps before sequencing (e.g., final Bioanalyzer run, pooling of libraries, and loading of the flow cell) and the final sequencing run was performed at the Core Unit Systems Medicine (University of Würzburg) by members of staff. For a quick overview, a

scheme summarizing the adapter and PCR primer sequences added to the RNA fragment at each step of the library preparation is depicted below.



## 5.15 Processing of 5' Exo-seq reads

The following analysis was carried out by Andreas Pittroff under my supervision. Raw reads were processed using *Cutadapt* (Martin, 2011) for adapter trimming (of the 3' adapter sequence TGG AATTCTCGGGTGCCAAGG) and retaining all reads with a minimum read length of 20 nts after trimming. Remaining rRNA- and tRNA reads were filtered out, using *SortMeRNA* (Kopylova et al., 2012) and *Bowtie* (Langmead et al., 2009) respectively. rRNA and tRNA reference files were obtained from the UCSC table browser (Haeussler et al., 2018; Karolchik et al., 2004). Processed reads were then aligned to the mouse genome (mm10, December 2009) using *STAR* (Dobin et al., 2013), allowing two mismatches and filtering out multimappers. 5'-end coverage of mapped reads was calculated for both strands separately using bedtools' *genomecov* utility (Quinlan & Hall, 2010) with `-bg` and `-5` parameters. TSScall (Henriques et al., 2018) was then run on the generated bedgraph files with standard parameters for identification of transcription start sites (TSS). We defined extragenic TSSs as TSSs that do not occur within a RefSeq annotated gene  $\pm 2$  kb. By using these criteria 129,161 extragenic TSSs were identified for Replicate 1 (KCl-treated #1) and 131,312 for Replicate 2 (KCl-treated #2). Called TSSs were then overlapped with *de novo* GRO-seq defined enhancer transcript units allowing an offset of  $\pm 200$  nt (GRO-seq based enhancer list was provided by Dr. Seung-Kyoon Kim, see section 3.1.2). A single TSS was then selected for each identified enhancer locus. This selection was based on read coverage and distance to the 5'-end of the respective GRO-seq transcript. This resulted in 977 TSSs for Replicate 1 and 1,039 for Replicate

2. Activity-induced enhancers were defined based on the GRO-seq data fold-change ( $FC > 1.5$ ) between stimulated (KCl-treated) and unstimulated (untreated) conditions. In a final step, 39 high quality eRNA TSSs were selected for structure mapping. Among these 33 eRNA TSSs were derived from activity-induced eRNAs, as defined by GRO-seq data (Gorbovytska et al., 2021). The metagene plot shown in Figure 3-3A was generated by myself using the *deepTools* (Ramírez et al., 2016) `computeMatrix` reference-point utility with `--referencePoint TSS -b 500 -a 1000` parameters, bigWig files of the filtered Exo-seq reads 5' end coverage (control and KCl-treated samples) as input and RefSeq genes (for mm10 mouse genome, obtained from UCSC table browser) .bed file as reference. The generated matrix was used as input for the `plotProfile` utility of *deepTools*.

## 5.16 RNA structure probing by SHAPE-MaP

### 5.16.1 In vitro SHAPE-MaP

10x HEMK buffer (stock):                    500 mM K-HEPES pH 7.0, 1 mM Na-EDTA, 1.5 M KCl, and 150 mM MgCl<sub>2</sub>

To map the secondary structures of 39 eRNAs, we chose SHAPE-MaP as a method for chemical probing of secondary structure, which is combined with a read out by next-generation sequencing. The eRNAs (1-200 nt fragments) used for SHAPE-MaP were produced by *in vitro* transcription (section 5.9). The eRNAs were flanked with a 20 nt-long 5' linker sequence (5'-GGC CAT CTT CGG ATG GCC AA) and 43 nt-long 3' linker sequence (5'-TCG ATC CGG TTC GCC GGA TCC AAA TCG GGC TTC GGT CCG GTT C), based on previous protocols (Merino et al., 2005). In front of chemical probing, each purified eRNA (10 pmol in 18  $\mu$ L) was incubated for folding in 1x HEMK buffer at 37 °C for 30 min (adapted from (F. Liu et al., 2017)). The chemical probing and library preparation according to the small RNA workflow were adapted from (Smola, Rice, et al., 2015). The folded RNA sample was split into two, and each sample was treated either with 1  $\mu$ L of pure DMSO (DMSO control sample) or 1  $\mu$ L of 100 mM 1-methyl-7-nitroisatoic anhydride (1M7 modified sample) (final concentration of 10 mM 1M7) at 37 °C for 5 min. The treatment was repeated for a second round of modification to achieve higher modification rates. The sample volume was adjusted to 30  $\mu$ L with DEPC-H<sub>2</sub>O before the RNA was cleaned up with 1.8x AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions and eluted in 20  $\mu$ L of DEPC-H<sub>2</sub>O. 10  $\mu$ L of the modified RNA

sample were subjected to reverse transcription with SuperScript II enzyme (Invitrogen) in the presence of  $Mn^{2+}$  ions, as described (Smola, Rice, et al., 2015). A specific SHAPE-MaP RT primer was used, complementary to the 3' linker region. For eRNAs, which did not yield any cDNA product under the normal conditions, the temperature during reverse transcription was increased to 50 °C. The cDNA was cleaned up with 1.8x AMPure XP beads and eluted with 35  $\mu$ L DEPC- $H_2O$ . The subsequent first and second PCR were performed as described (Smola, Rice, et al., 2015). For the DMSO control and the 1M7 modified libraries two different Index-primers were used. The DNA from the first PCR was purified with 1.0x AMPure XP beads and from the second PCR with 0.8x. Final libraries were eluted with 25  $\mu$ L. All primers used for SHAPE-MaP are listed in the supplementary Table 7-11. Each sample was checked on the Fragment Analyzer (Agilent), and concentrations were fluorometrically determined on a Qubit (ThermoFisher Scientific). These steps were carried out by Michaela Hochholzer and Andrea Kripal at the Keylab Genomics & Bioinformatics of PD Dr. Alfons Weig (University of Bayreuth). All individual samples (39 DMSO and 39 1M7 samples) were diluted with library dilution buffer (10 mM Tris-HCl pH 8.0, 0,1% (v/v) Tween 20) to 5 nM and pooled equimolar. Libraries were sequenced on a NextSeq 500 platform with a Mid-Output Next-Seq kit in a 150 bp, paired-end mode, and ca. 30% PhiX due to the low complexity of the final library mix. The sequencing and data acquisition were performed at the Core Unit Systems Medicine (University of Würzburg) by Dr. Kristina Döring.

### 5.16.2 *In vivo* SHAPE-MaP

The treatment of primary cortical neurons with DMSO (control) and 1M7 was conducted by Dr. Seung-Kyoon Kim. Briefly, primary mouse cortical neurons were treated for 20 min with 1M7 (10 mM; added to the medium) during a 30 min KCl induction. Then nuclei were isolated and divided into two. One set (two 10 cm plates amount, 12 million neurons/plate) was directly lysed by Trizol ("Nuclei" sample). The other set was subjected to nuclear run-on for 20 min with 1 mM 1M7 and then lysed with Trizol ("Run-on" samples).

Samples were forwarded to our laboratory and from here on I produced the SHAPE-MaP libraries as described for *in vitro* SHAPE-MaP with some alternations. RNA was extracted from "Nuclei" and "Run-on" samples according to the manufacturer's protocol for Trizol and samples were additionally DNase I treated. Four different eRNAs from our set of 39 eRNAs

were selected for trial libraries to test SHAPE-MaP on an extended eRNA region 1-400 nt. eRNA-specific reverse transcription (RT) primer and PCR primer were designed to amplify two overlapping regions (1-250 and 150-400 nt) (supplementary Table 7-12). For each target eRNA fragment 4 µg of extracted RNA were subjected to reverse transcription and half of the cDNA (~2 µg) were used for the first PCR.

### 5.17 Analysis of SHAPE-MaP data

SHAPE-MaP sequencing reads (.fastq files) were analyzed by the ShapeMapper 2 tool (Busan & Weeks, 2017). The software was executed with default parameters. It aligns the reads to a given eRNA reference sequences file (.fasta file of eRNA sequences found in supplementary Table ) using Bowtie 2 (default option) and calculates the read-depth per each nucleotide for every eRNA along with the mutation rates and SHAPE reactivities. eRNA secondary structures were predicted with RNAstructure (Reuter & Mathews, 2010) using the SHAPE reactivities per nucleotide, from ShapeMapper 2, as constraints. The MaxExpect (maximum expected accuracy) structure (Reuter & Mathews, 2010), was selected for visualization of the secondary structure using the StructureEditor, that belongs to the RNAstructure software package.

The median SHAPE reactivity (Figure 3-8A) for each eRNA was calculated from SHAPE reactivities (using the .shape files generated by SHAPMapper 2). The minimum free energy for the lowest energy structure of each eRNA, which was plotted in Figure 3-8B, was retrieved from .efn2 (energy function 2) files, which were generated from .ct files in RNAstructure (J. Zuber & Mathews, 2019). The mutation rates and read depth for each eRNA in the SHAPE-MaP experiment were retrieved from histogram.pdf and from .map files, which are generated files of the ShapeMapper 2. The boxplots shown in Figures 3-6, 3-7 and 3-8 were generated in Rstudio (Version 1.3.1056). Testing the DMSO and 1M7 mutation rates for significant difference was performed using the paired Mann-Whitney-U test (wilcox.test in R).

The boxplot shown in Figure 3-15E was produced by Andreas Pittroff. Sequences of our 39 eRNAs set were extracted from position 1 to 1000 nucleotides using gffread (Pertea & Pertea, 2020). Sequences were then binned in non-overlapping bins of 200 nucleotides and analyzed for their sequence content. Differences in base content were tested for significance with a pairwise t-test using a custom R script (written by Andreas Pittroff).

## 5.18 Purification of endogenous Pol II

0 M HepR buffer:	50 mM Tris (pH 7.6 at room temperature; pH 7.9 at 4°C), 1 mM EDTA, 10 $\mu$ M ZnCl <sub>2</sub> , 10% (v/v) glycerol (prepare 7 L for one purification)
0.6 M HepR buffer:	50 mM Tris (pH 7.6 at room temperature; pH 7.9 at 4°C), 1 mM EDTA, 10 $\mu$ M ZnCl <sub>2</sub> , 10% (v/v) glycerol, 0.6 M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> (prepare 2.5 L for one purification)

Mammalian *Sus scrofa* Pol II was purified as previously described (Bernecky et al., 2016, 2017) with minor alterations, using pig thymus collected from the Bayerische Landesanstalt für Landwirtschaft (LfL) in Poing, Germany (provided by Dr. Kunz and Marcel Bowens). It was critical to freeze the thymus in liquid nitrogen (kindly carried out by Claus-D. Kuhn) immediately after the extraction from the pig body. 500-600 g of thymus (stored at -80°C) were used for one round of purification. All ammonium sulfate ((NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>)-containing buffers for the purification were prepared from 0 M and 0.6 M HepR buffer stocks, which were pre-cooled overnight to 4°C before adjusting the pH to 7.9. Unless otherwise noted, all steps were conducted at 4 °C, in the cold room, or on ice. All buffers and beakers were pre-cooled to 4°C before usage. The prepared buffers were supplemented with the appropriate proteinase inhibitor mix, as described (Bernecky et al., 2016, 2017), just before their usage.

The purification protocol comprises homogenization of the thymus, polyethyleneimine (PEI) precipitation of nucleic acids with bound Pol II, purification over a MacroPrepQ ion-exchange column, ammonium sulfate precipitation, antibody (8WG16) column against RPB1 CTD and eventually purification over an UnoQ column. In the following, only alternations to the original protocol or more detailed descriptions are specified. The frozen thymus was broken down to small pieces using a meat mallet and was homogenized in 2x 750 mL (per 2x 250 g thymus) of 0 M HepR buffer in a 2 L blender (Waring). All centrifugation steps before the MacroPrepQ column were carried out at 10,000 rpm (~ 17,700 g) and 4°C for 30 min in an F10-6x500y fixed angle rotor (Thermo Scientific, Sorvall). PEI pellets were washed once with 0 M HepR buffer (1.2 L) and kept with some buffer on top in the cold room overnight before they were resuspended on the next day in 0.15 M HepR buffer (700 mL) for loading on a self-packed MacroPrepQ anion-exchange column (1 column volume (CV)= ~200 mL).

After loading the cleared input on the column, the column was washed with 3 CVs (600 mL) of 0.2 M HepR and bound proteins were eluted with 0.4 M HepR buffer (3 CV). Following the absorbance at 280 nm, a single elution fraction (ca. 200-300 mL) comprising the main protein

peak was collected in a separate beaker and proteins were precipitated by ammonium sulfate (50% saturation), which was slowly added to the protein solution while stirring. The amount of finely grounded ammonium sulfate to be added was calculated using the Encor calculator tool (<https://www.encorbio.com/protocols/AM-SO4.htm>). After stirring for 1 h, the ammonium sulfate pellets were obtained by centrifugation at 15,000 rpm (~27,000 g) and 4°C for 30 min in an SS-34 fixed angle rotor (Thermo Scientific, Sorvall). The pellets were stored overnight with some buffer left on top and were resuspended on the next day with 30 mL of 0 M HepR buffer. After adjusting of the sample's conductivity to fit 0.15 M HepR, the sample was purified over an 8WG16 ( $\alpha$ RPB1 CTD) antibody-couplechromatogse column (~3 mL CV). The last size exclusion chromatography (SEC) step was omitted. Instead, Pol II peak fractions, eluted from the Uno Q-1 anion exchange column (Bio-rad), were pooled and the buffer was exchanged to the final Pol 2 buffer (25 mM Na-HEPES pH 7.5, 150 mM NaCl, 10  $\mu$ M ZnCl<sub>2</sub>, 10% (v/v) glycerol, 2 mM DTT) by diafiltration using 100-kDa MWCO centrifugal filters (Amicon Ultra-4, Millipore). Pol II was concentrated up to 2.6 mg/mL (5  $\mu$ M), aliquoted, flash-frozen in liquid nitrogen and stored at -80°C. It is very critical to keep Pol II in glycerol (10% (v/v))-containing buffer. Exchanging the buffer to a glycerol-free one renders it fully enzymatically inactive.

### 5.19 Expression of NELF variants and DSIF in insect cells

The recombinant expression and purification of WT NELF and DSIF were carried out by me, Silke Spudeit, or Dr. Felix Klatt. The NELF $\Delta$ RRM variant was expressed and purified by Lisa-Marie Schneider, and Dr. Felix Klatt carried out the expression and purification of the NELF patch mutant and double mutant. The expression plasmids of all NELF variants were a kind gift from Dr. Seychelle Vos and Prof. Dr. Patrick Cramer (Max-Planck-Institute for Biophysical Chemistry, Göttingen). The expression plasmid for DSIF was kindly given by Prof. Dr. Birgitta Wöhrl (Chair of Biopolymers, University of Bayreuth).

The protein expression in insect cells was performed using the MultiBacTurbo system (Berger et al., 2004; Fitzgerald et al., 2006) with an adapted protocol described by Dr. Felix Klatt (Klatt, 2020). Briefly, bacmids for protein expression were made in DH10EMBaCY cells and isolated as described. For the V0 baculovirus generation, bacmids were transfected into Sf21 insect cells in SF-4 Baculo Express medium (Bioconcept) at a density of  $0.8 \times 10^6$  cells/mL.

YFP fluorescence was monitored as a proxy for transfection efficiency and cells were kept at a density of  $0.8 \times 10^6$  cells/mL until 100% of cells showed YFP fluorescence. The supernatant, containing the V0 virus, was subsequently isolated and utilized to infect High Five cells at a 1:10 through 1:20 ratio. For protein expression High Five cells were kept at a density of  $1.0$ - $1.2 \times 10^6$  cells/mL. Protein expression was allowed to proceed for 48 h before cells were harvested by centrifugation in a Fiberlite F10-4x1000 LEX rotor (15 min, 700 g and room temperature). Cell pellets were resuspended in 60 mL (per 1 L expression culture) lysis buffer (for NELF: 20 mM Na-HEPES pH 7.4, 300 mM NaCl, 10% (v/v) glycerol, 30 mM imidazole, 2 mM DTT; for DSIF: 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM DTT) supplemented with the protease inhibitors pepstatin A (1  $\mu$ g/mL), leupeptin (1  $\mu$ g/mL) benzamidine hydrochloride (0.2 mM) and phenylmethylsulfonyl fluoride (PMSF) (0.2 mM). Cells were directly used for protein purification or flash frozen in liquid nitrogen and stored at  $-80$  °C.

## 5.20 Purification of NELF variants

The purification protocol for the WT NELF and all NELF variants (NELF $\Delta$ RRM, NELF patch mutant and NELF double (patch +  $\Delta$ RRM) mutant) were adapted by Lisa-Marie Schneider and Jonathan V. Patzke (Master's thesis Scheider, 2017; Patzke, 2020) from a prior protocol (Vos et al., 2016). Briefly, cells were lysed by sonication and lysates were cleared by ultracentrifugation at 33,000 rpm and  $4$  °C for 90 min in a Type 45 Ti rotor (Beckman coulter). The supernatant was purified by Ni-NTA gravity-flow affinity chromatography, the NELF containing protein fractions were pooled, diluted to match 300 mM NaCl, filtered and loaded onto a pre-equilibrated (20 mM Na-HEPES pH 7.4, 300 mM NaCl, 10 % (v/v) glycerol, 2 mM DTT) Resource Q column (6 mL CV; GE Healthcare). As the 4-subunit NELF complex does not bind to the column, the flow through was collected, diluted to 150 mM NaCl, reloaded onto the Resource Q column, and eluted with a linear salt gradient to 1 M NaCl (the first Resource Q step allows for the removal of trimeric NELF-ABD complexes). Peak fractions were pooled, concentrated using 30 kDa Amicon concentrators (Millipore) and loaded onto a Superose 6 10/300 column (GE Healthcare) that had been equilibrated in SEC buffer (20 mM Na-HEPES pH 7.4, 150 mM NaCl, 2 mM DTT). Peak fractions after SEC were pooled and concentrated thromatox. 30  $\mu$ M ( $\sim$ 7.0 mg/mL). Finally, glycerol was added to 10 % (v/v) before the protein was aliquoted, flash-frozen in liquid nitrogen and stored at  $-80$ °C.

## 5.21 Purification of DSIF

DSIF carried a C-terminal 2xStrep-tag on SPT5. Cells were lysed by sonication and lysates were then cleared as described for the purification of NELF. Subsequently, the supernatant was applied to a gravity-flow column with 1 mL Strep-Tactin beads (IBA) that had been equilibrated in binding buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM DTT). Beads were then washed with the same buffer (40 CV) and the protein was eluted with 20 mL (1 mL fractions) elution buffer (binding buffer supplemented with 2.5 mM desthiobiotin). DSIF-containing fractions were pooled and applied to a Resource Q column. DSIF was eluted with a linear salt gradient from 150 mM to 1 M NaCl. Peak fractions were pooled, concentrated using 30 kDa Amicon concentrators (Millipore) and applied to a Superose 6 10/300 column in SEC buffer (20 mM Na-HEPES pH 7.4, 150 mM NaCl, 2 mM DTT). Peak fractions were pooled and concentrated to 7.5  $\mu$ M (~1.1 mg/mL). Glycerol was added to 10 % (v/v) before the protein was aliquoted, flash frozen in liquid nitrogen and stored at -80 °C.

## 5.22 SDS-PAGE

6x SDS sample buffer	375 mM Tris-HCl pH 6.8, 6% SDS, 48% glycerol, 0.025% (w/v) bromophenol blue (BPB)
Coomassie staining solution	0.1 % (w/v) Coomassie Brilliant blue R-250, 50 % (v/v) methanol, 10 % (v/v) glacial acetic acid
Destaining solution	40 % (v/v) methanol, 10 % (v/v) glacial acetic acid

SDS polyacrylamide gel electrophoresis (SDS-PAGE) was performed to separate proteins according to their size. Discontinuous Bis-Tris gels or Tris-Glycine gels, comprising a resolving gel (typically 8-10%) and a short stacking gel (4%) were used. 0.75 mm gels were cast by hand using the Mini-Protean Tetra System (Bio-Rad) and run at 155 V for 45-60 min. Protein samples were prepared by combining 12  $\mu$ L of protein sample with 4  $\mu$ L of 6x SDS sample buffer containing DTT (3  $\mu$ L 6x buffer + 1  $\mu$ L 2M DTT) and were boiled for 5 min at 95 °C before loaded on a gel. Precision Plus Protein Standard (Bio-Rad) was used to assign bands. Gels were

stained for 5-10 min in Coomassie staining solution and subsequently destained in destain solution.

### 5.23 Electrophoretic mobility shift assays (EMSAs)

5x Annealing buffer:	100 mM Na-HEPES pH 7.4, 500 mM NaCl, 15 mM MgCl <sub>2</sub> , 50% (v/v) glycerol
5x EMSA buffer:	100 mM Na-HEPES pH 7.4, 500 mM NaCl, 125 mM KCl, 15 mM MgCl <sub>2</sub>
EMSA loading dye:	20 mM Na-HEPES pH 7.4, 60% (v/v) glycerol, 0.025% (w/v) Bromophenol blue
3.5 % native gel (0.5x TBE) (for 50 mL)	4.4 mL 40% Acrylamide/Bis-acrylamide (ROTIPHORESE Gel 40 (29:1)), 2.5 mL 10xTBE buffer, 42.7 mL H <sub>2</sub> O, 400 µL 10% APS, 20 µL TEMED

EMSA Experiments were designed by me and mainly performed by Filiz Kuybu. The PEC was formed on a nucleic acid bubble scaffold using a synthetic 25 nt RNA and a 49 nt template and non-template DNA oligo (supplementary Table 7-13). The sequence of the shorter 15 nt RNA, used in control experiments, matched the 15 nucleotides from the 3' end of the longer 25 nt RNA. Prior to the assay, 100 pmol of RNA were 5'-end labeled with [ $\gamma$ -<sup>32</sup>P]-ATP (3000 Ci/mmol, Perkin Elmer) in a 20 µL reaction using T4 polynucleotide kinase (NEB) according to the manufacturer's protocol. The labeled RNA sample was filled up to 100 µL with DEPC-H<sub>2</sub>O and purified by ethanol precipitation in the presence of glycogen (section 5.10). Subsequently, the RNA pellet was dissolved in 15 µL DEPC-H<sub>2</sub>O and annealed to 100 pmol of template DNA in annealing buffer in a final volume of 20 µL (see table below).

Component	Amount	20 µL reaction
5'-end labled RNA	100 pmol	15 µL
Template DNA (T-DNA) (100 µM)	100 pmol	1 µL
5x Annealing buffer		4 µL

To this, the mix was heated at 95°C for 5 min and cooling down to 20°C in 1°C /min steps in a thermocycler, as previously described (Vos et al., 2018). After annealing, the sample was diluted 1:1 with DEPC-H<sub>2</sub>O to achieve a final concentration of 2.5 nM for the RNA-DNA hybrid (in 40 µL). All amounts in the following refer to one individual EMSA sample used to load in one gel lane. EMSA samples were typically prepared in a *n*-fold master mix, which was split into *n* samples before addition of different eRNAs. Each incubation step, if not explicitly mentioned, was performed at 30 °C for 15 min. The final sample volume was 8 µL (see Table

below for exact volumes of sample components). The PEC was formed by incubating pre-annealed RNA-template DNA hybrid (0.8 pmol) with Pol 2 (1.2 pmol). Then non-template DNA (1.6 pmol) was added and incubated with the mix. Subsequently, 5x EMSA buffer (supplemented with DTT) and DEPC-H<sub>2</sub>O was added to achieve a final concentration of 1x (2 mM DTT). Eventually, DSIF (2.4 pmol) and then NELF (1.2 pmol) were added with each protein incubating with the reaction mix as described above. Finally, the eRNA (1 μL of pre-diluted eRNA) was added with increasing amounts (1.2, 2.4, 4.8, 7.2, 9.6, 14.4, 19.2 pmol) to yield final concentrations of (0.15, 0.3, 0.6, 0.9, 1.2, 1.8, 2.4 μM).

Component	Final amount	Volume (for 8 μL sample (1x))
RNA-DNA hybrid (0.5 μM)	0.8 pmol	1.6 μL
Pol II (5 μM)	1.2 pmol	0.24 μL
NT-DNA (5 μM)	1.6 pmol	0.32 μL
5x EMSA buffer	1x	1.6 μL
DEPC-H <sub>2</sub> O		2.72 μL
DSIF (7.5 μM)	2.4 pmol	0.32 μL
NELF (6 μM)	1.2 pmol	0.2 μL
eRNA	X pmol	1 μL

If an EMSA gel contains a series of six instead of seven different eRNA concentrations, then the highest concentration (2.4 μM) was omitted. The samples were incubated for 15-20 min at room temperature, subsequently supplemented with 1.5 μL of EMSA loading dye and loaded on a pre-chilled vertical 3.5% native acrylamide gel (0.5x TBE) (small 1.5 mm thick gel). The gel was pre-run for 30 min at 90 V and 4 °C (in pre-chilled 0.5x TBE running buffer; gel running chamber was kept on ice during electrophoresis). Samples were electrophoresed under the described conditions for 1.5 h. Subsequently, the gel was dried and exposed on storage phosphor screen for 3 h or overnight, depending on the strength of radioactive signal. The signal read-out was carried out on a CR 35 image plate reader (Elysia-Raytest). Gels were analyzed by densitometry using ImageJ software and the Pol II-DSIF fraction was plotted against the eRNA concentration with Prism 9.

$$(PolII - DSIF)fraction = \frac{(PolII - DSIF)}{(PolII - DSIF) + (PolII - DSIF - NELF)}$$

The produced pseudo binding curves were fitted with a single site quadratic binding equation as previously described (Vos et al., 2016). The obtained apparent K<sub>d</sub> values were used as an

estimation to compare the different eRNA condition and should not be seen as real dissociation constants.

The supershift assay (Figure 3-12E) was performed under the standard EMSA conditions with a final sample volume of 10  $\mu$ L. The final concentration of Arc eRNA 1-200 nt in the eRNA containing samples was 1  $\mu$ M. The NELF antibody (anti-NELF-E, #ab170104, abcam) and the DSIF antibody (anti-SPT5, #sc-133217X, Santa Cruz Biotech) were used with the final concentrations of 0.26  $\mu$ M (2  $\mu$ L of 1.3  $\mu$ M was added) and 0.48  $\mu$ M (2  $\mu$ L of 2.4  $\mu$ M), respectively. This corresponds to a ratio of about 2:1 (for NELF) and 4:1 (for DSIF) of the antibody to protein per sample. The NELF titration experiment (Figure 3-17D) using the four different NELF variants (WT, NELF $\Delta$ RRM, NELF patch mutant, NELF double mutant) was performed under the standard EMSA conditions. NELF variants were added with the following amounts: 0.24, 0.4, 0.8, 1.0, 1.2, 1.6, 2.4 pmol (final concentrations: 0.03, 0.05, 0.1, 0.125, 0.15, 0.2, 0.3  $\mu$ M).

## 5.24 Transcription assay on magnetic beads

5x Transcription buffer:	100 mM Na-HEPES pH 7.4, 750 mM NaCl, 15 mM MgCl <sub>2</sub> , 50 $\mu$ M ZnCl <sub>2</sub> , 20% (v/v) glycerol
5x BW buffer	100 mM Na-HEPES pH 7.5, 100 mM NaCl, 20% (v/v) glycerol, 0.2% (v/v) Tween20, 0.1% (v/v) IGEPAL CA-630
2x Stop buffer	6.4 M urea, 50 mM EDTA pH 8.0, 1x TTE buffer, 0.025% (w/v) Bromophenol blue (BPB)

Transcription assays were adapted from Vos et al., 2018 and performed with a fully complementary scaffold, similar to the nucleic acid scaffold used for EMSA experiments. While the RNA (25 nt) was the same as for EMSA experiments, the utilized template DNA (76 nt) and non-template DNA (80 nt) (supplementary Table 7-13) were longer to allow for a longer extension of the RNA. The non-template DNA had a 4 nt long overhang at the 5' end, which carried a biotin-tag to enable binding to magnetic streptavidin beads. The scaffold contains a 9-base pair (bp) DNA-RNA hybrid, 16 nts of exiting RNA bearing a 5'-<sup>32</sup>P label, 17 nts of upstream DNA and 50 nts of downstream DNA. Radioactively labeled RNA-template DNA hybrid was prepared as described for the EMSA experiment. Samples were generally prepared in a *n*-fold master mix (usually a 14x master mix was prepared, which was split into six 2.33x master mixes, from which each master mix was used to produce one time-course transcription experiment). All amounts in the following are related to a 1x mix.

The assembly of a transcription competent Pol II complex from was carried out analogous to the EMSA experiment, except for slightly altered amounts of the components (see table below). After incubation with non-template DNA, the reaction was supplemented with 5x transcription buffer (+DTT) to achieve a concentration of 1x (+2 mM DTT) in a final volume of 5  $\mu$ L. Addition of the ATP/CTP/UTP (HTP)-nucleotide mix and incubation at 30°C for 10 min, allowed Pol II to transcribe 4 nucleotides of the implemented G-less cassette before it stalled at +4 position because of GTP deprivation. Subsequently, DSIF and then NELF were added, and each time incubated at 30 °C for 15 min (as described in section 5.23).

Component	Final amount	Volume (for 1x reaction)
RNA-DNA hybrid (2.5 $\mu$ M)	2.5 pmol	1 $\mu$ L
Pol II (5 $\mu$ M)	3.75 pmol	0.75 $\mu$ L
NT-DNA (10 $\mu$ M)	5 pmol	0.5 $\mu$ L
5x Transcription buffer (+ DTT)	1x (+2 mM DTT)	1 $\mu$ L
DEPC-H <sub>2</sub> O		0.75 $\mu$ L
HTPs (100 $\mu$ M)	100 pmol	1 $\mu$ L
Mix and let	10 min at 30°C	
Transcribe until pause:		
DSIF (7.5 $\mu$ M)	7.5 pmol	1 $\mu$ L
NELF (30 $\mu$ M)	7.5 pmol	0.25 $\mu$ L

An *Input* control sample was taken before proceeding to the bead binding step. The master mix (14x) of the assembled PEC complex was then diluted with 1x Transcription buffer to a final volume of 180  $\mu$ L (1.5x of initial bead volume) and was bound to magnetic streptavidin beads (Dynabeads MyOne Streptavidin C1, ThermoFisher Scientific) (120  $\mu$ L beads per 14x master mix). Prior to this, beads were conditioned by washing (3x; first time with 1 mL, else with 120  $\mu$ L) with 1x BW buffer (supplemented with 2 mM DTT) and finally resuspended in 120  $\mu$ L of 1x Transcription buffer. The binding mix (300  $\mu$ L total volume) was incubated at room temperature on a tube rotator for 20-30 min. After taking off the supernatant, beads with the bound PEC complexes (usually 1/3 of initial RNA-DNA hybrid was bound, judged by the radioactivity ratio between supernatant and beads) were washed three times with 1x BW buffer (first time with 300  $\mu$ L, else with 120  $\mu$ L) and split into six samples (2.33x of original master mix). The BW buffer was taken off, and the beads were resuspended in 60  $\mu$ L of 5  $\mu$ M eRNA sample in 0.5x HEK buffer (10 mM Na-HEPES pH 7.4, 50 mM KCl, 0.5 mM EDTA) or only the buffer (for buffer control sample). The mix was incubated on a tube rotator for 15 min.

Subsequently, the supernatant was taken off and beads were resuspended in 24  $\mu\text{L}$  of 1x Transcription buffer (+ 2 mM DTT). The *0 min sample* (4.3  $\mu\text{L}$ ) was taken and added to pre-aliquoted 2x Stop buffer (5  $\mu\text{L}$ ), before the transcription on beads was resumed by addition of NTPs (3  $\mu\text{L}$  of 100  $\mu\text{M}$ ). Upon NTP-addition, the bead mix was immediately put back into a thermomixer and incubated at 30 °C and 900 rpm to avoid the sedimentation of beads. Time course samples (4.9  $\mu\text{L}$ ) were taken after 1, 3, 6, 14 and 25 min, and were immediately quenched with 2x Stop Buffer (5  $\mu\text{L}$ ). Multiple time course experiments were processed in parallel in a phased manner (10 min phasing). Finally, the samples were first proteinase K (1  $\mu\text{L}$ , 5  $\mu\text{g}$ ) treated at 37 °C for 30 min and then boiled at 90 °C for 4 min to release the bound molecules from the beads. The samples were then separated on a pre-ran (30 min at 400 V) denaturing 15 % Urea-gel (0.5xTTE, see section 5.11) for in 3 h at 500 V (first 30 min at 400 V). Usually, 4-6  $\mu\text{L}$  of the sample were loaded depending on the amount of radioactivity. After the gel run, gels were dried and exposed overnight as described for the EMSA experiments. For a quantification of the pause release by the eRNAs, the first band above the triple pausing band (as highlighted on the gels in the results part) was analyzed by densitometry using Image J. The single band intensity was normalized against the total band intensity of the corresponding lane and plotted against the time (min).

## 5.25 Protein-RNA Crosslinking coupled to mass spectrometry

### 5.25.1 Protein-RNA crosslinking and sample preparation

10x Complex buffer

200 mM Na-HEPES pH 7.4, 1 M NaCl, 30 mM  $\text{MgCl}_2$

The PEC was prepared as described for the EMSA (section 5.23), except that the nascent RNA was not labeled, preparative amounts and slightly different ratios of components were used (RNA-DNA hybrid was used in a 2fold excess over Pol II to saturate all Pol II molecules). For hybridization of nascent RNA and template DNA higher concentration were used than described above. 25  $\mu\text{L}$  of 100  $\mu\text{M}$  RNA were annealed with 25  $\mu\text{L}$  of 100  $\mu\text{M}$  template DNA in a total volume of 100  $\mu\text{L}$  (final RNA-DNA hybrid concentration: 25  $\mu\text{M}$ ). The reaction set-up for the PEC formation with *Nr4a1*-(b) 1-100 nt is shown in the table below (half of the amounts were used for the experiment with *Arc* eRNA 1-200). The PEC sample was supplemented with 130  $\mu\text{L}$  1x complex buffer (supplemented with 1 mM DTT and  $\text{MgCl}_2$  to achieve a final

concentration of 3 mM), before *Nr4a1*-(b) 1-100 nt eRNA was added and incubated for 15 min at 30 °C. The sample was then loaded on a Superose 6 10/300 column preequilibrated with 1x complex buffer (+1 mM DTT) to separate the full PEC plus eRNA from excessive amounts of nucleic acids and proteins. The complex fractions were pooled and concentrated with centrifugal filters (100 kDa MWCO, Amicon Ultra-4) up to 0.5 – 1 mg/mL.

Component	Final amount	Volume
RNA-T-DNA hybrid (25 μM)	1.4 nmol	56 μL
Pol II (5 μM)	0.7 nmol	140 μL
NT-DNA (100 μM)	2.8 nmol	28 μL
DSIF (30 μM)	1.8 nmol	60 μL
NELF (30 μM)	1.8 nmol	60 μL
1x complex buffer + MgCl <sub>2</sub> + 1 mM DTT		130 μL
eRNA ( <i>Nr4a1</i> -(b) 1-100 nt) (25 μM)	1.8 nmol	75 μL

### 5.25.2 Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis and identification of protein-RNA crosslinks

The eRNA-bound PEC fractions were handed over to Dr. Alexander Leitner (Institute of Molecular Systems Biology, ETH Zurich, Switzerland), who performed the RNA-protein UV-crosslinking and LC-MS/MS analysis.

### 5.26 quantitative PCR (qPCR)

qPCR was performed to test the success of 7SL RNA depletion from total RNA (ca. 1 μg). First, after the depletion the RNA was purified by ethanol precipitation (section 5.10) or spin-columns (RNA Clean & Concentrator-5 columns, Zymo-Research). The RNA pellet was dissolved in 6 μL, or the RNA was eluted from the column with 6 μL DEPC-H<sub>2</sub>O. The RNA was reverse transcribed in a thermocycler using the SuperScript III First-Strand Synthesis SuperMix (Invitrogen), according to the manufacturer's protocol (see also below).

Component	Volume (20 $\mu$ L)
RNA	6 $\mu$ L
Annealing buffer	1 $\mu$ L
Random hexamers	1 $\mu$ L
Mix and incubate	5 min at 65 °C At least 1 min on ice
Add:	
2x FS reaction mix	10 $\mu$ L
SuperScript III/RNase OUT Enzyme mix	2 $\mu$ L
Mix and incubate	10 min at 25 °C 50 min at 50 °C 5 min at 85 °C

The qPCR reactions were set up in duplicates, in white 96-well white PCR plates (Bio-Rad), as described in the table below using the PowerUp SYBR Green Master Mix (Thermo Fisher Scientific). The used qPCR Primer for *7SL* RNA and *Gapdh* RNA are listed in the supplementary Table 7-9. The qPCR was run with in the standard cycling mode of the manufacturer's protocol (see also below) on the CFX Connect Real Time PCR Detection System (BioRad). qPCR results were analyzed by the comparative ( $\Delta\Delta C_T$ ) method (Livak & Schmittgen, 2001). Means of  $C_T$  values from the duplicates were used for calculations.

Component	Volume (20 $\mu$ L)	Step	Volume (20 $\mu$ L)	Duration
cDNA	1 $\mu$ L (ca. 5 ng)	UDG activation	95 °C	2 min
qPCR primer mix (forward + reverse) (5 $\mu$ M each)	0.24 $\mu$ L	Dual-Lock DNA polymerase	95 °C	2 min
ddH <sub>2</sub> O	7.4 $\mu$ L	Denature	95 °C $\mu$ L	15 s
PowerUp SYBR Green Master Mix	10 $\mu$ L	Anneal/Extend	60 °C	1 min
				Cycle 40 x

## 5.27 Data availability

The sequencing data from the final Exo-seq (Sequencing Run#2) and *in vitro* SHAPE-MaP experiment are deposited in the Gene Expression Omnibus (GEO) and are part of the series GSE163113. Data will be publicly available by December 2021.

## 6 References

- Adelman, K., & Lis, J. T. (2012). Promoter-proximal pausing of RNA polymerase II: Emerging roles in metazoans. *Nature Reviews. Genetics*, *13*(10), 720–731. <https://doi.org/10.1038/nrg3293>
- Adelman, K., Marr, M. T., Werner, J., Saunders, A., Ni, Z., Andrulis, E. D., & Lis, J. T. (2005). Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Molecular Cell*, *17*(1), 103–112. <https://doi.org/10.1016/j.molcel.2004.11.028>
- Afik, S., Bartok, O., Artyomov, M. N., Shishkin, A. A., Kadri, S., Hanan, M., Zhu, X., Garber, M., & Kadener, S. (2017). Defining the 5' and 3' landscape of the Drosophila transcriptome with Exo-seq and RNaseH-seq. *Nucleic Acids Research*, *45*(11), e95. <https://doi.org/10.1093/nar/gkx133>
- Alberti, S., Gladfelter, A., & Mittag, T. (2019). Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell*, *176*(3), 419–434. <https://doi.org/10.1016/j.cell.2018.12.035>
- Amleh, A., Nair, S. J., Sun, J., Sutherland, A., Hasty, P., & Li, R. (2009). Mouse Cofactor of BRCA1 (Cobra1) Is Required for Early Embryogenesis. *PLOS ONE*, *4*(4), e5034. <https://doi.org/10.1371/journal.pone.0005034>
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., ... Sandelin, A. (2014a). An atlas of active enhancers across human cell types and tissues. *Nature*, *507*(7493), 455–461. <https://doi.org/10.1038/nature12787>
- Andersson, R., Refsing Andersen, P., Valen, E., Core, L. J., Bornholdt, J., Boyd, M., Heick Jensen, T., & Sandelin, A. (2014b). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nature Communications*, *5*, 5336. <https://doi.org/10.1038/ncomms6336>
- Arner, E., Daub, C. O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drabløs, F., Lennartsson, A., Rønnerblad, M., Hrydziuszko, O., Vitezic, M., Freeman, T. C., Alhendi, A. M. N., Arner, P., Axton, R., Baillie, J. K., Beckhouse, A., Bodega, B., Briggs, J., Brombacher, F., ... Hayashizaki, Y. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science (New York, N.Y.)*, *347*(6225), 1010–1014. <https://doi.org/10.1126/science.1259418>
- Austenaa, L. M. I., Barozzi, I., Simonatto, M., Masella, S., Della Chiara, G., Ghisletti, S., Curina, A., de Wit, E., Bouwman, B. A. M., de Pretis, S., Piccolo, V., Termanini, A., Prosperini, E., Pelizzola, M., de Laat, W., & Natoli, G. (2015). Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. *Molecular Cell*, *60*(3), 460–474. <https://doi.org/10.1016/j.molcel.2015.09.018>
- Bahrami, S., & Drabløs, F. (2016). Gene regulation in the immediate-early response process. *Advances in Biological Regulation*, *62*, 37–49. <https://doi.org/10.1016/j.jbior.2016.05.001>
- Baillat, D., Hakimi, M.-A., Näär, A. M., Shilatifard, A., Cooch, N., & Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with

- the C-terminal repeat of RNA polymerase II. *Cell*, 123(2), 265–276. <https://doi.org/10.1016/j.cell.2005.08.019>
- Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: Organizers of cellular biochemistry. *Nature Reviews. Molecular Cell Biology*, 18(5), 285–298. <https://doi.org/10.1038/nrm.2017.7>
- Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27(2 Pt 1), 299–308. [https://doi.org/10.1016/0092-8674\(81\)90413-x](https://doi.org/10.1016/0092-8674(81)90413-x)
- Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4), 823–837. <https://doi.org/10.1016/j.cell.2007.05.009>
- Beltran, M., Tavares, M., Justin, N., Khandelwal, G., Ambrose, J., Foster, B. M., Worlock, K. B., Tvardovskiy, A., Kunzelmann, S., Herrero, J., Bartke, T., Gamblin, S. J., Wilson, J. R., & Jenner, R. G. (2019). G-tract RNA removes Polycomb Repressive Complex 2 from genes. *Nature Structural & Molecular Biology*, 26(10), 899–909. <https://doi.org/10.1038/s41594-019-0293-z>
- Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature Reviews. Genetics*, 15(3), 163–175. <https://doi.org/10.1038/nrg3662>
- Berger, I., Fitzgerald, D. J., & Richmond, T. J. (2004). Baculovirus expression system for heterologous multiprotein complexes. *Nature Biotechnology*, 22(12), 1583–1587. <https://doi.org/10.1038/nbt1036>
- Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M., & Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. *Nature*, 529(7587), 551–554. <https://doi.org/10.1038/nature16482>
- Bernecky, C., Plitzko, J. M., & Cramer, P. (2017). Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. *Nature Structural & Molecular Biology*, 24(10), 809–815. <https://doi.org/10.1038/nsmb.3465>
- Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Stamatoyannopoulos, J. A., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., ... Transcriptional Regulatory Elements. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146), 799–816. <https://doi.org/10.1038/nature05874>
- Boettiger, A. N., & Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. *Science (New York, N.Y.)*, 325(5939), 471–473. <https://doi.org/10.1126/science.1173976>
- Bose, D. A., Donahue, G., Reinberg, D., Shiekhatar, R., Bonasio, R., & Berger, S. L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell*, 168(1), 135–149.e22. <https://doi.org/10.1016/j.cell.2016.12.020>
- Broussard, J. A., & Green, K. J. (2017). Research Techniques Made Simple: Methodology and Applications of Förster Resonance Energy Transfer (FRET) Microscopy. *Journal of Investigative Dermatology*, 137(11), e185–e191. <https://doi.org/10.1016/j.jid.2017.09.006>

- Brunelle, J. L., & Green, R. (2013). In vitro transcription from plasmid or PCR-amplified DNA. *Methods in Enzymology*, *530*, 101–114. <https://doi.org/10.1016/B978-0-12-420037-1.00005-1>
- Bulger, M., & Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell*, *144*(3), 327–339. <https://doi.org/10.1016/j.cell.2011.01.024>
- Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Molecular Cell*, *36*(4), 541–546. <https://doi.org/10.1016/j.molcel.2009.10.019>
- Busan, S., & Weeks, K. M. (2017). Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA*, rna.061945.117. <https://doi.org/10.1261/rna.061945.117>
- Calo, E., & Wysocka, J. (2013). Modification of enhancer chromatin: What, how, and why? *Molecular Cell*, *49*(5), 825–837. <https://doi.org/10.1016/j.molcel.2013.01.038>
- Cazenave, C., & Uhlenbeck, O. C. (1994). RNA template-directed RNA synthesis by T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(15), 6972–6976.
- Cheng, B., & Price, D. H. (2007). Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *The Journal of Biological Chemistry*, *282*(30), 21901–21912. <https://doi.org/10.1074/jbc.M702936200>
- Cheng, B., & Price, D. H. (2008). Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. *Nucleic Acids Research*, *36*(20), e135. <https://doi.org/10.1093/nar/gkn630>
- Cheung, A. C. M., & Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, *471*(7337), 249–253. <https://doi.org/10.1038/nature09785>
- Cheung, A. C. M., Sainsbury, S., & Cramer, P. (2011). Structural basis of initial RNA polymerase II transcription. *The EMBO Journal*, *30*(23), 4755–4763. <https://doi.org/10.1038/emboj.2011.396>
- Core, L., & Adelman, K. (2019a). Promoter-proximal pausing of RNA polymerase II: A nexus of gene regulation. *Genes & Development*. <https://doi.org/10.1101/gad.325142.119>
- Core, L., & Adelman, K. (2019b). Promoter-proximal pausing of RNA polymerase II: A nexus of gene regulation. *Genes & Development*, *33*(15–16), 960–982. <https://doi.org/10.1101/gad.325142.119>
- Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, *46*(12), 1311–1320. <https://doi.org/10.1038/ng.3142>
- Core, L. J., Waterfall, J. J., & Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (New York, N.Y.)*, *322*(5909), 1845–1848. <https://doi.org/10.1126/science.1162228>
- Cramer, P. (2019). Organization and regulation of gene transcription. *Nature*, *573*(7772), 45–54. <https://doi.org/10.1038/s41586-019-1517-4>
- Cramer, P., Bushnell, D. A., & Kornberg, R. D. (2001). Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science*, *292*(5523), 1863–1876. <https://doi.org/10.1126/science.1059493>

- Danko, C. G., Hah, N., Luo, X., Martins, A. L., Core, L., Lis, J. T., Siepel, A., & Kraus, W. L. (2013). Signaling Pathways Differentially Affect RNA Polymerase II Initiation, Pausing, and Elongation Rate in Cells. *Molecular Cell*, *50*(2), 212–222. <https://doi.org/10.1016/j.molcel.2013.02.015>
- Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K. J., Gooding, A. R., Lee, J. T., & Cech, T. R. (2015). Toward a Consensus on the Binding Specificity and Promiscuity of PRC2 for RNA. *Molecular Cell*, *57*(3), 552–558. <https://doi.org/10.1016/j.molcel.2014.12.017>
- Davidovich, C., Zheng, L., Goodrich, K. J., & Cech, T. R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nature Structural & Molecular Biology*, *20*(11), 1250–1257. <https://doi.org/10.1038/nsmb.2679>
- de Laat, W., & Grosveld, F. (2003). Spatial organization of gene expression: The active chromatin hub. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *11*(5), 447–459. <https://doi.org/10.1023/a:1024922626726>
- De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., Muller, H., Ragoussis, J., Wei, C.-L., & Natoli, G. (2010). A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. *PLoS Biology*, *8*(5). <https://doi.org/10.1371/journal.pbio.1000384>
- Dekker, J. (2008). Gene regulation in the third dimension. *Science (New York, N.Y.)*, *319*(5871), 1793–1794. <https://doi.org/10.1126/science.1152850>
- Dengl, S., & Cramer, P. (2009). Torpedo Nuclease Rat1 Is Insufficient to Terminate RNA Polymerase II in Vitro. *The Journal of Biological Chemistry*, *284*(32), 21270–21279. <https://doi.org/10.1074/jbc.M109.013847>
- Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., & Zhou, M. M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, *399*(6735), 491–496. <https://doi.org/10.1038/20974>
- Dienemann, C., Schwalb, B., Schilbach, S., & Cramer, P. (2019). Promoter Distortion and Opening in the RNA Polymerase II Cleft. *Molecular Cell*, *73*(1), 97–106.e4. <https://doi.org/10.1016/j.molcel.2018.10.014>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Fitzgerald, D. J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T. J., & Berger, I. (2006). Protein complex expression by using multigene baculoviral vectors. *Nature Methods*, *3*(12), 1021–1032. <https://doi.org/10.1038/nmeth983>
- Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., & Peterlin, B. M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Molecular and Cellular Biology*, *24*(2), 787–795. <https://doi.org/10.1128/mcb.24.2.787-795.2004>
- Gaertner, B., & Zeitlinger, J. (2014). RNA polymerase II pausing during development. *Development*, *141*(6), 1179–1183. <https://doi.org/10.1242/dev.088492>
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. <https://doi.org/10.1038/nmeth.1318>

- Gilchrist, D. A., Dos Santos, G., Fargo, D. C., Xie, B., Gao, Y., Li, L., & Adelman, K. (2010). Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation. *Cell*, *143*(4), 540–551. <https://doi.org/10.1016/j.cell.2010.10.004>
- Gilchrist, D. A., Nechaev, S., Lee, C., Ghosh, S. K. B., Collins, J. B., Li, L., Gilmour, D. S., & Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes & Development*, *22*(14), 1921–1933. <https://doi.org/10.1101/gad.1643208>
- Gilmour, D. S., & Lis, J. T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Molecular and Cellular Biology*, *6*(11), 3984–3989.
- Goldman, S. R., Ebright, R. H., & Nickels, B. E. (2009). Direct Detection of Abortive RNA Transcripts in Vivo. *Science (New York, N.Y.)*, *324*(5929), 927–928. <https://doi.org/10.1126/science.1169237>
- Gorbovytska, V., Kim, S.-K., Kuybu, F., Götze, M., Um, D., Kang, K., Pittroff, A., Schneider, L.-M., Leitner, A., Kim, T.-K., & Kuhn, C.-D. (2021). Enhancer RNAs stimulate Pol II pause release by harnessing multivalent interactions to NELF. *BioRxiv*, 2021.04.25.441328. <https://doi.org/10.1101/2021.04.25.441328>
- Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., & Fraser, P. (2000). Intergenic Transcription and Developmental Remodeling of Chromatin Subdomains in the Human  $\beta$ -globin Locus. *Molecular Cell*, *5*(2), 377–386. [https://doi.org/10.1016/S1097-2765\(00\)80432-3](https://doi.org/10.1016/S1097-2765(00)80432-3)
- Gross, D. S., & Garrard, W. T. (1988). Nuclease hypersensitive sites in chromatin. *Annual Review of Biochemistry*, *57*, 159–197. <https://doi.org/10.1146/annurev.bi.57.070188.001111>
- Gu, X., Mooers, B. H. M., Thomas, L. M., Malone, J., Harris, S., & Schroeder, S. J. (2015). Structures and Energetics of Four Adjacent G·U Pairs That Stabilize an RNA Helix. *The Journal of Physical Chemistry B*, *119*(42), 13252–13261. <https://doi.org/10.1021/acs.jpcc.5b06970>
- Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., & Young, R. A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, *130*(1), 77–88. <https://doi.org/10.1016/j.cell.2007.05.042>
- Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G. A., Teng, M., & Niu, L. (2008). Core Structure of the yeast Spt4-Spt5 Complex: A Conserved Module for Regulation of Transcription Elongation. *Structure (London, England : 1993)*, *16*(11), 1649–1658. <https://doi.org/10.1016/j.str.2008.08.013>
- Guo, S., Yamaguchi, Y., Schilbach, S., Wada, T., Lee, J., Goddard, A., French, D., Handa, H., & Rosenthal, A. (2000). A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature*, *408*(6810), 366–369. <https://doi.org/10.1038/35042590>
- Guo, X., Myasnikov, A. G., Chen, J., Crucifix, C., Papai, G., Takacs, M., Schultz, P., & Weixlbaumer, A. (2018). Structural Basis for NusA Stabilized Transcriptional Pausing. *Molecular Cell*, *69*(5), 816–827.e4. <https://doi.org/10.1016/j.molcel.2018.02.008>
- Gupte, R., Muse, G. W., Chinenov, Y., Adelman, K., & Rogatsky, I. (2013). Glucocorticoid receptor represses proinflammatory genes at distinct steps of the transcription cycle.

- Proceedings of the National Academy of Sciences of the United States of America*, 110(36), 14616–14621. <https://doi.org/10.1073/pnas.1309898110>
- Hah, N., Danko, C. G., Core, L., Waterfall, J. J., Siepel, A., Lis, J. T., & Kraus, W. L. (2011). A Rapid, Extensive, and Transient Transcriptional Response to Estrogen Signaling in Breast Cancer Cells. *Cell*, 145(4), 622–634. <https://doi.org/10.1016/j.cell.2011.03.042>
- Harlen, K. M., & Churchman, L. S. (2017). The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology*, 18(4), 263–273. <https://doi.org/10.1038/nrm.2017.10>
- Hartzog, G. A., Wada, T., Handa, H., & Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes & Development*, 12(3), 357–369. <https://doi.org/10.1101/gad.12.3.357>
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, 39(3), 311–318. <https://doi.org/10.1038/ng1966>
- Heinz, S., Romanoski, C. E., Benner, C., & Glass, C. K. (2015). The selection and function of cell type-specific enhancers. *Nature Reviews. Molecular Cell Biology*, 16(3), 144–154. <https://doi.org/10.1038/nrm3949>
- Henninger, J. E., Oksuz, O., Shrinivas, K., Sagi, I., LeRoy, G., Zheng, M. M., Andrews, J. O., Zamudio, A. V., Lazaris, C., Hannett, N. M., Lee, T. I., Sharp, P. A., Cissé, I. I., Chakraborty, A. K., & Young, R. A. (2021). RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell*, 184(1), 207–225.e24. <https://doi.org/10.1016/j.cell.2020.11.030>
- Henriques, T., Scruggs, B. S., Inouye, M. O., Muse, G. W., Williams, L. H., Burkholder, A. B., Lavender, C. A., Fargo, D. C., & Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. *Genes & Development*. <https://doi.org/10.1101/gad.309351.117>
- Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell*, 169(1), 13–23. <https://doi.org/10.1016/j.cell.2017.02.007>
- Ho, C. K., & Shuman, S. (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Molecular Cell*, 3(3), 405–411. [https://doi.org/10.1016/s1097-2765\(00\)80468-2](https://doi.org/10.1016/s1097-2765(00)80468-2)
- Ho, Y., Elefant, F., Liebhaber, S. A., & Cooke, N. E. (2006). Locus control region transcription plays an active role in long-range gene activation. *Molecular Cell*, 23(3), 365–375. <https://doi.org/10.1016/j.molcel.2006.05.041>
- Hsieh, C.-L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C. J., Lee, G.-S. M., Chen, S., Balk, S. P., Liu, X. S., Brown, M., & Kantoff, P. W. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proceedings of the National Academy of Sciences*, 111(20), 7319–7324. <https://doi.org/10.1073/pnas.1324151111>
- Hyman, A. A., Weber, C. A., & Jülicher, F. (2014). Liquid-Liquid Phase Separation in Biology. *Annual Review of Cell and Developmental Biology*, 30(1), 39–58. <https://doi.org/10.1146/annurev-cellbio-100913-013325>

- Jankowsky, E., & Harris, M. E. (2015). Specificity and non-specificity in RNA–protein interactions. *Nature Reviews. Molecular Cell Biology*, 16(9), 533–544. <https://doi.org/10.1038/nrm4032>
- Jin, C., & Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes & Development*, 21(12), 1519–1529. <https://doi.org/10.1101/gad.1547707>
- Junwei, S., & Vakoc, C. R. (2014). The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Molecular Cell*, 54(5), 728–736. <https://doi.org/10.1016/j.molcel.2014.05.016>
- Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J., & Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314), 430–435. <https://doi.org/10.1038/nature09380>
- Kang, J. Y., Mishanina, T. V., Bellecourt, M. J., Mooney, R. A., Darst, S. A., & Landick, R. (2018). RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Molecular Cell*, 69(5), 802–815.e5. <https://doi.org/10.1016/j.molcel.2018.01.018>
- Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S., & Ebright, R. H. (2006). INITIAL TRANSCRIPTION BY RNA POLYMERASE PROCEEDS THROUGH A DNA-SCRUNCHING MECHANISM. *Science (New York, N.Y.)*, 314(5802), 1144–1147. <https://doi.org/10.1126/science.1131399>
- Kim, I. V., Ross, E. J., Dietrich, S., Döring, K., Sánchez Alvarado, A., & Kuhn, C.-D. (2019). Efficient depletion of ribosomal RNA for RNA sequencing in planarians. *BMC Genomics*, 20(1), 909. <https://doi.org/10.1186/s12864-019-6292-y>
- Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D., & Ren, B. (2005). A high-resolution map of active promoters in the human genome. *Nature*, 436(7052), 876–880. <https://doi.org/10.1038/nature03877>
- Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G., & Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465(7295), 182–187. <https://doi.org/10.1038/nature09033>
- Kim, Y. W., Lee, S., Yun, J., & Kim, A. (2015). Chromatin looping and eRNA transcription precede the transcriptional activation of gene in the  $\beta$ -globin locus. *Bioscience Reports*, 35(2). <https://doi.org/10.1042/BSR20140126>
- Klatt, F. (2020). *Activation of the human Mediator kinase CDK8 by MED12* [Doctoral thesis]. <https://epub.uni-bayreuth.de/5202/>
- Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T. K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A. L., Heidemann, M., Hintermair, C., Eick, D., Gut, I., Ferrier, P., & Andrau, J.-C. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural & Molecular Biology*, 18(8), 956–963. <https://doi.org/10.1038/nsmb.2085>
- Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics (Oxford, England)*, 28(24), 3211–3217. <https://doi.org/10.1093/bioinformatics/bts611>

- Kraus, A. J., Brink, B. G., & Siegel, T. N. (2019). Efficient and specific oligo-based depletion of rRNA. *Scientific Reports*, *9*(1), 12281. <https://doi.org/10.1038/s41598-019-48692-2>
- Kuroda, K. O., Meaney, M. J., Uetani, N., & Kato, T. (2008). Neurobehavioral basis of the impaired nurturing in mice lacking the immediate early gene FosB. *Brain Research*, *1211*, 57–71. <https://doi.org/10.1016/j.brainres.2008.02.100>
- Kwak, H., & Lis, J. T. (2013). Control of Transcriptional Elongation. *Annual Review of Genetics*, *47*, 483–508. <https://doi.org/10.1146/annurev-genet-110711-155440>
- Kyrpides, N. C., Woese, C. R., & Ouzounis, C. A. (1996). KOW: A novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends in Biochemical Sciences*, *21*(11), 425–426. [https://doi.org/10.1016/s0968-0004\(96\)30036-4](https://doi.org/10.1016/s0968-0004(96)30036-4)
- Lagha, M., Bothma, J. P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D. S., Zeitlinger, J., & Levine, M. S. (2013). Paused Pol II Coordinates Tissue Morphogenesis in the *Drosophila* Embryo. *Cell*, *153*(5), 976–987. <https://doi.org/10.1016/j.cell.2013.04.045>
- Lai, F., Gardini, A., Zhang, A., & Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature*, *525*(7569), 399–403. <https://doi.org/10.1038/nature14906>
- Lai, F., Orom, U. A., Cesaroni, M., Beringer, M., Taatjes, D. J., Blobel, G. A., & Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*, *494*(7438), 497–501. <https://doi.org/10.1038/nature11884>
- Lam, M. T. Y., Cho, H., Lesch, H. P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M. U., Kim, A. S., Kosaka, M., Lee, C. Y., Watt, A., Grossman, T. R., Rosenfeld, M. G., Evans, R. M., & Glass, C. K. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, *498*(7455), 511–515. <https://doi.org/10.1038/nature12209>
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Leppek, K., Das, R., & Barna, M. (2018). Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews. Molecular Cell Biology*, *19*(3), 158–174. <https://doi.org/10.1038/nrm.2017.103>
- Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Current Biology: CB*, *20*(17), R754–763. <https://doi.org/10.1016/j.cub.2010.06.070>
- Li, B., Carey, M., & Workman, J. L. (2007). The Role of Chromatin during Transcription. *Cell*, *128*(4), 707–719. <https://doi.org/10.1016/j.cell.2007.01.015>
- Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., Oh, S., Kim, H.-S., Glass, C. K., & Rosenfeld, M. G. (2013a). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, *498*(7455), 516–520. <https://doi.org/10.1038/nature12210>
- Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., Oh, S., Kim, H.-S., Glass, C. K., & Rosenfeld, M. G. (2013b). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, *498*(7455), 516–520. <https://doi.org/10.1038/nature12210>
- Li, W., Notani, D., & Rosenfeld, M. G. (2016). Enhancers as non-coding RNA transcription units: Recent insights and future perspectives. *Nature Reviews Genetics*, *17*(4), 207–223. <https://doi.org/10.1038/nrg.2016.4>

- Li, X., Kazan, H., Lipshitz, H. D., & Morris, Q. D. (2014). Finding the target sites of RNA-binding proteins. *Wiley Interdisciplinary Reviews. RNA*, 5(1), 111–130. <https://doi.org/10.1002/wrna.1201>
- Lindstrom, D. L., Squazzo, S. L., Muster, N., Burckin, T. A., Wachter, K. C., Emigh, C. A., McCleery, J. A., Yates, J. R., & Hartzog, G. A. (2003). Dual Roles for Spt5 in Pre-mRNA Processing and Transcription Elongation Revealed by Identification of Spt5-Associated Proteins. *Molecular and Cellular Biology*, 23(4), 1368–1378. <https://doi.org/10.1128/MCB.23.4.1368-1378.2003>
- Ling, J., Baibakov, B., Pi, W., Emerson, B. M., & Tuan, D. (2005). The HS2 enhancer of the beta-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a cis-linked globin promoter. *Journal of Molecular Biology*, 350(5), 883–896. <https://doi.org/10.1016/j.jmb.2005.05.039>
- Liu, F., Somarowthu, S., & Pyle, A. M. (2017). Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nature Chemical Biology*, 13(3), 282–289. <https://doi.org/10.1038/nchembio.2272>
- Liu, P., Kenney, J. M., Stiller, J. W., & Greenleaf, A. L. (2010). Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Molecular Biology and Evolution*, 27(11), 2628–2641. <https://doi.org/10.1093/molbev/msq151>
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods*, 25(4), 402–408. <https://doi.org/10.1006/meth.2001.1262>
- Lubas, M., Andersen, P. R., Schein, A., Dziembowski, A., Kudla, G., & Jensen, T. H. (2015). The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Reports*, 10(2), 178–192. <https://doi.org/10.1016/j.celrep.2014.12.026>
- Lucks, J. B., Mortimer, S. A., Trapnell, C., Luo, S., Aviran, S., Schroth, G. P., Pachter, L., Doudna, J. A., & Arkin, A. P. (2011). Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proceedings of the National Academy of Sciences*, 108(27), 11063–11068. <https://doi.org/10.1073/pnas.1106501108>
- Luo, Z., Lin, C., & Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. *Nature Reviews. Molecular Cell Biology*, 13(9), 543–547. <https://doi.org/10.1038/nrm3417>
- Lyskov, S., Chou, F.-C., Conchúir, S. Ó., Der, B. S., Drew, K., Kuroda, D., Xu, J., Weitzner, B. D., Renfrew, P. D., Sripakdeevong, P., Borgo, B., Havranek, J. J., Kuhlman, B., Kortemme, T., Bonneau, R., Gray, J. J., & Das, R. (2013). Serverification of molecular modeling applications: The Rosetta Online Server that Includes Everyone (ROSIE). *PloS One*, 8(5), e63906. <https://doi.org/10.1371/journal.pone.0063906>
- Ma, D. K., Jang, M.-H., Guo, J. U., Kitabatake, Y., Chang, M.-L., Pow-Anpongkul, N., Flavell, R. A., Lu, B., Ming, G.-L., & Song, H. (2009). Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science (New York, N.Y.)*, 323(5917), 1074–1077. <https://doi.org/10.1126/science.1166859>
- Mahat, D. B., Kwak, H., Booth, G. T., Jonkers, I. H., Danko, C. G., Patel, R. K., Waters, C. T., Munson, K., Core, L. J., & Lis, J. T. (2016). Base-Pair Resolution Genome-Wide Mapping

- Of Active RNA polymerases using Precision Nuclear Run-On (PRO-seq). *Nature Protocols*, 11(8), 1455–1476. <https://doi.org/10.1038/nprot.2016.086>
- Malik, A. N., Vierbuchen, T., Hemberg, M., Rubin, A. A., Ling, E., Couch, C. H., Stroud, H., Spiegel, I., Farh, K. K.-H., Harmin, D. A., & Greenberg, M. E. (2014). Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nature Neuroscience*, 17(10), 1330–1339. <https://doi.org/10.1038/nn.3808>
- Mandal, S. S., Chu, C., Wada, T., Handa, H., Shatkin, A. J., & Reinberg, D. (2004). Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proceedings of the National Academy of Sciences*, 101(20), 7572–7577. <https://doi.org/10.1073/pnas.0401493101>
- Margueron, R., & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature*, 469(7330), 343–349. <https://doi.org/10.1038/nature09784>
- Marshall, N. F., & Price, D. H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Molecular and Cellular Biology*, 12(5), 2078–2090.
- Marshall, N. F., & Price, D. H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *The Journal of Biological Chemistry*, 270(21), 12335–12338. <https://doi.org/10.1074/jbc.270.21.12335>
- Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C., & Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity: RNA polymerase-Spt4/5 complex architecture. *The EMBO Journal*, 30(7), 1302–1310. <https://doi.org/10.1038/emboj.2011.64>
- Melo, C. A., Drost, J., Wijchers, P. J., van de Werken, H., de Wit, E., Oude Vrielink, J. A. F., Elkon, R., Melo, S. A., Léveillé, N., Kalluri, R., de Laat, W., & Agami, R. (2013). ERNAs are required for p53-dependent enhancer activity and gene transcription. *Molecular Cell*, 49(3), 524–535. <https://doi.org/10.1016/j.molcel.2012.11.021>
- Merino, E. J., Wilkinson, K. A., Coughlan, J. L., & Weeks, K. M. (2005). RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *Journal of the American Chemical Society*, 127(12), 4223–4231. <https://doi.org/10.1021/ja043822v>
- Michaelis, C., Ciosk, R., & Nasmyth, K. (1997). Cohesins: Chromosomal proteins that prevent premature separation of sister chromatids. *Cell*, 91(1), 35–45. [https://doi.org/10.1016/s0092-8674\(01\)80007-6](https://doi.org/10.1016/s0092-8674(01)80007-6)
- Minatohara, K., Akiyoshi, M., & Okuno, H. (2015). Role of Immediate-Early Genes in Synaptic Plasticity and Neuronal Ensembles Underlying the Memory Trace. *Frontiers in Molecular Neuroscience*, 8, 78. <https://doi.org/10.3389/fnmol.2015.00078>
- Mishiro, T., Ishihara, K., Hino, S., Tsutsumi, S., Aburatani, H., Shirahige, K., Kinoshita, Y., & Nakao, M. (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *The EMBO Journal*, 28(9), 1234–1245. <https://doi.org/10.1038/emboj.2009.81>
- Missra, A., & Gilmour, D. S. (2010). Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proceedings of the National Academy of Sciences*, 107(25), 11301–11306. <https://doi.org/10.1073/pnas.1000681107>

- Mito, Y., Henikoff, J. G., & Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. *Science (New York, N.Y.)*, *315*(5817), 1408–1411. <https://doi.org/10.1126/science.1134004>
- Moore, M. J., & Proudfoot, N. J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*, *136*(4), 688–700. <https://doi.org/10.1016/j.cell.2009.02.001>
- Moteki, S., & Price, D. (2002). Functional Coupling of Capping and Transcription of mRNA. *Molecular Cell*, *10*(3), 599–609. [https://doi.org/10.1016/S1097-2765\(02\)00660-3](https://doi.org/10.1016/S1097-2765(02)00660-3)
- Mousavi, K., Zare, H., Dell’orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G. L., & Sartorelli, V. (2013). ERNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Molecular Cell*, *51*(5), 606–617. <https://doi.org/10.1016/j.molcel.2013.07.022>
- Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D. A., Adams, C. M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B. J., Levitt, M., & Kornberg, R. D. (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. *Science (New York, N.Y.)*, *342*(6159), 1238724. <https://doi.org/10.1126/science.1238724>
- Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., Zeitlinger, J., & Adelman, K. (2007). RNA polymerase is poised for activation across the genome. *Nature Genetics*, *39*(12), 1507–1511. <https://doi.org/10.1038/ng.2007.21>
- Mustoe, A. M., Busan, S., Rice, G. M., Hajdin, C. E., Peterson, B. K., Ruda, V. M., Kubica, N., Nutiu, R., Baryza, J. L., & Weeks, K. M. (2018). Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell*, *173*(1), 181-195.e18. <https://doi.org/10.1016/j.cell.2018.02.034>
- Nair, S. J., Yang, L., Meluzzi, D., Oh, S., Yang, F., Friedman, M. J., Wang, S., Suter, T., Alshareedah, I., Gamliel, A., Ma, Q., Zhang, J., Hu, Y., Tan, Y., Ohgi, K. A., Jayani, R. S., Banerjee, P. R., Aggarwal, A. K., & Rosenfeld, M. G. (2019). Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nature Structural & Molecular Biology*, *26*(3), 193–203. <https://doi.org/10.1038/s41594-019-0190-5>
- Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D., Hasegawa, J., Omori, M., Inukai, N., Endoh, M., Yamada, T., & Handa, H. (2003). Human Transcription Elongation Factor NELF: Identification of Novel Subunits and Reconstitution of the Functionally Active Complex. *Molecular and Cellular Biology*, *23*(6), 1863–1873. <https://doi.org/10.1128/MCB.23.6.1863-1873.2003>
- Nechaev, S., Fargo, D. C., dos Santos, G., Liu, L., Gao, Y., & Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science (New York, N.Y.)*, *327*(5963), 335–338. <https://doi.org/10.1126/science.1181421>
- Pagano, J. M., Kwak, H., Waters, C. T., Sprouse, R. O., White, B. S., Ozer, A., Szeto, K., Shalloway, D., Craighead, H. G., & Lis, J. T. (2014). Defining NELF-E RNA Binding in HIV-1 and Promoter- Proximal Pause Regions. *PLOS Genetics*, *10*(1), 11.
- Palangat, M., Renner, D. B., Price, D. H., & Landick, R. (2005). A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIIS. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(42), 15036–15041. <https://doi.org/10.1073/pnas.0409405102>

- Pei, Y., & Shuman, S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *The Journal of Biological Chemistry*, 277(22), 19639–19648. <https://doi.org/10.1074/jbc.M200015200>
- Pertea, G., & Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. *F1000Research*, 9. <https://doi.org/10.12688/f1000research.23297.2>
- Peterlin, B. M., & Price, D. H. (2006). Controlling the Elongation Phase of Transcription with P-TEFb. *Molecular Cell*, 23(3), 297–305. <https://doi.org/10.1016/j.molcel.2006.06.014>
- Petrov, A., Wu, T., Puglisi, E. V., & Puglisi, J. D. (2013). RNA purification by preparative polyacrylamide gel electrophoresis. *Methods in Enzymology*, 530, 315–330. <https://doi.org/10.1016/B978-0-12-420037-1.00017-8>
- Phillips, J. E., & Corces, V. G. (2009). CTCF: Master weaver of the genome. *Cell*, 137(7), 1194–1211. <https://doi.org/10.1016/j.cell.2009.06.001>
- Polymenidou, M. (2018). The RNA face of phase separation. *Science (New York, N.Y.)*, 360(6391), 859–860. <https://doi.org/10.1126/science.aat8028>
- Ponting, C. P. (2002). Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Research*, 30(17), 3643–3652. <https://doi.org/10.1093/nar/gkf498>
- Pool, M. R. (2005). Signal recognition particles in chloroplasts, bacteria, yeast and mammals (Review). *Molecular Membrane Biology*, 22(1–2), 3–15. <https://doi.org/10.1080/09687860400026348>
- Qian, X., Zhao, J., Yeung, P. Y., Zhang, Q. C., & Kwok, C. K. (2019). Revealing lncRNA Structures and Interactions by Sequencing-Based Approaches. *Trends in Biochemical Sciences*, 44(1), 33–52. <https://doi.org/10.1016/j.tibs.2018.09.012>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Rahnamoun, H., Lee, J., Sun, Z., Lu, H., Ramsey, K. M., Komives, E. A., & Lauberth, S. M. (2018). RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation. *Nature Structural & Molecular Biology*, 25(8), 687–697. <https://doi.org/10.1038/s41594-018-0102-0>
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, 44(W1), W160–165. <https://doi.org/10.1093/nar/gkw257>
- Rao, Jampani N., Neumann, L., Wenzel, S., Schweimer, K., Rösch, P., & Wöhrle, B. M. (2006). Structural studies on the RNA-recognition motif of NELF E, a cellular negative transcription elongation factor involved in the regulation of HIV transcription. *The Biochemical Journal*, 400(3), 449–456. <https://doi.org/10.1042/BJ20060421>
- Rao, Jampani Nageswara, Schweimer, K., Wenzel, S., Wöhrle, B. M., & Rösch, P. (2008). NELF-E RRM undergoes major structural changes in flexible protein regions on target RNA binding. *Biochemistry*, 47(12), 3756–3761. <https://doi.org/10.1021/bi702429m>
- Rasmussen, E. B., & Lis, J. T. (1993). In vivo transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proceedings of the National Academy of Sciences of the United States of America*, 90(17), 7923–7927. <https://doi.org/10.1073/pnas.90.17.7923>

- Rasmussen, E. B., & Lis, J. T. (1995). Short transcripts of the ternary complex provide insight into RNA polymerase II elongational pausing. *Journal of Molecular Biology*, 252(5), 522–535. <https://doi.org/10.1006/jmbi.1995.0517>
- Rawat, P., Boehning, M., Hummel, B., Aprile-Garcia, F., Pandit, A. S., Eisenhardt, N., Khavaran, A., Niskanen, E., Vos, S. M., Palvimo, J. J., Pichler, A., Cramer, P., & Sawarkar, R. (2021). Stress-induced nuclear condensation of NELF drives transcriptional downregulation. *Molecular Cell*, 81(5), 1013-1026.e11. <https://doi.org/10.1016/j.molcel.2021.01.016>
- Reines, D., Ghanouni, P., Li, Q. Q., & Mote, J. (1992). The RNA polymerase II elongation complex. Factor-dependent transcription elongation involves nascent RNA cleavage. *The Journal of Biological Chemistry*, 267(22), 15516–15522.
- Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, 11, 129. <https://doi.org/10.1186/1471-2105-11-129>
- Reyes, A., & Huber, W. (2018). Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Research*, 46(2), 582–592. <https://doi.org/10.1093/nar/gkx1165>
- Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L. N., Beck, M., Schmidt, A., Mueller, M., & Aebersold, R. (2008). Identification of cross-linked peptides from large sequence databases. *Nature Methods*, 5(4), 315–318. <https://doi.org/10.1038/nmeth.1192>
- Rosenberg, M., Blum, R., Kesner, B., Aeby, E., Garant, J.-M., Szanto, A., & Lee, J. T. (2021). Motif-driven interactions between RNA and PRC2 are rheostats that regulate transcription elongation. *Nature Structural & Molecular Biology*, 28(1), 103–117. <https://doi.org/10.1038/s41594-020-00535-9>
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., & Weissman, J. S. (2014a). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, 505(7485), 701–705. <https://doi.org/10.1038/nature12894>
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., & Weissman, J. S. (2014b). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, 505(7485), 701–705. <https://doi.org/10.1038/nature12894>
- Roy, R., Hohng, S., & Ha, T. (2008). A Practical Guide to Single Molecule FRET. *Nature Methods*, 5(6), 507–516. <https://doi.org/10.1038/nmeth.1208>
- Sadovskaya, N., Mustafaev, O., Tyurin, A., & Goldenkova-Pavlova, I. (2020). *JetGene – web database and toolkit for analysis of regulatory regions or nucleotide contexts at differently translated transcripts of plants*. <https://doi.org/10.28983/PLAMIC2020.212>
- Saha, R. N., Wissink, E. M., Bailey, E. R., Zhao, M., Fargo, D. C., Hwang, J.-Y., Daigle, K. R., Fenn, J. D., Adelman, K., & Dudek, S. M. (2011). Rapid activity-induced transcription of Arc and other IEGs relies on poised RNA polymerase II. *Nature Neuroscience*, 14(7), 848–856. <https://doi.org/10.1038/nn.2839>
- Sainsbury, S., Niesser, J., & Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, 493(7432), 437–440. <https://doi.org/10.1038/nature11715>
- Saldaña-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E. P., Bruneau, B. G., Tsirigos, A., Furlan-Magaril, M., Skok, J., & Reinberg, D. (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Molecular Cell*, 76(3), 412-422.e5. <https://doi.org/10.1016/j.molcel.2019.08.015>

- Sanchez de Groot, N., Armaos, A., Graña-Montes, R., Alriquet, M., Calloni, G., Vabulas, R. M., & Tartaglia, G. G. (2019). RNA structure drives interaction with proteins. *Nature Communications*, *10*(1), 3246. <https://doi.org/10.1038/s41467-019-10923-5>
- Sartorelli, V., & Lauberth, S. M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. *Nature Structural & Molecular Biology*, *27*(6), 521–528. <https://doi.org/10.1038/s41594-020-0446-0>
- Saunders, A., Core, L. J., Sutcliffe, C., Lis, J. T., & Ashe, H. L. (2013). Extensive polymerase pausing during *Drosophila* axis patterning enables high-level and pliable transcription. *Genes & Development*, *27*(10), 1146–1158. <https://doi.org/10.1101/gad.215459.113>
- Schaukowitch, K., Joo, J.-Y., Liu, X., Watts, J. K., Martinez, C., & Kim, T.-K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. *Molecular Cell*, *56*(1), 29–42. <https://doi.org/10.1016/j.molcel.2014.08.023>
- Schier, A. C., & Taatjes, D. J. (2020). Structure and mechanism of the RNA polymerase II transcription machinery. *Genes & Development*, *34*(7–8), 465–488. <https://doi.org/10.1101/gad.335679.119>
- Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., & Cramer, P. (2017). Structures of transcription pre-initiation complex with TFIID and Mediator. *Nature*, *551*(7679), 204–209. <https://doi.org/10.1038/nature24282>
- Schwab, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., & Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science (New York, N.Y.)*, *352*(6290), 1225–1228. <https://doi.org/10.1126/science.aad9841>
- Shechner, D. M., Hacisuleyman, E., Younger, S. T., & Rinn, J. L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nature Methods*, *12*(7), 664–670. <https://doi.org/10.1038/nmeth.3433>
- Sheridan, R. M., Fong, N., D’Alessandro, A., & Bentley, D. L. (2019). Widespread backtracking by RNA pol II is a major effector of gene activation, 5’ pause release, termination and transcription elongation rate. *Molecular Cell*, *73*(1), 107–118.e4. <https://doi.org/10.1016/j.molcel.2018.10.031>
- Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E., & Weeks, K. M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nature Methods*, *11*(9), 959–965. <https://doi.org/10.1038/nmeth.3029>
- Sigova, A. A., Abraham, B. J., Ji, X., Molinie, B., Hannett, N. M., Guo, Y. E., Jangi, M., Giallourakis, C. C., Sharp, P. A., & Young, R. A. (2015). Transcription factor trapping by RNA in gene regulatory elements. *Science*, *350*(6263), 978–981. <https://doi.org/10.1126/science.aad3346>
- Smola, M. J., Calabrese, J. M., & Weeks, K. M. (2015). Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry*, *54*(46), 6867–6875. <https://doi.org/10.1021/acs.biochem.5b00977>
- Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A., & Weeks, K. M. (2015). Selective 2’-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nature Protocols*, *10*(11), 1643–1669. <https://doi.org/10.1038/nprot.2015.103>
- Smola, M. J., & Weeks, K. M. (2018). In-cell RNA structure probing with SHAPE-MaP. *Nature Protocols*, *13*(6), 1181–1195. <https://doi.org/10.1038/nprot.2018.010>

- Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J.-W., Kuchelmeister, H. Y., Batista, P. J., Torre, E. A., Kool, E. T., & Chang, H. Y. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, *519*(7544), 486–490. <https://doi.org/10.1038/nature14263>
- Sultan, M., Amstislavskiy, V., Risch, T., Schuette, M., Dökel, S., Ralser, M., Balzereit, D., Lehrach, H., & Yaspo, M.-L. (2014). Influence of RNA extraction methods and library selection schemes on RNA-seq data. *BMC Genomics*, *15*(1), 675. <https://doi.org/10.1186/1471-2164-15-675>
- Takahashi, H., Lassmann, T., Murata, M., & Carninci, P. (2012). 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nature Protocols*, *7*(3), 542–561. <https://doi.org/10.1038/nprot.2012.005>
- Talkish, J., May, G., Lin, Y., Woolford, J. L., & McManus, C. J. (2014). Mod-seq: High-throughput sequencing for chemical probing of RNA structure. *RNA*, *20*(5), 713–720. <https://doi.org/10.1261/rna.042218.113>
- Tome, J. M., Tippens, N. D., & Lis, J. T. (2018). Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nature Genetics*, *50*(11), 1533–1541. <https://doi.org/10.1038/s41588-018-0234-5>
- Trotta, E. (2014). On the Normalization of the Minimum Free Energy of RNAs by Sequence Length. *PLoS ONE*, *9*(11). <https://doi.org/10.1371/journal.pone.0113380>
- Varani, G., & McClain, W. H. (2000). The G-U wobble base pair. *EMBO Reports*, *1*(1), 18–23. <https://doi.org/10.1093/embo-reports/kvd001>
- Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., Park, T. J., Deaville, R., Erichsen, J. T., Jasinska, A. J., Turner, J. M. A., Bertelsen, M. F., Murchison, E. P., Flicek, P., & Odom, D. T. (2015). Enhancer evolution across 20 mammalian species. *Cell*, *160*(3), 554–566. <https://doi.org/10.1016/j.cell.2015.01.006>
- Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., & Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, *457*(7231), 854–858. <https://doi.org/10.1038/nature07730>
- Vos, S. M., Farnung, L., Urlaub, H., & Cramer, P. (2018). Structure of paused transcription complex Pol II-DSIF-NELF. *Nature*, *560*(7720), 601–606. <https://doi.org/10.1038/s41586-018-0442-2>
- Vos, S. M., Pöllmann, D., Caizzi, L., Hofmann, K. B., Rombaut, P., Zimniak, T., Herzog, F., & Cramer, P. (2016, June 10). *Architecture and RNA binding of the human negative elongation factor*. *ELife*; eLife Sciences Publications Limited. <https://doi.org/10.7554/eLife.14981>
- Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., & Handa, H. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & Development*, *12*(3), 343–356. <https://doi.org/10.1101/gad.12.3.343>
- Wang, C., Lee, J.-E., Lai, B., Macfarlan, T. S., Xu, S., Zhuang, L., Liu, C., Peng, W., & Ge, K. (2016). Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(42), 11871–11876. <https://doi.org/10.1073/pnas.1606857113>

- Wang, X., Goodrich, K. J., Gooding, A. R., Naeem, H., Archer, S., Paucek, R. D., Youmans, D. T., Cech, T. R., & Davidovich, C. (2017). Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Molecular Cell*, *65*(6), 1056-1067.e5. <https://doi.org/10.1016/j.molcel.2017.02.003>
- Watkins, A. M., Rangan, R., & Das, R. (2020). FARFAR2: Improved De Novo Rosetta Prediction of Complex Global RNA Folds. *Structure (London, England: 1993)*, *28*(8), 963-976.e6. <https://doi.org/10.1016/j.str.2020.05.011>
- Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., Abraham, B. J., Cohen, M. A., Nabet, B., Buckley, D. L., Guo, Y. E., Hnisz, D., Jaenisch, R., Bradner, J. E., Gray, N. S., & Young, R. A. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*, *171*(7), 1573-1588.e28. <https://doi.org/10.1016/j.cell.2017.11.008>
- Wen, Y., & Shatkin, A. J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes & Development*, *13*(14), 1774-1779. <https://doi.org/10.1101/gad.13.14.1774>
- Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K., & Peters, J.-M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, *451*(7180), 796-801. <https://doi.org/10.1038/nature06634>
- Werner, F. (2012). A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *Journal of Molecular Biology*, *417*(1-2), 13-27. <https://doi.org/10.1016/j.jmb.2012.01.031>
- Wilkinson, K. A., Merino, E. J., & Weeks, K. M. (2006). Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nature Protocols*, *1*(3), 1610-1616. <https://doi.org/10.1038/nprot.2006.249>
- Xu, C., Park, J.-K., & Zhang, J. (2019). Evidence that alternative transcriptional initiation is largely nonadaptive. *PLOS Biology*, *17*(3), e3000197. <https://doi.org/10.1371/journal.pbio.3000197>
- Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T., & Handa, H. (2006). P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Molecular Cell*, *21*(2), 227-237. <https://doi.org/10.1016/j.molcel.2005.11.024>
- Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., & Handa, H. (2002). Evidence that Negative Elongation Factor Represses Transcription Elongation through Binding to a DRB Sensitivity-Inducing Factor/RNA Polymerase II Complex and RNA. *Molecular and Cellular Biology*, *22*(9), 2918-2927. <https://doi.org/10.1128/MCB.22.9.2918-2927.2002>
- Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., & Handa, H. (1999). NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation. *Cell*, *97*(1), 41-51. [https://doi.org/10.1016/S0092-8674\(00\)80713-8](https://doi.org/10.1016/S0092-8674(00)80713-8)
- Yang, S., Oksenberg, N., Takayama, S., Heo, S.-J., Poliakov, A., Ahituv, N., Dubchak, I., & Boffelli, D. (2015). Functionally conserved enhancers with divergent sequences in distant vertebrates. *BMC Genomics*, *16*. <https://doi.org/10.1186/s12864-015-2070-7>

- Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., & Young, R. A. (2007). RNA Polymerase Stalling at Developmental Control Genes in the *Drosophila* Embryo. *Nature Genetics*, 39(12), 1512–1516. <https://doi.org/10.1038/ng.2007.26>
- Zentner, G. E., Tesar, P. J., & Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research*, 21(8), 1273–1283. <https://doi.org/10.1101/gr.122382.111>
- Zhang, B. H., Pan, X. P., Cox, S. B., Cobb, G. P., & Anderson, T. A. (2006). Evidence that miRNAs are different from other RNAs. *Cellular and Molecular Life Sciences: CMLS*, 63(2), 246–254. <https://doi.org/10.1007/s00018-005-5467-7>
- Zhao, Y., Wang, L., Ren, S., Wang, L., Blackburn, P. R., McNulty, M. S., Gao, X., Qiao, M., Vessella, R. L., Kohli, M., Zhang, J., Karnes, R. J., Tindall, D. J., Kim, Y., MacLeod, R., Ekker, S. C., Kang, T., Sun, Y., & Huang, H. (2016). Activation of P-TEFb by Androgen Receptor-Regulated Enhancer RNAs in Castration-Resistant Prostate Cancer. *Cell Reports*, 15(3), 599–610. <https://doi.org/10.1016/j.celrep.2016.03.038>
- Zuber, J., & Mathews, D. H. (2019). Estimating uncertainty in predicted folding free energy changes of RNA secondary structures. *RNA*, 25(6), 747–754. <https://doi.org/10.1261/rna.069203.118>
- Zuber, P. K. (2018). Structure and nucleic acid binding properties of KOW domains 4 and 6–7 of human transcription elongation factor DSIF. *Scientific Reports*, 12.
- Zuker, M., & Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1), 133–148. <https://doi.org/10.1093/nar/9.1.133>

## 7 Supplement

### 7.1 Materials

#### 7.1.1 Chemicals and Reagents

Chemicals and reagents used in this study were purchased from Carl Roth GmbH & Co (Karlsruhe, Germany), exceptions are specified in the Table below.

**Table 7-1 |** Chemicals and reagents

<b>Chemicals/Reagents</b>	<b>Manufacturer</b>
1-Methyl-7-nitroisatoic anhydride (1M7)	abcr GmbH
Agar (bacteriological)	Chemsolute
Agarose LE	GERBU Biotechnik GmbH
EZ Vision One DNA loading dye (6x)	Amresco
Glycogen (20 mg/ml)	Roche
HYDRANAL-Formamide (dry)	Fluka
SYBR Gold Nucleic Acid Gel Stain	Invitrogen
dNTP solution mix (10 mM each), DMSO	New England Biolabs
Manganese(II)-chloride (MnCl <sub>2</sub> ) solution (1M)	Fisher BioReagents
Polyethyleneimine (PEI) solution (50% in water)	Sigma-Aldrich
[ $\gamma$ - <sup>32</sup> P]-ATP 3000 Ci/mmol, 10mCi/ml, 250 $\mu$ Ci	PerkinElmer
F-4 Baculo Express ICM (insect cell medium)	BioConcept AG
XtremeGENE™ HP	Roche

#### 7.1.2 Critical commercial kits for nucleic acid applications

**Table 7-2 |** Commercial kits and beads used for purification of DNA and RNA and for depletion of rRNA.

<b>Kit</b>	<b>Manufacturer</b>
QIAprep Spin Miniprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
RNA Clean & Concentrator-5	Zymo Research
RiboCop rRNA depletion kit V1.2 (human/mouse/rat)	Lexogen
Ribo-Zero rRNA Removal Kit	Illumina
SuperScript III First-Strand Synthesis SuperMix	Invitrogen
PowerUp SYBR Green Master Mix	ThermoFisher Scientific
Agencourt AMPure XP beads	Beckman Coulter

Dynabeads MyOne Silane beads	Invitrogen
Dynabeads MyOne Streptavidin C1 beads	ThermoFisher Scientific

### 7.1.3 Enzymes

**Table 7-3 | Enzymes used in this study**

<b>Enzyme</b>	<b>Manufacturer / Source</b>
Benzonase Endonuclease	selfmade (by Silke Spudeit, AG Kuhn)
Calf Intestinal Alkaline Phosphatase (CIP)	New England Biolabs (NEB)
DNase I, recombinant (RNase free; 10 U/μL)	Roche
FastAP thermosensitive Alkaline Phosphatase (1 U/μL)	Thermo Scientific
Phusion High-Fidelity DNA Polymerase (2 U/μL)	NEB
Proteinase K	Roche
SuperScript II reverse transcriptase	Invitrogen
SuperScript III reverse transcriptase	Invitrogen
Taq DNA Ligase (40 U/μL)	NEB
Taq DNA polymerase (5 U/μL)	NEB
T5 Exonuclease (10 U/μL)	NEB
T4 Polynucleotide Kinase (PNK) (10 U/μL)	NEB
T4 RNA Ligase 1 (ssRNA Ligase)	New England (NEB)
T7 RNA polymerase (ca. 5 mg/mL)	selfmade (by Silke Spudeit, AG Kuhn)

### 7.1.4 Antibodies

**Table 7-4 | Antibodies used in this study**

<b>Antibody</b>	<b>Manufacturer</b>	<b>Identifier</b>
anti-NELF-E	Abcam	Cat# ab170104
anti-SPT5 (D-3) X	Santa Cruz Biotech	Cat# sc-133217X
anti-Pol II CTD (8WG16)	Core Facility Monoclonal Antibodies, German Research Center for Environmental Health, Munich	

### 7.1.5 Nucleic acid and Protein Standards

DNA, RNA and protein standards used in this work listed along with their corresponding manufacturer. The DNA standards were used for agarose gels, the RNA and ssDNA standards were used for Urea gels, and the protein standard was used for SDS-PAGE.

**Table 7-5 | DNA, RNA and protein standards used in this study**

<b>Standard</b>	<b>Manufacturer</b>
<b>RNA and ssDNA standard:</b>	
Low Molecular Weight Marker 10-100 nt (ssDNA)	Affymetrix
DNA oligo length standard 20/100 nt (ssDNA)	Integrated DNA Technologies
RiboRuler low range RNA ladder (ssRNA)	Thermo Scientific
<b>DNA standard:</b>	
1 kb DNA ladder	NEB
100 bp DNA ladder	NEB
<b>Protein standard:</b>	
Precision Plus Protein Standard (Unstained)	Bio-Rad

### 7.1.6 Materials and Consumables

General plastic consumables/disposables like pipette tips and serological pipettes Greiner Bio-One, Sarstedt GmbH, or Brand GmbH.

**Table 7-6 | Materials and Consumables**

<b>Materials/Consumables</b>	<b>Manufacturer</b>
Amicon Ultra-4 and -15 centrifugal filter units (3, 10 and 30, 100 kDa MWCO)	Merck Millipore
Centrifuge bottles Nalgene (PPCO) with sealing closure (30 mL, 500 mL, 1 L)	Thermo Fisher Scientific
Centrifuge tubes 15 and 50 mL	Greiner Bio-One
<b>Chromatography columns (empty, for gravity-flow or ÄKTA):</b>	
Econo-Column 2,5 × 10 cm (glass)	Bio-Rad
Econo-Pac 1,5 × 12 cm (plastic)	Bio-Rad
XK-50/20 column	GE Healthcare
<b>Filters (for sterile filtration):</b>	
Durapore membrane, PVDF, 0,22 µm	Merck Millipore
Syringe filters 0,2 µm Filtropur S 0.2	Sarstedt
Filtertips (Sapphire) 10, 20, 300, 1250 µL	Greiner Bio-One
Multiplate 96-Well PCR Plates, low profile, unskirted, white	Bio-Rad
PCR-tube stripes	A. Hartenstein
Pipette tips 10, 200, 1000 µL	Brand
Pipette tips GELoader® (0.5-20 µL)	Eppendorf
Pipette tips Gelloader (200 µL)	Sarstedt
Reaction tubes 1,5 and 2 mL	Sarstedt

Reaction tubes 1,5 mL DNA LoBind (low binding)	Eppendorf
<b>Resins and chromatography columns for ÄKTA-system:</b>	
Ni-NTA agarose	Qiagen
Macro-Prep High Q Resin	Bio-Rad
NHS-Activated Sepharose 4 Fast Flow Affinity Coupling Media	Cytiva
Strep-Tactin Superflow resin	IBA Lifesciences
Uno Q-1 (1.3 mL CV)	Bio-Rad
Resource Q (6mL CV)	GE Healthcare
Superdex 200 Increase 10/300 GL	GE Healthcare
Superose 6 10/300 column (GE Healthcare)	GE Healthcare

### 7.1.7 Equipment list

**Table 7-7 |** Equipment used in this work

<b>Equipment</b>	<b>Manufacturer</b>
ÄKTA pure chromatography system	GE Healthcare
CFX Connect Real Time PCR Detection System	Bio-Rad
Digital Rocker Wisemix RK-1D	Witeg Labortechnik GmbH
Digital tube roller Stuart SRT6D	Bibby Scientific
Gel documentation system Felix 1020	Biostep
High-speed image plate scanner CR-35 Bio (used as Phosphor-imager)	Raytest
Incubator Shaker New Brunswick Innova42	Eppendorf
Insect cell shaker, Multitron Pro (throw 50 mm)	Infors HT
Pipettes PIPETMAN Classic P2, P20, P200 and P1000	Gilson
Sonifier	Branson Ultrasonics
Spectrophotometer Eppendorf BioSpectrometer basic	Eppendorf
Sterile Bench HERAsafe KS18 1/PE AC Thermo Fischer Scientific	Thermo Fischer Scientific
Thermocycler peqSTAR 2X Universal	Peqlab
Water bath TW2	Julabo
<b>Centrifuges:</b>	
Eppendorf 5418 R (fixed angle rotor; for 1,5 mL tubes)	Eppendorf
Eppendorf 5810 R (swing-bucket rotor; for 15 and 50 mL tubes)	Eppendorf
Eppendorf Mini-spin (fixed angle rotor)	Eppendorf
Eppendorf 5418 (fixed angle rotor)	Eppendorf
Sorvall RC6 plus (Rotor: F10S-6x500y; F10-4x1000 LEX; SS-34)	Thermo Scientific
Ultracentrifuge Optima XPN (Rotor: Ti-45)	Beckman Coulter

**Gel electrophoresis equipment:**

Mini-PROTEAN Tetra system (for vertical PAA gels)	Bio-Rad
SE400 (18 x 16 cm) and SE-410 (18 x 24 cm) Air-Cooled Vertical Electrophoresis Unit	Hoefer

**Thermomixer:**

Eppendorf ThermoMixer C	Eppendorf
Eppendorf ThermoMixer F1.5	Eppendorf
Eppendorf Thermomixer comfort	Eppendorf
Water bath TW2	Julabo

## 7.2 Sequences

### 7.2.1 Oligos

All oligonucleotides (oligos) used in this thesis were purchased from Biomers.net GmbH (Ulm, Germany) or Eurofins Genomics Germany GmbH (Ebersberg, Germany).

**Table 7-8 |** PCR primer used to produce linear PCR templates for T7-mediated IVT

Primer	Sequence 5'- 3'	Application
RP0007	GCCAAGCTTTAATACGACTCAC	Universal forward primer for all eRNA productions (T7 Promoter specific)
ER0132	GCCAGTTAGAGGGTGGCGT	Reverse primer Arc eRNA 1-55
ER0009	GCTGGCCGATGAGACACC	Reverse primer Arc eRNA 1-100
ER0035	CCAGTCTGAGTGCCACCTA	Reverse primer Arc eRNA 1-200
ER0152	ACACAACAAAACCCTCTGTA ACT	Reverse primer Nr4a1-(a) 1-50
ER0153	GAGGTTTAAAGAAAGATTACCCAC	Reverse primer Nr4a1-(a) 1-100
LS0074	TTGAGTGTTTCCACATCTGTTGTC	Reverse primer Nr4a1-(a) 1-200
ER0154	AAGGAACGCGGCGGGC	Reverse primer Nr4a1-(b) 1-50
ER0155	CGAGCTCCCTTCCTGG	Reverse primer Nr4a1-(b) 1-100
LS0075	TGCGAAGTCGTAATTGCACTC	Reverse primer Nr4a1-(b) 1-200

**Table 7-9 |** Biotinylated antisense oligos and qPCR primer for 7SL RNA depletion and depletion analysis

Oligo_name	Sequence 5'- 3'	T <sub>m</sub> (°C)
<b>depletion oligos</b>		
anti SRP_ oligo1	Biotin/TACAGCCCAGAACTCCTGGACTCAAGCGATCCTCCTG	80.1
anti SRP_ oligo2	Biotin/ATCCCACTACTGATCAGCACGGGAGTTTTGACCTGCTC	79.1
anti SRP_ oligo3	Biotin/TCACCATATTGATGCCGAACCTAGTCCGGACACCCGATC	78.8
anti SRP_ oligo4	Biotin/CTATGTTGCCAGGCTGGAGTGCAGTGGCTATTACAG	79.9
<b>qPCR Primer</b>		
7SL_qPCR_f4 (forward)	ATCGGGTGTCCGCACTAAGTT	61.2
7SL_qPCR_r4 (reverse)	CAGCACGGGAGTTTTGACCT	60.5
Gapdh_qPCR_f1 (forward)	AGGTCGGTGTGAACGGATTTG	61.2
Gapdh_qPCR_r1 (reverse)	TGTAGACCATGTAGTTGAGGTCA	60.9

**Table 7-10 |** Oligos used for Exo-seq library construction (Modifications 5Phos = 5' phosphate; ddC = dideoxycytidine)

Oligos	Sequence 5'- 3'	Comment
3' Adapter (linker 1)	5Phos/TGGAATTCTCGGGTGCCAAGG/ddC	based on RA3 adapter (TruSeq Small RNA sample Kit, Illumina)
RT Primer	GCCTTGGCACCCGAGAATTCCA	RTP primer (TruSeq Small RNA sample Kit, Illumina)
5' Adapter (linker 2)	5Phos/GATCGTCCGACTGTAGAACTCTGAAC/ddC	based on RA5 adapter (TruSeq Small RNA sample Kit, Illumina)
forward PCR Primer (universal)	AATGATACGGCGACCACCGAGATCTACACGTTTCAGAG TTCTACAGTCCGA	RP1 primer (TruSeq Small RNA sample Kit, Illumina)
reverse PCR Primer (Index)	CAAGCAGAAGACGGCATAACGAGAT#####GTGACTG GAGTTCCTTGGCACCCGAGAATTCCA	RPI# Index primer (TruSeq Small RNA sample Kit, Illumina)

**Table 7-11 |** Oligos for *in vitro* SHAPE-MaP library construction

Oligos	Sequence 5'- 3'	Comment/Internal designation
SHAPE-MaP Primer for reverse transcription	GAACCGGACCGAAGCCC	RP0008
SHAPE-MaP 1 <sup>st</sup> -PCR forward Primer	GTTTCAGAGTTCTACAGTCCGACGATCGGCCATCT TCGGATGGCCA	with partial Illumina primer sequences
SHAPE-MaP 1 <sup>st</sup> -PCR reverse Primer	CCTTGGCACCCGAGAATTCCAGAACCGGACCGA AGCCCG	based on RA5 adapter (TruSeq Small RNA sample Kit, Illumina)
SHAPE-MaP 2 <sup>nd</sup> -PCR forward PCR Primer (same as for Exo-seq)	AATGATACGGCGACCACCGAGATCTA CACGTTTCAGAGTTCTACAGTCCGA	RP1 primer (TruSeq Small RNA sample Kit, Illumina)

SHAPE-MaP 2 <sup>nd</sup> -PCR reverse PCR Primer (same as for Exo-seq)	CAAGCAGAAGACGGCATACGAGAT#####GTG ACTGGAGTTCCTTGGCACCCGAGAATTCCA	RPI# Index primer (TruSeq Small RNA sample Kit, Illumina)
--	--	---

**Table 7-12** | Primer used for *in vivo* SHAPE-MaP libraries of *Gadd45b*, *Nr4a1-(a)*, *Fosb* and *Fos-e2* eRNA (1-250 and 150-400 eRNA regions). Reverse transcription (RT) primer and primer for the first PCR are designated with [1] and [2] for amplification of eRNA regions 1-250 nt and 150-400 nt, respectively. The comprise overhang sequences complementary to Illumina PCR primer and contain a randomized N7 (NNNNNNN) sequence used to sort out PCR duplicates.

Primer	Sequence 5'- 3'	Comment
ER0123	AGCAATCCTGCTTCAGCCTCC	RT primer [1] Gadd45b
ER0124	CTCAGAAGGAAGCCTGGTGG	RT Primer [2] Gadd45b
ER0125	GAAGGCCACTGGATTGGAGG	RT primer [1] Nr4a1-(a)
ER0126	TCATAGGTCCCTTCCTCAGC	RT Primer [2] Nr4a1-(a)
ER0129	TGCAGAGTGAATTCAGACAAC	RT primer [1] Fosb
ER0130	CAACCTGACCAAAAGTACAACC	RT primer [2] Fosb
ER0070	GAGAAGATAATGGCATTGCTGAGC	RT primer [1] Fos-e2 (alias 0090)
ER0120	TGCACAGAATTAGGACCAGACG	RT primer [2] Fos-e2 (alias 0090)
Gadd45b_N7_for_1	GTTTCAGAGTTCTACAGTCCGACGATCNNNNNNNGTTTTTGC GTATGCACTTGTAG	1 <sup>st</sup> -PCR primer forward [1] Gadd45b
Gadd45b_N7_rev_1	CCTTGGCACCCGAGAATTCCANNNNNNNAGCAATCCTGCTTC AGCCTC	1 <sup>st</sup> -PCR primer reverse [1] Gadd45b
Gadd45b_N7_for_2	GTTTCAGAGTTCTACAGTCCGACGATCNNNNNNNCCATGGAAT AACTAGCCCTGAG	1 <sup>st</sup> -PCR primer forward [2] Gadd45b
Gadd45b_N7_rev_2	CCTTGGCACCCGAGAATTCCANNNNNNNCTCAGAAGGAAGC CTGGTGG	1 <sup>st</sup> -PCR primer reverse [2] Gadd45b
0069_N7_for_1	GTTTCAGAGTTCTACAGTCCGACGATCNNNNNNNGCGGTGAG AGTGATTTTGTAAC	1 <sup>st</sup> -PCR primer forward [1] Nr4a1-(a) (alias 0069)
0069_N7_rev_1	CCTTGGCACCCGAGAATTCCANNNNNNNGAAGGCCACTGGA TTGGAGG	1 <sup>st</sup> -PCR primer reverse [1] Nr4a1-(a) (alias 0069)
0069_N7_for_2	GTTTCAGAGTTCTACAGTCCGACGATCNNNNNNNAGATGATCC TCAGTAGCAAGAAC	1 <sup>st</sup> -PCR primer forward [2] Nr4a1-(a) (alias 0069)
0069_N7_rev_2	CCTTGGCACCCGAGAATTCCANNNNNNNTCATAGGTCCCTTC CTCAGC	1 <sup>st</sup> -PCR primer reverse [2] Nr4a1-(a) (alias 0069)
Fosb_N7_for_1	GTTTCAGAGTTCTACAGTCCGACGATCNNNNNNNGAGGAGAC CCTGGAGCTGT	1 <sup>st</sup> -PCR primer forward [1] Fosb
Fosb_N7_rev_1	CCTTGGCACCCGAGAATTCCANNNNNNNTGCAGAGTGAATTC CAGACAAC	1 <sup>st</sup> -PCR primer reverse [1] Fosb
Fosb_N7_for_2	GTTTCAGAGTTCTACAGTCCGACGATCNNNNNNNCTGGCAGCT GTAGATAGATATC	1 <sup>st</sup> -PCR primer forward [2] Fosb
Fosb_N7_rev_2	CCTTGGCACCCGAGAATTCCANNNNNNNCAACCTGACCAAAA GTACAACC	1 <sup>st</sup> -PCR primer reverse [2] Fosb
0090_N7_for_1	GTTTCAGAGTTCTACAGTCCGACGATCNNNNNNNATTTCTGTA ACAGATCTTGGAGGC	1 <sup>st</sup> -PCR primer forward [1] Fos-e2 (alias 0090)

0090_N7_rev_1	CCTTGGCACCCGAGAATTCCANNNNNNNGAGAAGATAATGG CATTGCTGAGC	1 <sup>st</sup> -PCR primer reverse [1] Fos-e2 (alias 0090)
0090_N7_for_2	GTTACAGAGTTCTACAGTCCGACGATCNNNNNNNAGTCTTGCT CCTCGTTCTCAG	1 <sup>st</sup> -PCR primer forward [2] Fos-e2 (alias 0090)
0090_N7_rev_2	CCTTGGCACCCGAGAATTCCANNNNNNNTGCACAGAATTAGG ACCAGACG	1 <sup>st</sup> -PCR primer reverse [2] Fos-e2 (alias 0090)

**Table 7-13** | Oligos used for transcription bubble formation in EMSAs and Protein-RNA UV-crosslinking

Nucleic acids	Sequence 5'- 3'	Length (nt)	Internal designation
Template DNA (T-DNA)	CTGTAGACTGACCAAGTTGTCCCGTAGGAGTGTA GAGATATATGGTAG	49	GpSp2 T-DNA
Non-template DNA (NT-DNA)	CTACCATATATCTTACTCTACGCGGACA GGTCAGTCTACAG	49	GpSp2 NT-DNA
RNA (25)	UCUAGAACUAAAAAAAAUCUUACACUC	25	GpSp2 RNA 25
RNA (15)	UUUUUUUCUUACACUC	15	GpSp2 RNA 15

**Table 7-14** | Oligos used for transcription bubble formation in *in vitro* transcription assays. The Non-template DNA strand is biotinylated at the 5' end

Nucleic acids	Sequence 5'- 3'	Length (nt)	Internal designation
Template DNA (T-DNA)	GATCAAGCAGTAATCGTTGCGATCTGTAGACTGACCAAGT TGTCGGTAGGAGTGTAAGAGATATATGGTAGTACC	76	GpSp13 T-DNA
Non-template DNA (NT-DNA)	5Biotin-TEG/ GTCTGGTACTACCATATATCTTACTCTACGCGGACAA CTTGGTCAGTCTACAGATCGCAACGATTACTGCTTGATC	80	GpSp13 NT-DNA (biotin)
RNA (25)	Same as for EMSA experiment	25	GpSp2 RNA 25

**Table 7-15** | Repetitive, low complexity RNAs used for EMSAs (purchased from biomers GmbH)

RNA	sequence	Length (nt)
poly(A) <sub>80</sub>	(A) <sub>80</sub>	80
poly(C) <sub>80</sub>	(C) <sub>80</sub>	80
poly(U) <sub>80</sub>	(U) <sub>80</sub>	80
poly(CU) <sub>40</sub>	(CU) <sub>40</sub>	80
poly(GU) <sub>40</sub>	(GU) <sub>40</sub>	80
poly(GA) <sub>20</sub>	(GA) <sub>20</sub>	40



**Table 7-17 |** DNA template oligos for *in vitro* synthesis of *Nr4a1*-(a) G-to-A and G-to-C mutants (see Table 7-16 for description). The template oligos for the remaining G-mutants (see Figure 3-15D for sequence) are not listed but designed in the same manner as shown below.

Oligo (Name)	Sequence 5'- 3'	Comment
Nr4a1a_1-50 _Gless_plusT7	ATATAATAAAATTTTTGTAATTTTAGAGTTATAAAATTATTTTA TTGCCTATAGTGAGTCGTATTAATTTTC	Nr4a1-(a) 1-50 G-to-A mutant [template-strand]
Nr4a1a_1-50 _Gless_ss_plusT7	GAAATTAATACGACTCACTATAGGACAATAAAAATAATTTTATAA CTCTAAAATTACAAAAAATTTTATTATAT	Nr4a1-(a) 1-50 G-to-A mutant [non-template strand]
Nr4a1a_1-100 _Gless_plusT7	GGTTTAAAGAAAGATTATTTATAAATTTAAATTTAGGGTTTTGAT TAAAAATATAATAAAATTTTTGTAATTTTAGAGTTATAAAATTAT TTTTATTGCCTATAGTGAGTCGTATTAATTTTC	Nr4a1-(a) 1-100 G-to-A mutant [template-strand]
Nr4a1a_1-100 _Gless_ss_plusT7	GAAATTAATACGACTCACTATAGGACAATAAAAATAATTTTATAA CTCTAAAATTACAAAAAATTTTATTATATTTTTAATCAAAAACCCT AATTTTAATTTATAAATAATCTTTCTTTAAACC	Nr4a1-(a) 1-100 G-to-A mutant [non-template strand]
Nr4a1a_100_GtoC	GGTTTAAAGAAAGATTAGGGAGAAATGAAAAGGAGGGGTGTT GATTA AAAAGAGAAGAAAAGGGTGTGTAAGTTTAGAGTTAGAA AATGAGTGTGAGGGCCTATAGTGAGTCGTATTAATTTTC	Nr4a1-(a) 1-100 G-to-C mutant [template-strand]
Nr4a1a_100_GtoC_ ss	GAAATTAATACGACTCACTATAGGCCCTCACACTCATTTTCTAA CTCTAAACTTACACACCCTTTTCTCTCTTTTAAATCAACACCCT CCTTTTCATTTCTCCCTAATCTTTCTTTAAACC	Nr4a1-(a) 1-100 G-to-C mutant [non-template strand]

**Table 7-18 |** Vector- and insert-specific primers used to clone eRNAs (1-200 nt) into pUC18 plasmids, that were needed for general eRNA production (not for SHAPE-MaP).

Primer	Sequence 5'- 3'	Comment
ER0039	TTGTCTTCGAATTCGTAATCATGG	forward primer linearization of pUC18
ER0065	CTATAGTGAGTCGTATTAAGCTTG	reverse primer linearization of pUC18
LS0067	GCTTTAATACGACTCACTATAGGCGGTGAGAGTGATTTTGT AACT	eRNA_1_0069a f alias: eRNA_Nr4a1-(a)
LS0068	ATTTGGATCCGGCGAACCGGATCGATAGGGAGTTCAAGG CCAGC	eRNA_1_0069a r alias: eRNA_Nr4a1-(a)
LS0069	CACTATAGGGCCATCTTCGGATGGCCAAGAAGGTGAGGG GGGATTGG	eRNA_1_0069b f alias: eRNA_Nr4a1-(b)
LS0070	ATTTGGATCCGGCGAACCGGATCGACGTGAGGATTTAGGA GGAATCA	eRNA_1_0069b r alias: eRNA_Nr4a1-(b)
ER0033	ATGTGCTGCAAGGCGATTAAG	Universal forward primer for colony PCR and Sanger sequencing (for pUC18 plasmids)
ER0034	TTTACACTTTATGCTTCCGGCTC	Universal reverse primer for colony PCR and Sanger sequencing (for pUC18 plasmids)

**Table 7-19 |** Vector- and insert-specific primers used to clone eRNAs (1-200 nt; flanked with linker sequences) into pUC18 plasmids for the SHAPE-MaP experiment. The overhang of the primers is highlighted in bold. Forward (f); lense (r) primer; [\_1\_] = eRNA in sense direction/(+) strand; [\_2\_] = eRNA in antisense direction/(-) strand

Primer	Sequence 5'- 3'	Application
RP0002	GTTCCGCCGATCCAAATCGGGCTTCGGTCCGGTCTTGTCTTCGAATTCGT AATCATGG	initial linearization of pUC18 f primer
RP0022	TTGGCCATCCGAAGATGGCCCTATAGTGAGTCGTATTAAGCTTG	initial linearization of pUC18 f primer
RP0015	GATCCGGTTCGCCGGATC	linearization of pUC18 (with already cloned in linker sequence) f
RP0023	CATCCGAAGATGGCCCTATAG	linearization of pUC18 (with already cloned in linker sequence) r
RP0024	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> AGAGCGCGGGAAGAACA	eRNA_1_0364 f
RP0025	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TAGGGAGTTCAAGGCCAGC	eRNA_1_0364 r
RP0026	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GAAGGTGAGGGGGATTGG	eRNA_2_0376 f
RP0027	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CGTGAGGATTTAGGAGGAATCA	eRNA_2_0376 r
RP0028	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GCTTTGAGGCTGAGTCTCAC	eRNA_1_0242 f
RP0029	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TTACAATGCCCTTACCGGATAG	eRNA_1_0242 r
RP0030	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> AAAACACAGCCGCCAGGGT	eRNA_1_Arc f
RP0031	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CCAGTCTGAGTGCCACC	eRNA_1_Arc r
RP0032	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> ATTTCTGTAACAGATCTTGGAGGC	eRNA_1_0090 f eRNA_Fos_e2
RP0033	<b>TTTGGATCCGGCGAACCGGATCGA</b> TGACACTAAAATACCGAGCTGTAAC	eRNA_1_0090 r eRNA_Fos_e2
RP0034	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> CGTACGTGAGCGCGCGG	eRNA_2_0117a f eRNA_Junb-(a)
RP0035	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GGAAGGTCTTCCCAGCAGG	eRNA_2_0117a r eRNA_Junb-(a)
RP0036	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> AGTGCTGCGGAACGAGAGA	eRNA_2_0117b f eRNA_Junb-(b)
RP0037	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GCTGTTCCATGGCAACGG	eRNA_2_0117b r eRNA_Junb-(b)
RP0038	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> GCGGTGAGAGTGATTTTGTAAAC	eRNA_1_0069a f eRNA_Nr4a1-(a)
RP0039	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TTGAGTGTTTCCACATCTGTTGTC	eRNA_1_0069a r eRNA_Nr4a1-(a)
RP0040	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> AGTAGGGGCATGGCTCGC	eRNA_1_0069b f eRNA_Nr4a1-(b)
RP0041	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TGCGAAGTCCGTAATTGCACTC	eRNA_1_0069b r eRNA_Nr4a1-(b)
RP0042	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> AGCTGGTCTCTGAATCTCC	eRNA_1_0146 f

RP0043	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CACCCATGAGATTTTGCAGATTTG	eRNA_1_0146 r
RP0044	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GTCACTGCTGCCTCCTCC	eRNA_2_0145 f
RP0045	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GAACAAGTGCCCTGGGCT	eRNA_2_0145 r
RP0046	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GCAGTTTTCTTCTGGGCTTGAG	eRNA_1_0194 f
RP0047	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CATTTATGCTTGCTGAAGAGGATG	eRNA_1_0194 r
RP0048	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GGAAGGCAGAGAAGGAGAC	eRNA_2_0449 f
RP0049	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CTGTATTTTGTAGGAGACAGAGC	eRNA_2_0449 r
RP0050	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GAGGAGACCCTGGAGCTGTA	eRNA_1_Fosb f
RP0051	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CACCGAGGGGGGAAAAAAGG	eRNA_1_Fosb r
RP0052	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GTTTTTGCGTATGCACTTGTAG	eRNA_1_Gadd45b f
RP0053	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TTCAGGTTAAGCTGTGCTGG	eRNA_1_Gadd45b r
RP0054	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> AGCAGTGTGAGCGAATGCTGG	eRNA_2_0526 f
RP0055	<b>ATTTGGATCCGGCGAACCGGATCGA</b> AACATGGATCCCCTTACCTAGC	eRNA_2_0526 r
RP0058	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GAGTCAGAGCTGCTTTTCCTT	eRNA_1_0206-(a) f
RP0059	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TCTAAGATCCCAAAGTATCTGAG	eRNA_1_0206-(a) r
RP0060	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> AGGAGTCTGACAGAACCAGGA	eRNA_1_0206b f
RP0061	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TCCCAGACAACATACTGTAGTAAC	eRNA_1_0206b r
RP0064	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> TGTCTGCAGGAGCGAGCTGAC	eRNA_2_0124b f
RP0065	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GCAAGCCGAAAGAAAGCTTCG	eRNA_2_0124b r
RP0066	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GCAGGGACGCCGAGACC	eRNA_2_0124c f
RP0067	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CCTGTACTTCATTATCCACCTG	eRNA_2_0124c r
RP0068	<b>TAGGGCCATCTTCGGATGGCCAA</b> GCCACTGCTTCTCTCTGA	eRNA_1_0014 f
RP0069	<b>AAGCCCGATTTGGATCCGGCGAACCGGATCGA</b> TTTGGCAGCAAGGTGGCA AATC	eRNA_1_0014 r
RP0070	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GCCGTGCGATGCAACATTGA	eRNA_1_0133 f
RP0071	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TGGGGGAATATTACTACCAG	eRNA_2_0133 r
RP0072	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GGCAAAGGCACACTCTGG	eRNA_2_0367 f
RP0073	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TTCCCAGGAGTGTGCCCAAG	eRNA_2_0367 r
RP0074	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> CTTAGTTATTTAATCCTCAGTGACC	eRNA_1_0643 f
RP0075	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GTTTGGTTGACTTTCTTTCATTTACC	eRNA_2_0643 r
RP0076	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GGGACTTCTGGTCCCCAC	eRNA_2_0160a f
RP0077	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CTCTCGATCTTCTAACTTGAGA	eRNA_2_0160a r
RP0078	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> AGTTCTGTCTAAGTCTGTCT	eRNA_2_0310a f
RP0079	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CCAGCGTGCTGCCACGTAT	eRNA_2_0310a r
RP0080	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> GTGTGGACTTTGCAATGCGTTG	eRNA_2_0098 f
RP0081	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CATCCCCTCTAAATGCCACAC	eRNA_2_0098 r
RP0082	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> AGGGATCTGGTTAAGGCTGTG	eRNA_1_0254a f

RP0083	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GTCCAGCAAACATTATAACCC	eRNA_1_0254a r
RP0084	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> AAGGCTGTGTTTCAGCCTGGG	eRNA_1_0254b f
RP0085	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TTTCCCCTGATTGTCCCAGCA	eRNA_1_0254b r
RP0086	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> GGATCAGAAATACTAGAATTCTGAG	eRNA_1_0558 f
RP0087	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GACACAGGTAAGGCTTTACTC	eRNA_1_0558 r
RP0088	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> AGAGAGTAAAAGACCCTGGCAG	eRNA_1_0165a f
RP0089	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GTGTCTGGTGGTCAAAGTCCT	eRNA_1_0165a r
RP0090	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> AAATGGGGAAGACTGCAGAGAG	eRNA_1_0165b f
RP0091	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GTCCTGATCATCTCCCATACAC	eRNA_1_0165b r
RP0092	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> CGTGTGCAACTGCTTGAAGT	eRNA_2_0139a f
RP0093	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CCTTATGAGTCTCCCTTCCTTC	eRNA_2_0139a r
RP0094	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> ACAGAATCTCCAGGAGCCC	eRNA_1_0250a f
RP0095	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TAGAGAGGCAGAAAGAGACAAGA	eRNA_1_0250a r
RP0096	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> GTTAAAAAACTATCTCCAGGCCG	eRNA_1_0659a f
RP0097	<b>ATTTGGATCCGGCGAACCGGATCGA</b> TCTGTGGGGGCCATGGATG	eRNA_1_0659a r
RP0098	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> ACTTCCGGCAAGCGTTGCTC	eRNA_2_0216a f
RP0099	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GGAGCCCACGAAAGGTCAG	eRNA_2_0216a r
RP0100	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> GAGAAAAGCTGGACTTATACACC	eRNA_1_0202 f
RP0101	<b>ATTTGGATCCGGCGAACCGGATCGA</b> AGTTTGTTCATCTGAATCGTTGG	eRNA_1_0202 r
RP0102	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> GGCTCTTAGGAGCCAAACCT	eRNA_1_0479 f
RP0103	<b>ATTTGGATCCGGCGAACCGGATCGA</b> ACCAGTTCAACCGTGAGGCA	eRNA_1_0479 r
RP0104	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> TGGAGGCTGAGTTTGTAGGG	eRNA_2_0301a f
RP0105	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GGGAAGGCCAAAAACAAAACAAAAGC	eRNA_2_0301a r
RP0106	<b>ACTATAGGGCCATCTTCGGATGGCCAA</b> AGTTCACGCTGTATTCTC	eRNA_2_0301b f
RP0107	<b>ATTTGGATCCGGCGAACCGGATCGA</b> CACGGGTTGAGAATCCATATTC	eRNA_2_0301b r
RP0108	<b>CACTATAGGGCCATCTTCGGATGGCCAA</b> ACTGGAGAAAGGAAAAGCGAAGT	eRNA_2_0248a f
RP0109	<b>ATTTGGATCCGGCGAACCGGATCGA</b> GTATTGTATGGTGTGCAGGGGA	eRNA_2_0248a r

### 7.3 Software

The software used in this thesis is listed in the Table below.

**Table 7-20 |** Software used in this study

Software	Reference	Source
BEDTools	(Quinlan and Hall, 2010)	<a href="https://bedtools.readthedocs.io/en/latest/index.html">https://bedtools.readthedocs.io/en/latest/index.html</a>
R Package v3.5.1	R Project	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
deepTools	(Ramírez et al., 2016)	<a href="https://deeptools.readthedocs.io/en/develop/index.html">https://deeptools.readthedocs.io/en/develop/index.html</a>
STAR	(Dobin et al., 2013)	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
ShapeMapper 2	(Busan and Weeks, 2018)	<a href="https://github.com/Weeks-UNC/shapemapper2">https://github.com/Weeks-UNC/shapemapper2</a>
RNAstructure software package (graphical and text interface) and Structure Editor	(Reuter and Mathews, 2010)	<a href="https://rna.urmc.rochester.edu/RNAstructure.html">https://rna.urmc.rochester.edu/RNAstructure.html</a>
cutadapt (v 1.15)	(Martin, 2011)	<a href="https://github.com/marcelm/cutadapt">https://github.com/marcelm/cutadapt</a>
SortMeRna (v 2.1)	(Kopylova et al., 2012)	<a href="https://github.com/biocore/sortmerna">https://github.com/biocore/sortmerna</a>
TSScall	(Henriques et al., 2018)	<a href="https://github.com/lavenderca/TSScall">https://github.com/lavenderca/TSScall</a>
gffread	(Pertea and Pertea, 2020)	<a href="https://github.com/gpertea/gffread">https://github.com/gpertea/gffread</a>
CLC Main Workbench 7	QIAGEN	
ImageJ (v 2.0.0-rc-69)	(Rueden et al., 2017)	<a href="https://imagej.net/ImageJ2_development_releases">https://imagej.net/ImageJ2_development_releases</a>
PyMOL (v 1.8.4.2)	Schrodinger LLC	<a href="https://pymol.org">https://pymol.org</a>
Prism 9 (v 9.0.0)	GraphPad Software	<a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a>
UCSC Genome Browser	(Kent et al., 2002)	<a href="http://genome.ucsc.edu">http://genome.ucsc.edu</a>
UCSC Table Browser	(Karolchik et al., 2004)	<a href="http://genome.ucsc.edu/cgi-bin/hgTables?org=Mouse&amp;db=mm10">http://genome.ucsc.edu/cgi-bin/hgTables?org=Mouse&amp;db=mm10</a>
IGV Browser (v 2.3.91)	(Robinson et al., 2011)	<a href="http://software.broadinstitute.org/software/igv/download">http://software.broadinstitute.org/software/igv/download</a>

## 7.4 Supplementary Data

### 7.4.1 RNA sequences

**Table 7-21** | Sequences of the cloned set of 39 eRNA (1-200 nt)

cloned eRNA ID	sequence
eRNA_0014	GCCACUGCUUUCUCUCUGACCCCCUUGGGAGUCGGGAGUUGCCUUCUUGUGCUUGAAGUACUUUUUCUAGCCAGAAGAUGCUGGCUAGUCAUGGCCUGC CAUGACUUUGAAGGCUAACUUUUUUCUGUUCACAGUGGUCGAGGAAAGUCCUCGCUUUCACUGCCUACCUCAGGAGAUUUGCCACCUUGCUGCCAAA
eRNA_Nr4a1-(a)	GCGGUGAGAGUGAUUUUGUAAUCUCUAAAAGUUACAGAGGGUUUUGUUGUUUUAAUCAAGAGCCCUGGUUUUGAUUUGUGGGUAAUCUUUCUUAAACCU CGCCCAUUUCCGGCCUGGUGACCUGGGCCACUUUCUCAUGUUGACCCAGAUGAUCCUCAGUAGCAAGAACAAAAGGACAACAGAUGUGGAAACACUCA
eRNA_Nr4a1-(b)	AGUAGGGGCAUGGCUCGCCCAGGCUCGGUGCUGCCCCGCCGCCGCGUUCUUCGGCUUGGUGUUCGGUGGGGGACCAGGGAAGGGAAGCUCGG GCCAAAUCCGGUCUCGGGCGAGGUGACAGAAAUUUGGAAUCUGACACCCGGAGGAUGGCUGAGCUCGGCUCUUCUGGGAGUGCAAUUACGGACUUCGCA
eRNA_Fos e2	AUUUCUGUAAACAGAUUCUUGGAGGCUCGGUCUGGAUCCUCGCCAAGAACCAGAUCCAGGAGAAAACGUGCUC AACGUGCAGCUCUGCUCCUACUGAUUUA AGCCCCACAGAUGACAUCGCUCCAUAUGUCACACCAAGUCUCCUGUGGGAGUCUUGCUCUUCGCUUCACAGUGUCUGUACAGCUCGGUAUUUUAGUGUCA
eRNA_0133	GCCGUCGCAUGCAACAUAUGAUUGUCUGUUAUAUUACAUUUAUUAAAGAAUGAUAAACUUGUAAAUAUUUGGAUGUAAAUAUUAAAGAACUCUCUCCUAGUUAUU GCCGUCUCAAUUAUAACUGUCGCUGGCUGGCAGUUAUAUAUGAUUAUAUGAGGGUUGUAUUUUUAUGCAGCCGACCGUGACCUGGUGAGUAAUAUUCGCCCA
eRNA_0146	AGCUGGUCCUCUGAAUCUCCGGAGCGACCUGCUACAUCACUCAGUCCUGUCUCUCCCUUAUUCUACUCCCUCCUGUGGAAUUUGUUUUUCUUGUUAUGUUG CCGAGAGAUGCAACCAUCAGCUCUCCGGUUGUCCUGGUUUUCUUAUCCUGUUCUCUGACUUUUUUUUUUUUUUUUUUUGACACAAAUCUGCAAAAUCUCA UGGGUG
eRNA_0165-(a)	AGAGAGUAAAAGACCCUGGCAGCAAUGGUCUGUGACACGGAUGAGAGUAAGACUCAGAAGUGACGAGAUGUCACAACAAAUAACUUAACACAACUGGAUAG AGACGCUGCCCACAGUGAGAACACAGCAGCAUAAUGGAGAUAGAAGGCCGACUGGGCCAGGAGUGUAUGGGAGAUGAUCAGGACUUUGACCACCAGACAC
eRNA_0165-(b)	AAAUGGGGAAGACUGCAGAGAGUAAAAGACCCUGGCAGCAAUGGUCUGUGACACGGAUGAGAGUAAGACUCAGAAGUGACGAGAUGUCACAACAAAUAAC UUACACAACUGGAUAGAGACGCUGCCCACAGUGAGAACACAGCAGCAUAAUGGAGAUAGAAGGCCGACUGGGCCAGGAGUGUAUGGGAGAUGAUCAGGAC
eRNA_0194	GCAGUUUUUCUUGGGCUUGAGCAGUAGGUGGACAAUCAUACAGUCCUGCGAACCACACAUUUUAGAGGGUGUGCCUACGGUAUGCAGUACUAGUGUGCU UCCACCUCUAAUGUAGCACACAAAAGGGAACAUACUUGUUGUUCUGGUUGGUUUUCAAAACAGCUUUUCAUCCUCUUCAGCAAGCAUAAAUG
eRNA_0202	GAGAAAAGCUGGACUUUAACACCCCAUUGAAAAGGCUUAUCUGGUCAGGAGAAAAAUCCUGGGCUAGGCUUGCCUUGCUGGGAGGGAAGAACCAAGCUGA GGCUGCUUAGGUGCCCUCGGUCAUCUGGAACAGUUCACAUGGCUCUUGAGGCAUGAUUUGAUGGGAGCCAACGAUUCAGAUGAGAACAACU
eRNA_0206-(a)	GAGUCAGAGCUCUUUUUCUUUUUAUUGCCAGCUCACAGUGGGAACUUGACCUUGAAAACCCUGGCUCUGUAGAUUUUGACGUAAGGAUUUAACGUCCA CAUUUCUUGAUUGGUGUUGACUAGGUUAGACCCAUCCCUUCCCAUGUUUAGUUACUACAGUAUGUUGUCUGGGACCUCAGAUAUUUGGGAUCUUAGA

---

eRNA_0206-(b)	AGGAGUCUGACAGAACCAGGAAGUGAGUCAGAGCUGCUUUUCCUUUUUUAUUGCCCAGCUCACAGUGGGAACUUGACCUUGAAACCCUGGCUUCUGUAGAUAU UUGACGUAAGGAUUAUAACGUCCACAUUUCUUGAUUGGUGUUGACUAGGUUAGACCCAUCCCUUCCCAUGUUUAGUUACUACAGUAUGUUGUCUGGGA
eRNA_Fos e1	GCUUUGAGGCUGAGUCUCACUCUUGAGCAUGCUGUCAUAAUAUGUCAUCAGCUCAGACACACACACACACACACAGAGCUCUUGGGGAAUGAGGAGAGG UGUGUGCUGAGGAGGAGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUUCUGCCAGCUUGACUAUACUAUCCGGUAAGGGCAUUGUA
eRNA_0250	ACAGAAUCUUCAGGAGCCCCAGGCGCAGUUAGGAGGGGGGAUAUCGCGGCGGAAUAAGGAAACCGAGGCACGCAGACACCCCGCCCCGGGGGAGAUU GCAACAAGGGCUCUUGAGCUGACUGCAAGGUAGCAUGCCGCAACUUCAGCUUGGAUCUCUCCCCUCCCCAACUCCUCUUGUCUCUUUCUGCCUCUCUA
eRNA_0254-(a)	AGGGAUCUGGUUAAGGCUGUGUUCAGCCUGGGCUACAGAGGAGGAGAUCUUGUUUAAAACGCAAGACAAAUCAAAACAAGAAAAACCAACGAGGGGGAGAGA AAAGGGUCAGGCAUCAGGAUUAUACAUGUAUGUAGUAGAUCUAAAGUGGAAAUUUCUGGUUUUUUAAGACUUGGGUUUAUAUGUUUUGCUGGGAC
eRNA_0254-(b)	AAGGCUGUGUUCAGCCUGGGCUACAGAGGAGGAGAUCUUGUUUAAAACGCAAGACAAAUCAAAACAAGAAAAACCAACGAGGGGGAGAGAAAAGGGUCAGGC AUCAGGAUUAUACAUGUAUGUAGUAGAUCUAAAGUGGAAAUUUCUGGUUUUUUAAGACUUGGGUUUAUAUGUUUUGCUGGGACAAUCAGGGGAAA
eRNA_0364	AGAGCGCGGGAAGAACACUAAUUUAGGCCAAAAAUGGGUUCAGGGUCGCCUGUGAAAGUGGCCUGGAUUGGGUGGAAAAGGGUGAGGUGUCGCGCCUGAGG CCCAGAAGGUGGGAUAGGCGUCUCUAUGAAGAUUCACACAUUUUUUUUUUCCUUUUGGGACAGAAGCCCCAGUAGCCCAGGCUGGCCUUGAACUCCCUA
eRNA_0558	GGAUCAGAAAUACUAGAAUUCUGAGGCUAAGGGCACCCAUGAGAAUGCAUGCACACUGUGAAGGUCCACAGAGAGAUUAGUAGGCAGUUUAUGGAGACACA AGAGUGGUUAUACCAGCUUUGCAUAGCUCAGGAGGGGCAAGAAGUCCUCUCACUCUUCGGAGGUAUCCUCUGAUUCAGCGAGUAAAGCCUUACCUGUGUC
eRNA_0643	CUUAGUUUUUUAAUCCUCAGUGACCCUAGCAAAGAAGCUUCGAGGCUCACAGCAAAGUAGCACCUCACAGCCUACAUUUGUCAGAUAAAGGAAACUGAAAUUG GGAGCAGUUAGUUCUUCAGGAUCACACAUCGAGGAGUGAAAUUAGAUUAAAAGGUUAUUGGUUUUUUAUGUGGUAAAUGAAAGAAAGUCAACCAAAC
eRNA_0659	GUUAAAAACUAUCUCCAGGCCGGAGGGGAGAAGAGGAAGCCGCAGGGGCCCCACAGGGCAGAGGACUGCUCUUGCCCUGGAGCUGUAGGGUGGGAAGGAG AGAGGGACUGGAACCUUGGUUCCUCUGGGUACCACAGUCACUGGGUGGGAAAGGGGUUCUUCUGUAUUUUGAGGCUACUUCAUCCAUGGCCCCACAGA
eRNA_Fosb	GAGGAGACCCUGGAGCUGUAAGAAGAAGAAGAAAAAACAAGUCCUUCUGACUACCGGGACAGUCAGUGGAUUUGGAGGCUGGGUUAAAGUUUAC CCCGUUCUCAAAGCGGAGACAGAAUACCAGGUUGCAGCUGGCAGCUGUAGAUAGAUUCUUUAUUUCUUUUUCCUUCUUUCCUUUUUCCCCCUCGG UGU
eRNA_Gadd45b	GUUUUUUGCGUAUGCACUUGUAGAAAAGUCUUUUAAUAUGCGACCCUCAUUUCACGCAUGGAGAAACUGAAGCUGGGUCCUGCCUACCUUACCAGCUCAU GGUUGGAAAUUAUUAUUAUCCCUAGCCCGUAAAACUGGGUCCAGCGCCUCCAUGGAAUAACUAGCCCUGAGUUCUCUCCCCAGCACAGCUUAAACCUGAA
eRNA_0098	GUGUGGACUUUGCAAUGCGUUGCCAAGGCAGUGAGGCCUGCUGGGUGAGUUGGCACAGGGCAGGGAAGGCAGACAAAACCCACACUGCAUCCAGAAACC ACCACUCACAGCACAGCAGGAGAGACGGGGCUGCAGCUGUAAUCUCUGCUGGGGUGGGGGGUGGCAGCGGUGAGGGGUGUGGCAUUUAAGAGGGGAUG
eRNA_Junb-(a)	CGUACGUGAGCGCGGGGGUAGCCGCGGGAGCGCUGGGGCCAGUCUGCGGAACGAGAGACUUUGGAGCGCAGGUGAAACUGGCAGGGGACUGCCGUUA ACCCCUUCCUGCCCAGCCGAAACUGGGGCCCCAGGGCCGAACGAACGGGGCCUCCCCAGCUCGCUGGGCGGGGACAGUCCUCUGGGGAAGACCUUCC
eRNA_Junb-(b)	AGUGCUGCGGAACGAGAGACUUUGGAGCGCAGGUGAAACUGGCAGGGGACUGCCGUUAACCCCUUCCUGCCCAGCCGAAACUGGGGCCCCAGGGCCGAAC GAACGGGGCCUCCCCAGCUCGCUGGGCGGGGAGUCCUCUGGGAAGACCUUCCUCUUGUGAGCCGAAACACCCCCCCCGUUGCCAUGGAAACAGC

---

---

eRNA_Fos e5-(a)	UGUCUGCAGGAGCGAGCUGACCUGGAGCAAAGGGCCGGAAGCUGGGAAGGGUUAGGUGACACCUUGGCCUUGGCAGAGCAGGGUGAGGGCAUCGGGAGAGU GCGGAGAGGGAGGGUGCAGGGACGCCGAGACCCCUUGGGAGGAGCUUGGGAAGCCGGCAGAGAGGAGCUUUUCUACCCAUCGAAGCUUUCUUCGGCUUGC
eRNA_Fos e5-(b)	GCAGGGACGCCGAGACCCCUUGGGAGGAGCUUGGGAAGCCGGCAGAGAGGAGCUUUUCUACCCAUCGAAGCUUUCUUCGGCUUGCAAGGUUGUAGAGGACUU GGUUUUCUUACCAGCCGCUAGCAUGCAGUCGUUGGGGACACUUUACCCUGGGCUCAGAUUCCAGGGCCGCCACUCCAGGUGGAUAAUGAAGUACAGG
eRNA_0139	CGUGUGCAACUGCUUGGAAGUGAGGCUGAGCUGGAGUGUUCAGGAUCUGUUAGGGAGAGACCCGGUAGAGUGGUAUUCUAAGAGAGAUAGAGGGAGCCCCG GGAGUCUGGAAGGGUUGAAGCACCCUAAGCUUCCUCUGGCUGCCCAGGGGCAAUGAGAUGCUCGGGGUGAACUGGAAGAAGGAAGGGAGACUCAUAAGG
eRNA_0145	GUCACUGCUGCCUCCUCCUCCUAGCUGCAAGGACCUCAGCUUCAUGGGUUUACUGACACUCUGCUCAGAAGCCUUUCUACCACUUAUCAUCUCCGAGAGA CAGCAUGGCCGCAGUGUCUCCCAGCUGCCAUCUCCUUCCCGCAUGGCUUCUCCAGACUACUGAAAGCAUGCCCCGAGAGUCAGGGUGGCAGCCCAGGGGAC ACUUGUUC
eRNA_0160	GGGACUUCUGGUCCCCACGCGGGUCCUGUAUUGCGUUCUUUAUGUAGCUCGUUCUCCUCUCAGCCCCUGGGCGCUCCGCUAGAAGGAACGGACAGAGAGUGU CAGGAUGUAGUAGAACCUCGCGCAAGAGGCUCGGGCUCUUCGCUCUUGGCUUUAAGUUCUGUGGUCCCCGUAAGGUUUCUCAAGUUAGGAAGAUCGAGAG
eRNA_0216	ACUUCCGCAAGCGUUGCUCUGUGGGAGGCGUUCUGGGUGGUCCACCCCCGAGAAGCGUUCACAGCGUCCCGGCCGAGGCGAGGCGGGUCUAGUCCCCUCG CAGGUCUCCGAGUCAGAAUUUCCCCUCGCGCCAGCGUGGCCUUAAAGGUCAGAGCUCGCCACUCUCCGUUCGCGCCGACCUGACCUUUCUGGGGUCC
eRNA_0248	ACUGGAGAAAAGGAAAAGCGAAGUGGGGUGUUAGAAAACUAUAGGAAACAGGAAGAAUUGCAGAGACAUUCAGAAGGACUGCAAAAAGGCCGACGAAGUGGCUCA GCGGAUAAAAGGUGCUUCGGUACCAGCCGGAGAAGCAACUCCAGAGUUAUCCUCUGAGCUUCACACAUAACAUCUCUCCCCUGCACACCAUACAUAUC
eRNA_0301-(a)	UGGAGGCUGAGUUUGUAGGGCGCGCUGGUAGGGGCACGAAAACAAAAACUGUGGGACUGGAAGUGACUGCUAGUUCCAGCCUGUAUUCUCCUCAGGUCA GAUCUGUCCAGCCUUGACUCUAAAUUGACAGUGUAAAAGAGAGAAGAGAGGGCGGCGAAAACGUAAAUUUGUUCGUCUUCGGGCAUUGCUUUUGUUUGUU UUUGCCUUC
eRNA_0301-(b)	AGUUCAGCCUGUAUUCUCCUCAGGUCAGAUUCUGUUCUCCAGCCUUGACUCUAAAUUGACAGUGUAAAAGAGAGAAGAGAGGGCGGCGAAAACGUAAAUUUG UUCGUCUUCGGGCAUUGCUUUUGUUUUUGCCUUCUCCUCCUAGAGCCUCCUGUAUUUCUCUUGUGGGUUCAUACUAUAGAAUUGGAUUCUCAACCC GUG
eRNA_0310	AGUUCUGUCCUAGUCCUGUUCUUGGCGGACCCCAGCAGGGCGCUCUGCUGUAGUAGCAGCAAUUAGAAAACGGAAAUAAGAUUGGAAGCUCGGGAUUA GCCCAGGGACUGCGGCAGAGGACUGACUUCAAAAGCAUUUAGAUAAAUGGCAUGACCUGUUCAGCCUAUCAAGACCCGAUACGUGGCAGCACGCUGG
eRNA_0376	GAAGGUGAGGGGGGAUUGGAGGGGAGAUGGGAAGGCUUAAAGGCACAGGACCCUCAUAUCCUUUAAAAAUGUGACCAGAGAUCUGCUCUUUUUAUCCACA CUGUCUCACUUCUAAUUUCAAAUGGGGCUUAGGAGUGGACCUAUUAGGUCUGCUCUGCAGUAAACCAAUCCUCUUCUGAUUCCUCCUAAAUCCUCACG
eRNA_0449	GGGAAGGCAGAGAAGGAGACGGGUAGAGUCCAGCCUGGAACAGAGAAAAGAACACGGGGUGCAGGGAGGGCAGAAGGACCCAUAAAUUCUUCUGCCACUUG UUAGACCGCCCCUCUGACGUCACCCCCCCCCCGCCUCUGGCAGGGUGGGGGUUGUGGAGGGUGAGCUCUUGGCAAGCUCUGUCUCCUACAAAAUA CAG
eRNA_0526	AGCAGUGUGAGCGAAUGCUGGCUUUUCUAAAAGUUCUGCUACACCCAAGGCGGUCAGAAGAACCUCGGCAAGUUCGGGCUUGCUAUGGCCGACUGUGAUUC ACUUCAGAGGGAAACACAGAAAGAUUCCAGCCUGCUCUCACAAUUUGGAGGCAUCCCAAUAGUGACUCAAAAGCACCCUGCUAGGUAAACGGGAUCCAUGUU

---

---

eRNA_Arc	AAAAACACAGCCGCCAGGGUGGGGGUGGCAGCCUUGAACGCCACCCUCUAAACUGGCAUGGGCAUCUGAACUCUCAGCUCUGGGGGUGUCUCAUCGGCCAGCU GGCAGUAGCCCCCCCCAACCCACCCUCGAUGGAGAACUGCUCUCUGUAUGAUCCAGCCCAGUGGGCGCACAGGGAGUUGGUAGGUGGGCACUCAGACUGG
----------	---

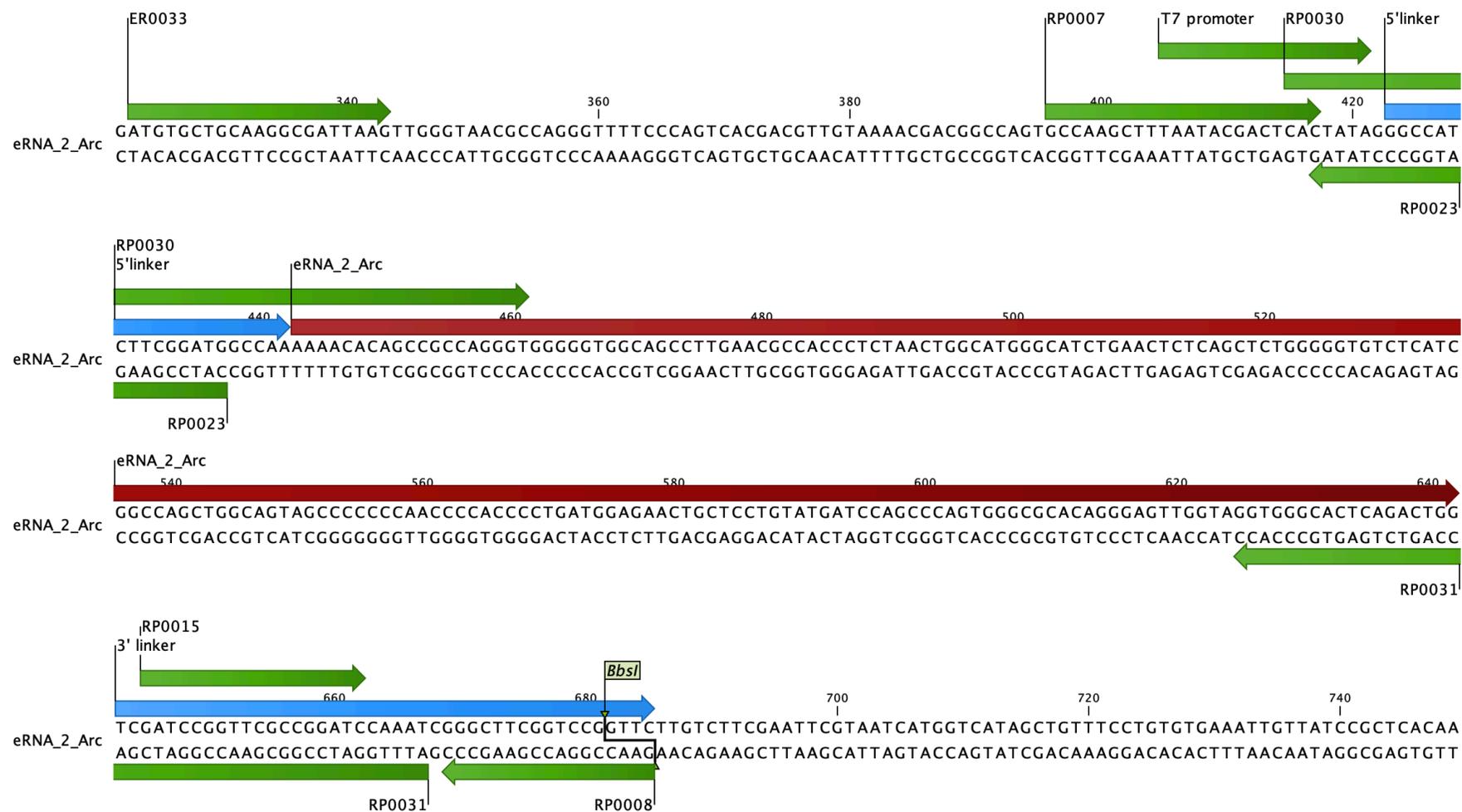
---

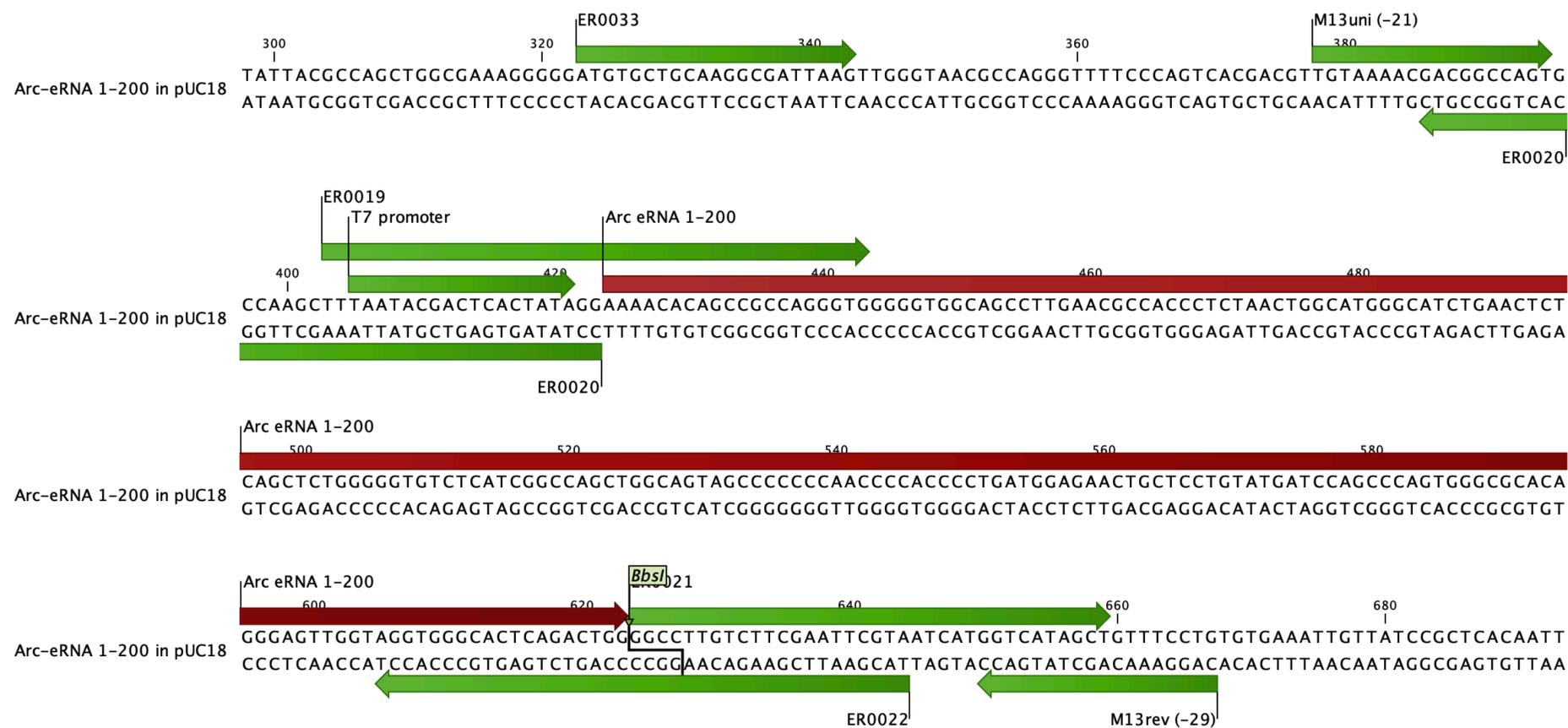
**Table 7-22 | Sequences of the mutant eRNAs (mutant eRNA sequences are cloned in pUC18 vectors)**

eRNA mutant	sequence
<i>Arc</i> $\Delta$ stem mutant (100 nt)	AAAACACACCACCAGAAUGAAAUAAAGCAGCCUUGAACUUACCCCGUACGCGCACCCAGCUGGCAGUAGCCCCCCCCAACCCACCCUCGAUAGCCC AGUG
<i>Nr4a1</i> -(a) $\Delta$ loop 1 mutant (170 nt)	GCGGUGAGAGUGAUUUUGUAACUCUAAAGUUACAGAGGGUUUGUUGUGUUUUAAUCAAGAGCCCUACCUCGCCC AUUUCGGCCUGGUGACCUG GGCCACUUUCUCAUGUUGACCCAGAUGAUCCUCAGUAGCAAGAACAAAAGGACAACAGAUGUGGAAACACUCAA
<i>Nr4a1</i> -(a) $\Delta$ loop 12 mutant	GCGGUGAGAGUGAUUUUGUAACUCUAAAGUUACAGAGGGUUUGUUGUGUUUUAAUCAAGAGCCCUACCUCGCCCUGGGCCACUUUCUCAUGUUG ACCCAG

## 7.4.2 pUC18 plasmid with cloned eRNA Insert

**Figure 7-1** | View of the *Arc* eRNA 1-200 insert sequence cloned into a pUC18 plasmid for eRNA production required for *in vitro* SHAPE-MaP (screenshot from CLC Main Workbench file). eRNA insert and flanking regions are displayed, all relevant primer binding regions etc. are annotated. All other eRNAs are cloned in the same manner.



**Figure 7-2 |** View of the *Arc* eRNA 1-200 insert sequence cloned into a pUC18 plasmid; not for SHAPE-MaP purposes (see Figure 7-1 for description)

### 7.4.3 Exo-seq and SHAPE-MaP supplementary data

**Table 7-23** | Final list of 39 eRNAs extracted from the Exo-seq data and the overlap with GRO-seq data. GRO-seq ID originates from the GRO-seq peak list provided by Dr. Seung-Kyoon Kim. eTSS\_IDs for the two KCl-treated Exo-seq libraries (Rep1 and 2) are the IDs of extragenic transcription start site (TSS) peaks called with TSScall.

cloned eRNA ID	GRO-seq ID	nearby gene	chr	strand	start	activity-induced (>1.5 fold change in GRO-seq)	eTSS_ID Rep1	eTSS_ID Rep2
eRNA_0014	chr19-243-0	Scyl1	chr19	+	5812393	yes	uTSS_271430	uTSS_275552
eRNA_Nr4a1-(a)	chr15-6853-0	Nr4a1	chr15	+	101246858	yes	uTSS_200653	uTSS_203771
eRNA_Nr4a1-(b)	chr15-6853-0	Nr4a1	chr15	+	101246579	yes	alternative TSS (not max reads)	alternative TSS (not max reads)
eRNA_Fos e2	chr12-6079-0	Fos	chr12	+	85455592	yes	uTSS_122651	uTSS_125250
eRNA_0133	chr5-8146-0	E130006D01Rik	chr5	+	111725130	yes	uTSS_430799	uTSS_437394
eRNA_0146	chr6-8296-0	Slc2a3	chr6	+	122761259	yes	uTSS_473851	uTSS_480777
eRNA_0165-(a)	chr6-8298-0	Slc2a3	chr6	+	122770015	yes	uTSS_473859	
eRNA_0165-(b)	chr6-8298-0	Slc2a3	chr6	+	122769999	yes		uTSS_480785
eRNA_0194	chr13-5893-0	Tmem161b	chr13	+	84344878	no	uTSS_150673	uTSS_153746
eRNA_0202	chr14-1970-0	Galnt15	chr14	+	31934333	no	not max reads	uTSS_167489
eRNA_0206-(a)	chr6-9242-0	Apold1	chr6	+	134994423	yes	uTSS_475609	uTSS_482499
eRNA_0206-(b)	chr6-9242-0	Apold1	chr6	+	134994399	yes	alternative TSS (not max reads)	alternative TSS (not max reads)
eRNA_Fos e1	chr12-6076-0	Fos	chr12	+	85436393	yes	uTSS_122647	uTSS_125247
eRNA_0250	chr7-1610-0	Ercc1	chr7	+	19333093	yes	not max reads	uTSS_488248
eRNA_0254-(a)	chr12-5899-0	Sipa11	chr12	+	82470842	yes	uTSS_121176	
eRNA_0254-(b)	chr12-5899-0	Sipa11	chr12	+	82470854	yes		uTSS_123809
eRNA_0364	chr15-5010-0	Trappc9	chr15	+	73067871	yes	uTSS_190798	uTSS_193566

## Chapter 7 – Supplement

eRNA_0558	chr8-1572-0	Tacc1	chr8	+	25382470	yes	uTSS_521506	uTSS_528081
eRNA_0643	chr2-11371-0	Ube2v1	chr2	+	167634784	yes	uTSS_330236	uTSS_335436
eRNA_0659	chr1-9565-0	Optc	chr1	+	133887290	yes	uTSS_024571	uTSS_025441
eRNA_Fosb	chr7-1607-0	Fosb	chr7	+	19320277	yes	uTSS_481386	uTSS_488242
eRNA_Gadd45b	chr10-5528-0	Gadd45b	chr10	+	80940083	yes	uTSS_054100	uTSS_055004
eRNA_0098	chr11-1902-1	Tob1	chr11	-	94160586	no	uTSS_699276	uTSS_708159
eRNA_Junb-(a)	chr8-3011-1	Junb	chr8	-	84986118	yes	uTSS_1129082	
eRNA_Junb-(b)	chr8-3011-1	Junb	chr8	-	84986075	yes		uTSS_1144065
eRNA_Fos e5-(a)	chr12-2413-1	Fos	chr12	-	85485987	yes	uTSS_726030	uTSS_735406
eRNA_Fos e5-(b)	chr12-2413-1	Fos	chr12	-	85485871	yes	alternative TSS (not max reads)	alternative TSS (not max reads)
eRNA_0139	chr7-9314-1	Zfp324	chr7	-	12961532	yes	uTSS_1075525	uTSS_1090498
eRNA_0145	chr5-2551-1	Mn1	chr5	-	111473604	yes	uTSS_1026275	uTSS_1040057
eRNA_0160	chr7-9315-1	Zfp324	chr7	-	12949505	yes	uTSS_1075516	uTSS_1090495
eRNA_0216	chr4-310-1	Park7	chr4	-	150914467	no	uTSS_1000230	uTSS_1013976
eRNA_0248	chr3-5223-1	Mef2d	chr3	-	88138356	yes	uTSS_949327	uTSS_961952
eRNA_0301-(a)	chr13-7065-1	Zkscan3	chr13	-	21408444	yes		uTSS_745603
eRNA_0301-(b)	chr13-7065-1	Zkscan3	chr13	-	21408373	yes	uTSS_736088	
eRNA_0310	chr3-4578-1	Anp32e	chr3	-	95951029	no	uTSS_951731	uTSS_964400
eRNA_0376	chr9-5233-1	Bcl9l	chr9	-	44457425	no	uTSS_1150319	uTSS_1165410
eRNA_0449	chr5-2106-1	Med13l	chr5	-	118422227	yes	uTSS_1029106	uTSS_1042888
eRNA_0526	chr7-8706-1	Igsf23	chr7	-	19933602	yes	uTSS_1077551	uTSS_1092549
eRNA_Arc	chr15-1944-1	Arc	chr15	-	74679213	yes	uTSS_798674	uTSS_809197

**Table 7-24** | *In vitro* SHAPE-MaP sequencing statistics showing read depth and mutation rates of DMSO and 1M7 samples for each eRNA individually

cloned eRNA ID	DMSO median depth	DMSO 5th percentile depth	DMSO 95th percentile mut rate (%)	DMSO median mut rate (%)	1M7 median depth	1M7 5th percentile depth	1M7 95th percentile mut rate (%)	1M7 median mut rate (%)	Sequencing Run
eRNA_0014	226,770	195,541	0.65	0.06	257,402	225,653	1.43	0.11	Run1
eRNA_Nr4a1-(a)	77,934	70,676	0.61	0.05	89,330	82,111	1.24	0.08	Run1
eRNA_Nr4a1-(b)	1,675,612	1,167,810	1.3	0	1,577,482	1,130,906	2.09	0.05	Run2
eRNA_Fos e2	65,155	50,786	0.47	0.04	55,431	44,600	0.95	0.06	Run1
eRNA_0133	1,166,158	675,651	0.74	0.05	1,204,846	721,476	1.97	0.05	Run2
eRNA_0146	651,138	4,876	1.01	0.05	788,133	15,791	2.08	0.1	Run2
eRNA_0165-(a)	1,095,838	751,523	0.94	0.06	1,299,301	880,832	3.13	0.12	Run2
eRNA_0165-(b)	1,665,080	700,458	0.53	0.05	1,873,591	809,769	2.6	0.11	Run2
eRNA_0194	279,758	226,725	0.59	0.05	305,237	250,615	1.14	0.09	Run1
eRNA_0202	48,726	40,035	0.64	0.05	59,097	49,418	1.14	0.11	Run1
eRNA_0206-(a)	842,700	675,766	1.04	0.08	776,239	626,361	2.24	0.15	Run2
eRNA_0206-(b)	755,587	510,494	1.01	0.07	852,397	611,773	2.05	0.14	Run2
eRNA_Fos e1	1,273,463	783,156	0.75	0.04	1,390,111	841,325	1.87	0.05	Run2
eRNA_0250	992,333	376,892	1.46	0.07	1,169,925	494,006	2.55	0.17	Run2
eRNA_0254-(a)	1,448,790	803,019	0.72	0.04	1,555,944	864,091	1.96	0.11	Run2
eRNA_0254-(b)	993,967	711,776	0.74	0.05	1,295,387	935,753	1.81	0.11	Run2
eRNA_0364	35,241	19,107	0.76	0.04	41,014	22,303	1.06	0.08	Run1
eRNA_0558	148,871	124,301	0.46	0.04	116,560	98,635	1.07	0.07	Run1
eRNA_0643	29,538	25,086	0.45	0.04	30,678	25,069	1.08	0.08	Run1
eRNA_0659	1,663,003	1,002,546	2.06	0.07	1,906,790	1,183,397	2.57	0.12	Run2
eRNA_Fosb	40,991	25,008	1.41	0.03	36,086	20,991	1.94	0.04	Run1
eRNA_Gadd45b	82,357	70,022	0.57	0.04	82,573	71,180	1.1	0.08	Run1
eRNA_0098	1,072,788	325,353	0.75	0.05	1,409,834	462,817	2.44	0.07	Run2

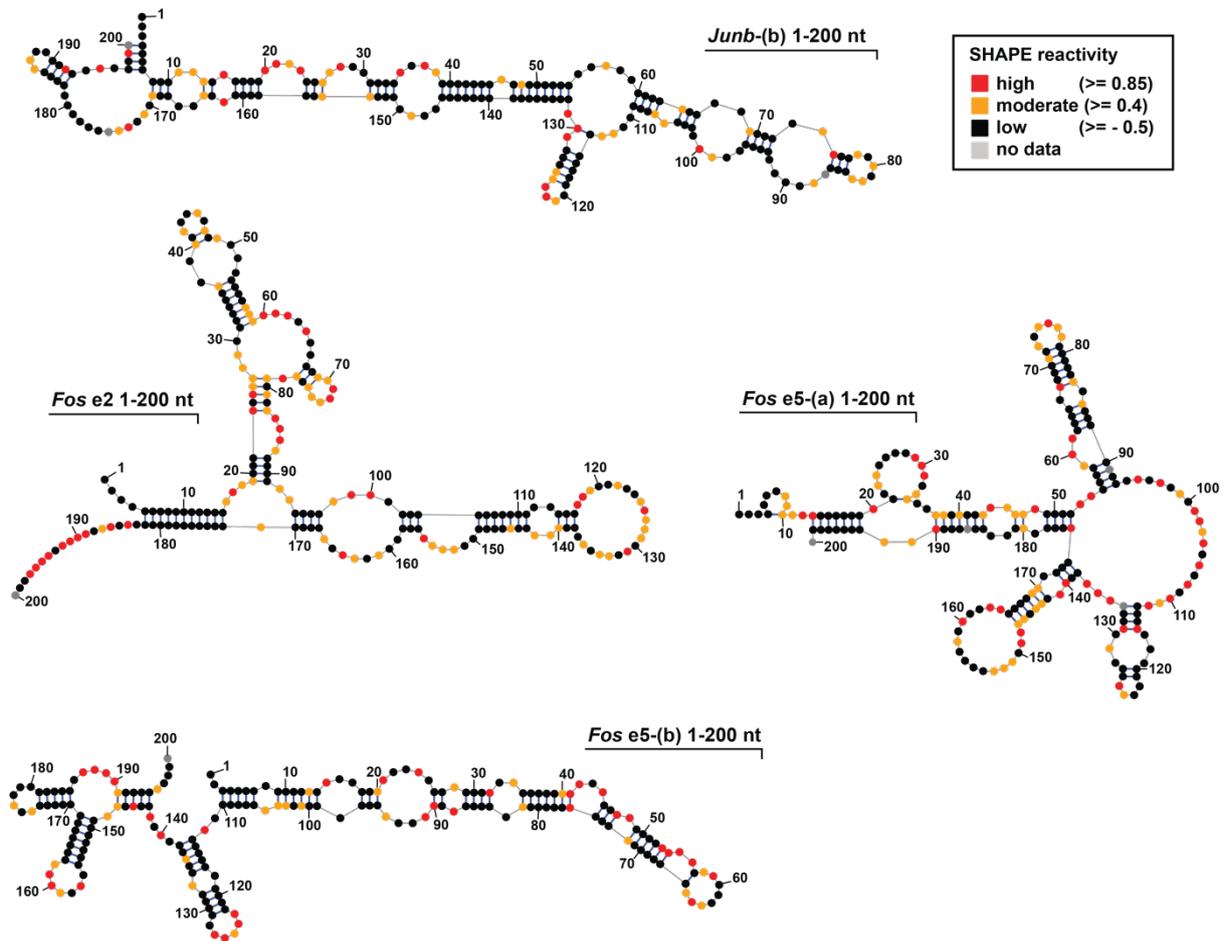
## Chapter 7 – Supplement

eRNA_Junb-(a)	1,417,926	1,020,288	1.71	0.07	1,996,751	1,450,014	2.2	0.14	Run2
eRNA_Junb-(b)	878,882	652,571	1.2	0.06	917,252	746,435	1.75	0.11	Run2
eRNA_Fos e5-(a)	15,603	8,903	2.13	0.07	48,300	26,937	4.13	0.15	Run2
eRNA_Fos e5-(b)	43,619	31,925	0.53	0.04	62,982	46,875	0.93	0.09	Run1
eRNA_0139	74,543	52,585	0.53	0.04	91,890	68,335	0.83	0.08	Run1
eRNA_0145	100,482	72,004	0.73	0.06	149,754	106,980	1.15	0.09	Run1
eRNA_0160	59,596	40,594	0.53	0.04	75,081	52,054	1.08	0.08	Run1
eRNA_0216	1,305,975	943,264	1.16	0.09	1,732,463	1,271,112	2.58	0.15	Run2
eRNA_0248	102,355	81,480	0.56	0.04	86,933	71,727	0.97	0.09	Run1
eRNA_0301-(a)	818,250	241,997	0.82	0.05	975,429	340,273	2.72	0.07	Run2
eRNA_0301-(b)	747,529	560,539	0.82	0.05	1,223,138	1,025,321	2.22	0.11	Run2
eRNA_0310	54,453	49,006	0.51	0.04	55,643	50,401	0.98	0.08	Run1
eRNA_0376	1,315,491	788,017	1.51	0.05	1,200,016	790,430	2.07	0.12	Run2
eRNA_0449	1,290,303	729,873	2.34	0.06	1,446,313	829,971	2.29	0.11	Run2
eRNA_0526	90,427	80,053	0.44	0.05	80,649	72,398	1.03	0.08	Run1
eRNA_Arc	1,453,732	876,823	1.34	0.07	1,345,769	836,112	1.86	0.1	Run2

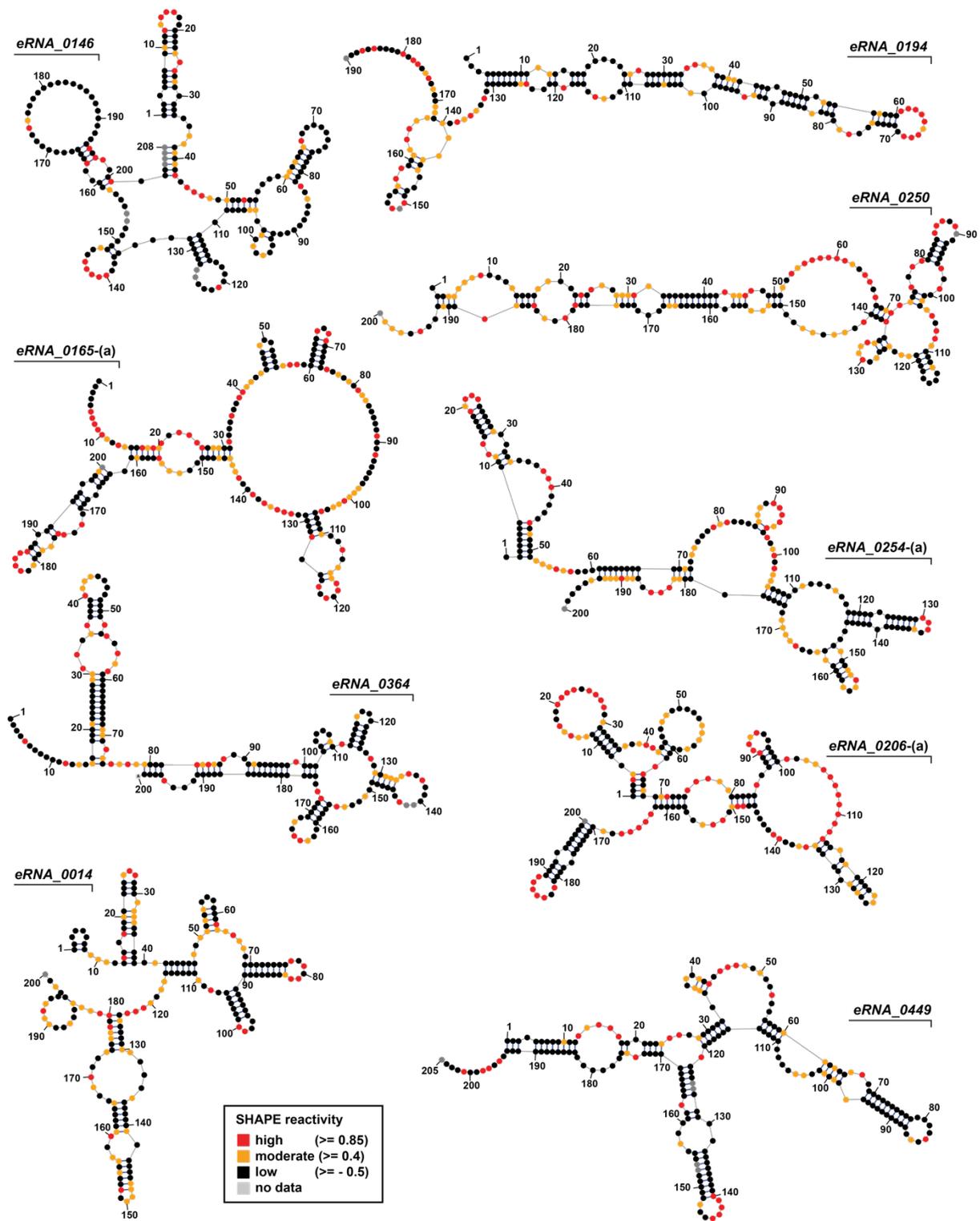
**Table 7-25** | *In vivo* SHAPE-MaP sequencing statistics showing read depth and mutation rates of DMSO and 1M7 samples for each eRNA individually

eRNA 1-400 nt	DMSO median depth	DMSO 5th percentile depth	DMSO 95th percentile mut rate (%)	DMSO median mut rate (%)	1M7 median depth	1M7 5th percentile depth	1M7 95th percentile mut rate (%)	1M7 median mut rate (%)	Condition
Nr4a1-(a)	385,648	191,314	0.77	0.03	273,289	155,640	0.86	0.04	Nuclei
Nr4a1-(a)	385,652	191,271	0.77	0.03	183,513	67,225	0.67	0.04	Run-on
Fos e2	473,983	337,242	0.58	0.03	333,010	201,059	0.58	0.04	Nuclei
Fos e2	473,983	337,242	0.58	0.03	339,641	225,109	0.58	0.04	Run-on
Fosb	78,059	16	0.76	0	136,750	0	0.72	0	Nuclei
Fosb	78,059	16	0.76	0	128,990	2	0.73	0	Run-on
Gadd45b	311,751	155,499	0.64	0.03	191,209	94,318	0.57	0.03	Nuclei
Gadd45b	311,751	155,485	0.64	0.03	204,733	121,477	0.65	0.03	Run-on

**Figure 7-3** | Additional SHAPE-MaP structures for eRNAs of the immediate early genes *c-Fos* and *Junb* (corresponding to Figure 3-8).



**Figure 7-4 |** Additional SHAPE-MaP structures for selected eRNAs from the set of 39 eRNAs (see Table 7-24).



### 7.4.4 Protein-RNA crosslinking data

**Table 7-26** | List of Protein-RNA crosslinks for the experiment with *Arc* eRNA 1-200. 1<sup>st</sup> column: crosslinked peptide, the crosslinked residue within the peptide and the additional mass of the crosslinked nucleotides (C=323; U=324; A=347; G=363; CU=629; UU=630; AU=653; GU=669); 2<sup>nd</sup> column (Protein1): crosslinked protein from the PEC; 3<sup>rd</sup> column (AbsPos1): Position of crosslinked residue in the protein. (nseen = number of recorded crosslinks). The table is sorted by the Id.Score. Data were recorded and provided by Dr. Alexander Leitner. Crosslinked complexes were previously assembled by me.

Peptide-Residue-Nucleotide	Protein1	AbsPos1	Error_rel. ppm.	nseen	Id.Score
MHGGGPPSGDSACPLR-G3-630	RPB1_HUMAN	3	-1.6	1	43.46
FAVALDSEQNNIHVK-A2-324	SPT5H_HUMAN	584	-1.6	1	42.75
MHGGGPPSGDSACPLR-G5-630	RPB1_HUMAN	5	-2.5	1	42.41
FAVALDSEQNNIHVK-V3-324	SPT5H_HUMAN	585	-2.2	1	41.5
FAVALDSEQNNIHVK-F1-324	SPT5H_HUMAN	583	-2.8	2	40.4
FHPKPSDLHLQTGYK-H2-324	RPB1_HUMAN	432	-3	3	39.83
FAVALDSEQNNIHVK-F1-324	SPT5H_HUMAN	583	-2	2	39.66
FHPKPSDLHLQTGYK-F1-324	RPB1_HUMAN	431	-2.4	4	39.38
FAVALDSEQNNIHVK-A2-324	SPT5H_HUMAN	584	-1.7	4	38.36
FAVALDSEQNNIHVK-F1-324	SPT5H_HUMAN	583	-1.2	3	38.04
ETFQVLNMYGK-Y9-324	SPT5H_HUMAN	564	-3.3	1	37.61
ETFQVLNMYGK-M8-324	SPT5H_HUMAN	563	-2.6	1	36.81
FHPKPSDLHLQTGYK-H2-630	RPB1_HUMAN	432	-2.7	2	36.66
FAVALDSEQNNIHVK-A2-324	SPT5H_HUMAN	584	-2.6	2	36.62
FHPKPSDLHLQTGYK-H2-324	RPB1_HUMAN	432	-3	1	36.26
MHGGGPPSGDSACPLR-M1-630	RPB1_HUMAN	1	-2	3	35.81
MHGGGPPSGDSACPLR-H2-630	RPB1_HUMAN	2	-1.8	1	35.53
FAVALDSEQNNIHVK-F1-324	SPT5H_HUMAN	583	-1.7	2	35.37
GPIQILNR-I5-630	RPB2_HUMAN	1069	-1.6	3	35.33
ETFQVLNMYGK-Y9-324	SPT5H_HUMAN	564	0	1	35.32
GPIQILNR-I5-324	RPB2_HUMAN	1069	-2.6	1	35.01
ARGPIQILNR-R2-630	RPB2_HUMAN	1064	-2.8	1	34.93
ARGPIQILNR-I5-630	RPB2_HUMAN	1067	-1.9	7	34.71
FHPKPSDLHLQTGYK-F1-630	RPB1_HUMAN	431	-2.6	1	34.71
GVAYKSR-Y4-653	SPT4H_HUMAN	108	0.2	1	34.23
ARGPIQILNR-R2-324	RPB2_HUMAN	1064	-1.9	1	34.03
ARGPIQILNR-P4-630	RPB2_HUMAN	1066	-2.5	1	33.88
ARGPIQILNR-R2-630	RPB2_HUMAN	1064	-2.5	3	33.62
ARGPIQILNR-I5-630	RPB2_HUMAN	1067	-2.4	2	33.29
FHPKPSDLHLQTGYK-H2-324	RPB1_HUMAN	432	-1.5	4	33.12
ARGPIQILNR-G3-630	RPB2_HUMAN	1065	-2.4	2	33.08
ARGPIQILNR-A1-324	RPB2_HUMAN	1063	-3.5	1	32.97
GPIQILNR-I5-324	RPB2_HUMAN	1069	-2	1	32.92
FHPKPSDLHLQTGYK-H2-630	RPB1_HUMAN	432	-2.7	3	32.88

ARGPIQILNR-G3-630	RPB2_HUMAN	1065	-2.8	4	32.84
FHPKPSDLHLQTYGK-F1-324	RPB1_HUMAN	431	-2.5	2	32.73
GPIQILNR-I5-630	RPB2_HUMAN	1069	-2.2	2	32.73
ISPWLLR-W4-363	RPB1_HUMAN	1210	0.2	1	32.71
LLVDSNNPK-V3-347	RPB1_HUMAN	119	-0.6	1	32.52
FAVALDSEQNNIHVK-V3-324	SPT5H_HUMAN	585	-2.3	3	32.46
TLLRK-L3-323	NELFA_HUMAN	248	-4.1	1	32.44
GPIQILNR-I5-630	RPB2_HUMAN	1069	-2.2	3	32.33
KQSSSSTTSQGGVK-Q2-323	NELFE_HUMAN	35	-0.3	3	32.25
LLVDSNNPK-V3-347	RPB1_HUMAN	119	0.3	1	32.08
LLVDSNNPK-V3-347	RPB1_HUMAN	119	0.6	1	31.86
FAVALDSEQNNIHVK-V3-324	SPT5H_HUMAN	585	-1.9	1	31.57
TLLRK-L3-323	NELFA_HUMAN	248	-3.9	2	31.52
GVAYKSR-S6-653	SPT4H_HUMAN	110	0.3	1	31.39
MHGGGPPSGDSACPLR-M1-324	RPB1_HUMAN	1	-2.7	1	31.39
FHPKPSDLHLQTYGK-H2-630	RPB1_HUMAN	432	-2.4	4	31.33
ARGPIQILNR-I5-324	RPB2_HUMAN	1067	-2.3	3	31.26
FHPKPSDLHLQTYGK-H2-630	RPB1_HUMAN	432	-2.9	2	31.18
FHPKPSDLHLQTYGK-H2-630	RPB1_HUMAN	432	-3.1	2	31.09
LFDAEKIR-A4-323	SPT5H_HUMAN	342	-3.5	1	31.05
FHPKPSDLHLQTYGK-H2-324	RPB1_HUMAN	432	-2.8	1	30.95
QGEVVQRLTRXVGK-T9-323	NELFB_HUMAN	205	-4	1	30.92
ARGPIQILNR-G3-324	RPB2_HUMAN	1065	-1.7	3	30.82
ITTPYMTKYER-T7-323	RPAB2_HUMAN	58	-1.1	1	30.8
GPIQILNR-I5-324	RPB2_HUMAN	1069	-3	1	30.75
FAVALDSEQNNIHVK-A4-324	SPT5H_HUMAN	586	-3.1	2	30.69
LKELINISK-I5-324	RPB1_HUMAN	1128	-2	1	30.38
GPIQILNR-Q4-324	RPB2_HUMAN	1068	-3.5	1	30.38
ETFQVLNMYGK-Y9-324	SPT5H_HUMAN	564	-1.7	3	30.37
IKYAKEVLQK-E6-363	RPB2_HUMAN	341	0.1	3	30.35
GPIQILNR-I5-630	RPB2_HUMAN	1069	-1.6	1	30.24
KQSSSSTTSQGGVK-Q2-323	NELFE_HUMAN	35	0.8	1	30.15
ARGPIQILNR-G3-630	RPB2_HUMAN	1065	-2	2	30.15
KQSSSSTTSQGGVK-Q2-323	NELFE_HUMAN	35	0.4	3	30.1
GPIQILNR-I5-324	RPB2_HUMAN	1069	-1.8	1	30.04
KNLSLTR-R7-323	NELFA_HUMAN	435	-3.6	2	30.04
GPIQILNR-P2-630	RPB2_HUMAN	1066	-1.6	3	29.92
FAVALDSEQNNIHVK-F1-324	SPT5H_HUMAN	583	-1.6	1	29.92
ARGPIQILNR-I5-630	RPB2_HUMAN	1067	-2.3	1	29.82
ARGPIQILNR-P4-630	RPB2_HUMAN	1066	-2.1	2	29.76
NLSLTREQMFAAQEMFK-F16-323	NELFA_HUMAN	445	0.1	1	29.75
FHPKPSDLHLQTYGK-H2-324	RPB1_HUMAN	432	-2.4	1	29.69
YQPRIRR-I5-323	RPB1_HUMAN	191	-1.8	2	29.68

LLALGQGAWDXIDSQVFK-V16-669	NELFB_HUMAN	359	0.6	1	29.63
RVSINKDELK-S3-323	NELFD_HUMAN	377	0.7	2	29.63
XLLAHVDLIEK-K11-363	RPAB5_HUMAN	58	-3.2	1	29.61
GPIQILNR-I5-630	RPB2_HUMAN	1069	-2.1	1	29.55
MLLAHVDLIEK-I9-363	RPAB5_HUMAN	56	-2.9	1	29.37
MHGGGPPSGDSACPLR-G5-630	RPB1_HUMAN	5	-2.5	1	29.33
ETFQVLNMYGK-Y9-324	SPT5H_HUMAN	564	-3	2	29.29
MHGGGPPSGDSACPLR-H2-630	RPB1_HUMAN	2	-2.7	2	29.24
GPIQILNR-Q4-630	RPB2_HUMAN	1068	-2	2	29.2
MHGGGPPSGDSACPLR-H2-630	RPB1_HUMAN	2	-2.3	1	29.15
DPEQLMNTLRKLR-M6-347	RPB2_HUMAN	584	0.4	2	29.02
SLYESFVSSDR-V7-363	NELFE_HUMAN	137	-0.1	1	28.94
ELINISKPK-I5-324	RPB1_HUMAN	1130	0	1	28.9
HMCDGDIVIFNR-N11-324	RPB1_HUMAN	459	-3.4	1	28.86
TLPHFIKDDYGPEFR-F5-363	RPB1_HUMAN	810	0.6	1	28.82
DVTNFTVGGFAPMSR-P12-324	SPT5H_HUMAN	664	-3	2	28.66
KQSSSSTTSQGGVK-S9-323	NELFE_HUMAN	42	0.9	1	28.52
KQSSSSTTSQGGVK-T7-323	NELFE_HUMAN	40	0.9	1	28.43
LKHMVDDK-D6-363	RPB2_HUMAN	1056	0.4	1	28.38
ITINKDTKVPK-V9-629	RPB1B_HUMAN	26	-0.4	2	28.37
LKIVGTRVDK-T6-323	RPB7_HUMAN	150	-2.8	1	28.35
ERFVDSK-V4-324	NELFB_HUMAN	275	0	3	28.25
QSLVDXAPK-L3-669	RPAB1_HUMAN	118	0.3	1	28.2
AXQKSGRPLKSLK-G6-363	RPB1_HUMAN	333	-0.4	7	28.17
ARGPIQILNR-R2-630	RPB2_HUMAN	1064	-2	1	28.15
QMDIIVSEVMIRDIR-I5-629	RPB2_HUMAN	597	-1.3	1	28.12
RQXDIIVSEVMIR-I6-323	RPB2_HUMAN	597	-2	1	28.04
QGEVVQRLTRXVGK-T9-323	NELFB_HUMAN	205	0.1	1	27.77
CSFEETVDVLMEEAAHGESDPMK-A13-324	RPB1_HUMAN	1442	-0.6	1	27.72
LLDISELDXVAGR-V10-363	NELFA_HUMAN	265	0.1	1	27.68
FAVALDSEQNNIHVK-V3-324	SPT5H_HUMAN	585	-1.8	2	27.49
HMCDGDIVIFNR-N11-324	RPB1_HUMAN	459	-2.9	1	27.4
HMCDGDIVIFNR-V8-324	RPB1_HUMAN	456	-2.1	1	27.36
ARGPIQILNR-R2-324	RPB2_HUMAN	1064	-2.3	1	27.34
ACQALGAXLSKGALNPADITVLFK-S10-363	NELFD_HUMAN	301	-3.3	1	27.26
LPSDLHPIK-P2-323	RPB1_HUMAN	1001	-2.6	1	27.24
DPEQLMNTLRKLR-M6-347	RPB2_HUMAN	584	0.2	2	27.23
LYDXVLQFLR-V5-363	NELFB_HUMAN	218	0.4	1	27.07
FAVALDSEQNNIHVK-V3-324	SPT5H_HUMAN	585	-0.7	2	27.03
QMDIIVSEVMIRDIR-I5-629	RPB2_HUMAN	597	-3.4	2	27.02
EQMFAAQEMFKTANK-A13-363	NELFA_HUMAN	448	-3.8	1	26.92
FAVALDSEQNNIHVK-F1-324	SPT5H_HUMAN	583	-3.3	2	26.92
EQMFAAQEXFKTANK-A5-347	NELFA_HUMAN	440	-3.4	1	26.89

IKYAKEVLQK-E6-363	RPB2_HUMAN	341	-1.4	3	26.85
NLSLTREQMFAAQEMFK-F16-323	NELFA_HUMAN	445	-2	1	26.78
AYAKKGFGK-K5-323	RPB3_HUMAN	167	-1.7	1	26.77
QAXGVYITNFHVR-Q1-324	RPB2_HUMAN	731	1	1	26.77
NLSLTREQMFAAQEMFK-F16-323	NELFA_HUMAN	445	-3.6	1	26.73
LKIVGTRVDK-T6-323	RPB7_HUMAN	150	-3.1	1	26.67
ERFVDSK-V4-324	NELFB_HUMAN	275	-1.2	1	26.6
ELINISKPK-S6-324	RPB1_HUMAN	1131	-0.8	1	26.47
ITINKDTKVPK-V9-629	RPB1B_HUMAN	26	0.3	2	26.45
FAVALDSEQNNIHVK-F1-630	SPT5H_HUMAN	583	-3	1	26.44
MATNTVYVFAKK-V8-324	RPB2_HUMAN	207	-2.4	1	26.44
XAGILFEDIFDVK-G3-363	RPAB3_HUMAN	3	-3	2	26.28
ECGYRIMYKK-M7-324	RPAB4_HUMAN	44	-2.3	1	26.26
LLDISELDVMGAGREAK-M9-323	NELFA_HUMAN	264	-3.5	1	26.17
SAGVPFHAKGRGLLR-R11-629	NELFA_HUMAN	202	-1.2	1	26.15
QMDIIVSEVSMIRDIR-I5-629	RPB2_HUMAN	597	-2.4	1	26.14
IKYAKEVLQK-E6-363	RPB2_HUMAN	341	-1.1	2	26.11
FAVALDSEQNNIHVK-A2-324	SPT5H_HUMAN	584	-3.2	2	26.02
DVTNFTVGGFAPMSPR-T6-324	SPT5H_HUMAN	658	-1.4	1	26.01
DVTNFTVGGFAPXSPR-V7-323	SPT5H_HUMAN	659	0.6	1	25.96
LLALGQGAWDXIDSQVFK-Q15-669	NELFB_HUMAN	358	0.9	2	25.96
HLQELVGQETLPR-L5-324	NELFB_HUMAN	325	-2.8	1	25.86
XIPRIDYDR-P3-323	SPT5H_HUMAN	309	-1.6	1	25.81
RTVDEMK-M6-324	NELFA_HUMAN	71	-1.1	1	25.8
GKNICEGGEEMDNK-E6-323	RPB1_HUMAN	155	-3.3	1	25.73
IKYAKEVLQK-E6-363	RPB2_HUMAN	341	-0.5	2	25.69
FHPKPSDLHLQTYGK-F1-630	RPB1_HUMAN	431	-1.3	1	25.69
FAVALDSEQNNIHVK-A2-630	SPT5H_HUMAN	584	-0.8	1	25.66
GASELVAELSTLYQCIR-L12-323	NELFD_HUMAN	414	-1.1	1	25.62
VRILPWSTFR-L4-323	RPB1_HUMAN	477	-1.9	1	25.58
KPKSATLR-T6-323	NELFA_HUMAN	173	-3.4	1	25.38
FLPMLMSFLVDDYTFNVDQK-L9-653	NELFB_HUMAN	381	-2.3	1	25.34
ARGPIQLNR-G3-324	RPB2_HUMAN	1065	-3.2	3	25.32
FAVALDSEQNNIHVK-V3-630	SPT5H_HUMAN	585	-2.4	1	25.24
RDCSTFLR-F6-323	RPB2_HUMAN	895	-1.5	1	25.24
LKIVGTRVDK-T6-323	RPB7_HUMAN	150	-2.5	1	25.23
QAXGVYITNFHVR-Q1-324	RPB2_HUMAN	731	0.7	1	25.21
QAXGVYITNFHVR-Q1-324	RPB2_HUMAN	731	0.8	2	25.14
IFVNGCWVGIHK-G9-324	RPB2_HUMAN	575	-2.2	1	25.14
GASELVAELSTLYQCIR-L9-323	NELFD_HUMAN	411	-3	1	25.11
FIIENTDLAVANSIR-D7-323	RPB3_HUMAN	27	0.6	1	25.1
NKTQISLVR-I5-669	RPB2_HUMAN	1146	-0.8	1	25.09
VRILPWSTFR-L4-323	RPB1_HUMAN	477	-1.9	1	25.08

**Table 7-27** | List of Protein-RNA crosslinks for the experiment with *Nr4a1*-(b) eRNA 1-100 (see description for Table 7-26).

Peptide-Residue-Nucleotide	Protein1	AbsPos1	Error_rel.ppm	nseen	ld.Score
FAVALDSEQNNIHVK-F1-324	SPT5H_HUMAN	583	-1.9	3	45.39
FAVALDSEQNNIHVK-A2-324	SPT5H_HUMAN	584	-0.9	2	39.39
ETFQVLNXYGK-Y9-324	SPT5H_HUMAN	564	-3.7	1	37.58
HAIYDKLDDDGLIAPGVR-A2-324	RPB2_HUMAN	843	-1.9	2	37.36
MHGGGPPSGDSACPLR-M1-630	RPB1_HUMAN	1	-3.6	2	37.08
ARGPIQLNR-P4-630	RPB2_HUMAN	1066	-1.4	2	36.83
DNDPNDYVEQDDILIVK-D6-347	RPB3_HUMAN	141	-2.8	1	36.32
VELDRK-R5-653	RPB1_HUMAN	1218	3.4	2	36.21
ARGPIQLNR-G3-630	RPB2_HUMAN	1065	-3	5	35.96
FHPKPSDLHLQTGYK-F1-630	RPB1_HUMAN	431	-3.6	2	35.63
VELDRK-R5-653	RPB1_HUMAN	1218	3.2	2	35.59
FHPKPSDLHLQTGYK-H2-630	RPB1_HUMAN	432	-1.5	2	35.52
FHPKPSDLHLQTGYK-H2-324	RPB1_HUMAN	432	-1.5	2	35.51
TLNYTAR-T5-323	RPB4_HUMAN	68	5.6	1	35.45
XHGGGPPSGDSACPLR-H2-630	RPB1_HUMAN	2	-1.7	2	35.42
ARGPIQLNR-G3-630	RPB2_HUMAN	1065	-2.1	3	35.41
VLIAQEKXATNTVYVFAK-T10-653	RPB2_HUMAN	202	3.7	1	35.2
FHPKPSDLHLQTGYK-F1-630	RPB1_HUMAN	431	-1.9	1	35.09
GFDQEEVFEKPTR-P11-363	RPB2_HUMAN	832	-1.7	1	35.01
HAIYDKLDDDGLIAPGVR-H1-630	RPB2_HUMAN	842	-4	1	34.57
GPIQLNR-I5-324	RPB2_HUMAN	1069	-1.1	1	34.46
DPIRCR-C5-323	RPAB4_HUMAN	36	0.8	1	34.24
DVTNFTVGGFAPXSPR-G8-324	SPT5H_HUMAN	660	0.3	1	34.18
SDSFPER-S3-363	NELFE_HUMAN	251	-0.8	1	34.13
GVAYKSR-Y4-653	SPT4H_HUMAN	108	0.8	1	34.02
GFDQEEVFEKPTR-P11-363	RPB2_HUMAN	832	-3	1	33.91
VELDRK-L3-653	RPB1_HUMAN	1216	4.3	2	33.86
TLLRK-L3-323	NELFA_HUMAN	248	-3.7	1	33.74
ARGPIQLNR-I5-630	RPB2_HUMAN	1067	-1.8	1	33.52
VYLLMKKLAF-K7-629	RPAB3_HUMAN	147	5.7	1	33.52
VYLLXKKLAF-K6-669	RPAB3_HUMAN	146	0.5	4	33.49
GVAYKSR-Y4-653	SPT4H_HUMAN	108	1.7	1	33.34
FHPKPSDLHLQTGYK-H2-630	RPB1_HUMAN	432	-3.7	4	33.26
GPIQLNR-P2-630	RPB2_HUMAN	1066	-2	1	33.25
SDSFPERR-P5-363	NELFE_HUMAN	253	-1.2	1	33.24
DVTNFTVGGFAPXSPR-V7-324	SPT5H_HUMAN	659	-2.6	1	33.16
FAVALDSEQNNIHVK-F1-324	SPT5H_HUMAN	583	-2.7	2	33.05
DILCR-C4-630	RPB1_HUMAN	1159	1	1	32.96

ATAISLXR-S5-347	SPT5H_HUMAN	195	5.4	1	32.84
GVAYKSR-Y4-653	SPT4H_HUMAN	108	1.2	2	32.79
GPIQLNR-I5-630	RPB2_HUMAN	1069	-1.6	2	32.74
XHGGGPPSGDSACPLR-H2-630	RPB1_HUMAN	2	-1.1	2	32.62
XHGGGPPSGDSACPLR-G4-630	RPB1_HUMAN	4	-2.5	1	32.61
VYLLXKKLAF-K6-669	RPAB3_HUMAN	146	-0.7	3	32.47
SDSFPER-F4-363	NELFE_HUMAN	252	-1	1	32.43
SDSFPER-S1-363	NELFE_HUMAN	249	-1.6	1	32.4
MHGGGPPSGDSACPLR-H2-630	RPB1_HUMAN	2	-2.7	2	32.31
IYAQKFIDRGK-G10-363	RPB2_HUMAN	426	5.4	2	32.28
ATAISLXR-L6-347	SPT5H_HUMAN	196	4.9	1	32.25
INNQLRR-Q4-629	RPB1_HUMAN	289	4.4	1	32.23
DVTNFTVGGFAPXSPR-G8-324	SPT5H_HUMAN	660	-1.6	1	32.1
HAIYDKLDDDGLIAPGVR-D9-324	RPB2_HUMAN	850	0.4	2	32.02
DILCR-C4-630	RPB1_HUMAN	1159	1.4	1	32.01
ARGPIQLNR-R2-630	RPB2_HUMAN	1064	-1.5	1	32
FAVALDSEQNNIHVK-A2-629	SPT5H_HUMAN	584	-2.6	1	31.96
VELDRK-R5-653	RPB1_HUMAN	1218	2.3	1	31.78
ARGPIQLNR-G3-630	RPB2_HUMAN	1065	-1.6	4	31.7
GEVVDVVVTQVVK-K13-630	RPB7_HUMAN	94	-1.1	1	31.66
GPIQLNR-P2-630	RPB2_HUMAN	1066	-1.5	1	31.64
ARGPIQLNR-G3-630	RPB2_HUMAN	1065	-1.4	2	31.63
DVTNFTVGGFAPXSPR-G9-324	SPT5H_HUMAN	661	-1.2	1	31.6
VELDRK-R5-653	RPB1_HUMAN	1218	3.9	1	31.55
GPIQLNR-G1-630	RPB2_HUMAN	1065	-1.8	1	31.48
VYLLXKKLAF-L8-669	RPAB3_HUMAN	148	0	1	31.42
XHGGGPPSGDSACPLR-G5-630	RPB1_HUMAN	5	-2.6	2	31.26
QPXEGRSRDGLR-S7-669	RPB2_HUMAN	1079	-0.9	1	31.24
IISDGLKYSLATGNWGDQK-I1-347	RPB2_HUMAN	439	4.3	2	31.22
MHGGGPPSGDSACPLR-H2-630	RPB1_HUMAN	2	-2.1	1	31.19
KQSSSSTTSQGGVK-Q2-323	NELFE_HUMAN	35	1.5	1	31.12
NLTYSAPLYVDITK-D11-363	RPB2_HUMAN	127	4.7	6	31
VERHMCDDGDIVFNRPQTLHK-P17-347	RPB1_HUMAN	462	5.9	2	30.85
ARGPIQLNR-I5-324	RPB2_HUMAN	1067	-1.2	1	30.83
IYAQKFIDRGK-G10-363	RPB2_HUMAN	426	4.3	1	30.82
ECGYRXYK-Y8-323	RPAB4_HUMAN	45	4.6	1	30.79
DXLEFPAQELRK-Q8-363	SPT5H_HUMAN	466	2.8	1	30.71
TLNYTAR-T5-323	RPB4_HUMAN	68	5.5	1	30.69
ALILGFMAAGSR-L4-347	NELFA_HUMAN	460	0.9	1	30.64
IYAQKFIDRGK-F6-363	RPB2_HUMAN	422	4.9	1	30.63
IAQEVQR-Q3-363	NELFD_HUMAN	263	2.8	1	30.62
LNSPIGR-P4-653	RPB2_HUMAN	488	3.6	1	30.6
IYAQKFIDRGK-F6-363	RPB2_HUMAN	422	4.6	1	30.58

VYLLXKKLAF-L8-669	RPAB3_HUMAN	148	-1.5	2	30.54
INNQLRR-Q4-629	RPB1_HUMAN	289	4.4	1	30.51
IYAQKFIDRGK-G10-363	RPB2_HUMAN	426	3.8	2	30.51
AKDILCR-L5-363	RPB1_HUMAN	1158	2.4	1	30.49
DILCRLEHTTLR-R12-324	RPB1_HUMAN	1167	0.7	1	30.48
QLVKSGAISAIAK-I8-323	NELFE_HUMAN	74	-1.9	1	30.45
ELQGFLDGVK-K10-324	NELFB_HUMAN	291	-3	2	30.38
NCAFVTYEK-K9-347	NELFE_HUMAN	304	-0.9	1	30.38
SFSLVK-L4-347	NELFB_HUMAN	103	3.6	1	30.3
ELQGFLDGVK-K10-324	NELFB_HUMAN	291	0.7	2	30.21
EVLQKEXLPHVGVSDFCETK-V11-630	RPB2_HUMAN	351	5.7	1	30.19
DXLEFPAQELRK-Q8-363	SPT5H_HUMAN	466	4	1	30.17
LRKGQELR-Q5-363	RPB3_HUMAN	157	-2.9	2	30.16
AHNNELEPTPGNTRLR-N12-347	RPB1_HUMAN	731	3.4	1	30.13
HAIYDKLDDDGLIAPGVR-I3-630	RPB2_HUMAN	844	-2.8	1	30.05
GPIQILNR-P2-630	RPB2_HUMAN	1066	-1.6	1	30.03
DVTNFTVGGFAPMSPR-P12-324	SPT5H_HUMAN	664	-1.8	1	30
VSRLHCESESEFK-V1-324	RPAB3_HUMAN	25	3.1	3	29.83
ARGPIQILNR-R2-630	RPB2_HUMAN	1064	-1.2	1	29.82
IISDGLKYSLATGNWGDQK-I1-363	RPB2_HUMAN	439	-0.6	2	29.81
FEQIYLSKPTHWER-S7-347	RPB2_HUMAN	94	2.9	1	29.8
GPIQILNR-I5-630	RPB2_HUMAN	1069	-1.3	3	29.78
EEELGEYMKKYAK-M9-630	SPT5H_HUMAN	142	0.9	1	29.77
DPEQLXNTRLR-L9-324	RPB2_HUMAN	587	1.9	1	29.77
XHGGGPPSGDSACPLR-G4-324	RPB1_HUMAN	4	-3.2	1	29.74
QDVIEVIEK-I4-323	RPB1_HUMAN	714	-2.2	1	29.72
ACQALGAMLSK-K11-323	NELFD_HUMAN	302	4.8	1	29.71
IISDGLKYSLATGNWGDQK-I1-363	RPB2_HUMAN	439	-2.8	1	29.69
GEVVDVAVTQVVK-V7-630	RPB7_HUMAN	88	-0.6	1	29.69
GEVVDVAVTQVVK-K13-630	RPB7_HUMAN	94	-1.3	1	29.64
LLDISELDXVGAGREAK-A16-363	NELFA_HUMAN	271	-0.3	1	29.64
IISDGLKYSLATGNWGDQK-I1-363	RPB2_HUMAN	439	-2.6	1	29.64
MHGGGPPSGDSACPLR-M1-629	RPB1_HUMAN	1	-1.5	2	29.62
FHPKPSDLHLQTGYK-H2-324	RPB1_HUMAN	432	-1.2	2	29.6
RLVVF DAR-L2-669	RPAB4_HUMAN	52	5.6	1	29.59
ILNLR-L4-653	SPT5H_HUMAN	1078	3.3	1	29.57
QLVKSGAISAIAK-I8-323	NELFE_HUMAN	74	-3	1	29.56
FHPKPSDLHLQTGYK-H2-630	RPB1_HUMAN	432	-3.4	1	29.55
GKNICEGGEEMDNK-I4-324	RPB1_HUMAN	153	-0.8	1	29.47
GPIQILNR-I5-630	RPB2_HUMAN	1069	-1.5	2	29.43
RLVVF DAR-L2-669	RPAB4_HUMAN	52	5.6	1	29.4
KFRFDYTNER-E9-363	RPB1_HUMAN	927	1.5	1	29.39
IYAQKFIDRGK-G10-363	RPB2_HUMAN	426	6	1	29.39

LNSPIGR-I5-653	RPB2_HUMAN	489	2.6	1	29.37
GEVVDVAVTQVVK-K13-630	RPB7_HUMAN	94	-2	1	29.32
DIVKVIDGPHSGR-V5-324	SPT5H_HUMAN	602	-3.9	1	29.3
RLIKSMESVMVK-K4-653	RPB1_HUMAN	866	2.7	1	29.25
ACQALGAMLSK-K11-323	NELFD_HUMAN	302	5.9	1	29.21
VYLLXKKLAF-L8-669	RPAB3_HUMAN	148	-1.6	3	29.17
EEELGEYMKKYAK-M9-630	SPT5H_HUMAN	142	3.4	1	29.14
SFSLVK-V5-363	NELFB_HUMAN	104	3.5	1	29.12
MATNTVYVFAK-F9-363	RPB2_HUMAN	208	-1.5	2	29.08
NLTYSAPLYVDITK-L8-363	RPB2_HUMAN	124	-0.2	5	29.06
IAQEVQRFAQEK-E4-324	NELFD_HUMAN	264	0.8	2	29.06
GEVVDVAVTQVVK-K13-630	RPB7_HUMAN	94	-0.7	1	29
ELQGFLDGVK-G8-669	NELFB_HUMAN	289	-3.7	1	29
FAVALDSEQNNIHVK-A2-629	SPT5H_HUMAN	584	-2.6	1	28.98
RQXDIIVSEVSMIRDIR-R1-630	RPB2_HUMAN	592	4.3	1	28.97
SLSEYNNFKSXVSGAK-N7-363	RPB1_HUMAN	765	-1.8	2	28.95
IAQEVQRFAQEK-E4-324	NELFD_HUMAN	264	-1.1	1	28.9
INNQLRR-R6-629	RPB1_HUMAN	291	5	1	28.9
RLIKSMESVMVK-M6-653	RPB1_HUMAN	868	2.5	1	28.89
MATNTVYVFAK-A10-363	RPB2_HUMAN	209	-1	1	28.89
ITTPYXTKYER-T2-363	RPAB2_HUMAN	53	3.3	1	28.87
ECGYRXYK-Y8-323	RPAB4_HUMAN	45	3.9	1	28.85
LTHVYDLCKGKNICEGGEEXDNK-V4-324	RPB1_HUMAN	144	-3.9	1	28.85
VELDRK-L3-653	RPB1_HUMAN	1216	3.8	2	28.84
IAQEVQRFAQEK-E4-323	NELFD_HUMAN	264	-2.6	2	28.71
INNQLRR-Q4-629	RPB1_HUMAN	289	5	1	28.66
SCLENSRPTSTI WVSMILAR-M17-653	RPB2_HUMAN	239	0.3	1	28.61
RHIDQLKER-Q5-629	RPB2_HUMAN	642	1.7	1	28.56
ARGPIQILNR-I5-324	RPB2_HUMAN	1067	-2.3	1	28.55
KQSSSSTTSQGGVK-Q2-323	NELFE_HUMAN	35	2.4	1	28.54
MATNTVYVFAK-F9-363	RPB2_HUMAN	208	-2.2	1	28.48
SLSEYNNFKSXVSGAK-L2-363	RPB1_HUMAN	760	-1.1	5	28.43
NLTYSAPLYVDITK-L8-363	RPB2_HUMAN	124	-3.9	5	28.39
LLALGQGAWDXIDSQVFK-S14-669	NELFB_HUMAN	357	1.5	1	28.38
DILCRLEHTTLR-L3-323	RPB1_HUMAN	1158	3.3	1	28.37
FSTRDYIXEPSIFNTLK-N14-324	NELFD_HUMAN	64	3	1	28.35
SLSEQPVM DATATEQAK-T12-323	NELFE_HUMAN	60	1.1	1	28.34
KAYFLGYXVHR-V9-669	RPB2_HUMAN	369	-1.2	1	28.27
SLSEYNNFKSXVSGAK-L2-363	RPB1_HUMAN	760	-2.1	3	28.15
SLKQR-Q4-323	RPB1_HUMAN	341	3.7	1	28.13
VYLLMKKLAF-M5-347	RPAB3_HUMAN	145	4.8	1	28.13
LLFQELMSMSIAPR-E5-629	RPB2_HUMAN	1161	-0.9	3	28.07
VYLLXKKLAF-F10-669	RPAB3_HUMAN	150	-1	1	28.07

DGGLRFGEXER-F6-347	RPB2_HUMAN	1086	-2.4	1	28.04
SVWGSLAVQNSPK-P12-323	NELFE_HUMAN	354	5.4	1	28.04
IPIMLR-L5-669	RPB2_HUMAN	156	-1.4	4	28.01
RLIKSMESVMVK-M6-653	RPB1_HUMAN	868	0	1	27.96
KLEDLLEKSFSLVK-E7-629	NELFB_HUMAN	98	5.4	1	27.9
LEFHQSVFDELR-V7-323	NELFB_HUMAN	65	5.8	1	27.88
DPEQLXNTRLR-R10-324	RPB2_HUMAN	588	-2.4	1	27.87
SCLENSSRPTSTI WV SMLAR-M17-653	RPB2_HUMAN	239	-0.2	1	27.86
LVENGGXFVCKTR-K11-363	SPT5H_HUMAN	639	3.9	1	27.86
ESDTGLWLHNK-G5-324	NELFA_HUMAN	10	1.6	1	27.84
ERGVKLLDISELDMVGAGR-L7-630	NELFA_HUMAN	257	2.1	1	27.82
VSRLHCESEFK-V1-324	RPAB3_HUMAN	25	1.8	2	27.79
NCAFVTYEK-K9-347	NELFE_HUMAN	304	0	1	27.79
EEVTELLAR-V3-347	RPAB1_HUMAN	156	-2.8	2	27.73
QPTLHKMSMMGHR-H12-347	RPB1_HUMAN	472	3.2	2	27.72
DVTNFTVGGFAPXSPR-G8-324	SPT5H_HUMAN	660	0.7	1	27.72
AELLQKSTETAQQLK-Q5-629	NELFA_HUMAN	180	-0.9	2	27.67
NLTYSAPLYVDITK-D11-363	RPB2_HUMAN	127	3.2	5	27.67
MHGGGPPSGDSACPLR-H2-324	RPB1_HUMAN	2	-1.2	1	27.66
LLALGQGAWDXIDSQVFK-S14-669	NELFB_HUMAN	357	2.8	2	27.66
LKHMVDDK-D6-363	RPB2_HUMAN	1056	0.5	1	27.65
LTHVYDLCKGK-V4-324	RPB1_HUMAN	144	-1.5	1	27.65
ALILGFMAGSR-L4-347	NELFA_HUMAN	460	0.7	1	27.64
QSSSTTSQGGVKR-K13-324	NELFE_HUMAN	47	2.2	1	27.62
VYLLMKKLAF-M5-347	RPAB3_HUMAN	145	5.5	1	27.62
QPTLHKMSMMGHR-H12-347	RPB1_HUMAN	472	4.1	2	27.59
ELYRACAVEVK-A7-363	NELFB_HUMAN	142	2.2	2	27.57
LAEAHPDCLXNFTVK-H5-347	NELFD_HUMAN	166	5.3	1	27.54
ETCQGXRHAIYDKLDDDGLIAPGVR-D12-630	RPB2_HUMAN	846	-1.8	1	27.47
KNLSLTR-L3-629	NELFA_HUMAN	431	-2.9	2	27.46
VELDRK-L3-653	RPB1_HUMAN	1216	3	1	27.44
VELDRK-D4-653	RPB1_HUMAN	1217	3.1	1	27.38
IISDGLKYSLATGNWGDQK-I1-363	RPB2_HUMAN	439	-2.7	2	27.3
DPEQLXNTRLR-R10-324	RPB2_HUMAN	588	0.9	1	27.3
NLTYSAPLYVDITK-Y9-363	RPB2_HUMAN	125	5.2	1	27.29
GVAYKSR-K5-347	SPT4H_HUMAN	109	-2.2	2	27.28
FSTRDYIXEPSIFNTLK-N14-323	NELFD_HUMAN	64	2.7	1	27.25
LGYWNQQMVPIK-Q6-363	SPT5H_HUMAN	252	4.6	1	27.25
RIAQEVQR-E5-363	NELFD_HUMAN	264	3.7	1	27.22
VSRLHCESEFK-V1-324	RPAB3_HUMAN	25	4.2	1	27.21
GYIYVEAYKQTHVK-Y4-324	SPT5H_HUMAN	226	3.9	1	27.2
IPQIGDKFASRHGQK-I1-669	RPB2_HUMAN	928	0.7	1	27.2
EEVTELLAR-V3-347	RPAB1_HUMAN	156	-3.1	1	27.17

SFSLVK-V5-347	NELFB_HUMAN	104	3.2	1	27.15
XPYACKLLFQELMSMSIAPRMMSV-F9-324	RPB2_HUMAN	1159	-2.1	1	27.13
NLTYSAPLYVDITK-L8-363	RPB2_HUMAN	124	4.4	2	27.13
NCAFVTYEK-K9-347	NELFE_HUMAN	304	0.6	1	27.11
QDVIEVIEK-I7-323	RPB1_HUMAN	717	-0.7	1	27.11
LEQLDHR-H6-630	NELFB_HUMAN	553	5.6	2	27.1
VERHMCDDGDIVFNQPTLHK-P17-347	RPB1_HUMAN	462	0.7	2	27.07
STYCLLNGLTDR-C4-324	RPB2_HUMAN	161	-0.3	2	27.05
MATNTVYVFAKK-K11-323	RPB2_HUMAN	210	-1.1	2	27
IPFGFKHRTLPHFIK-G4-363	RPB1_HUMAN	801	2.5	1	26.99
QPTLHKMSMMGHR-M9-347	RPB1_HUMAN	469	2.1	1	26.99
GYIYVEAYKQTHVK-Y4-323	SPT5H_HUMAN	226	-3.1	1	26.98
DMLEFPAQELRKYFK-E4-324	SPT5H_HUMAN	462	-3.6	1	26.98
DKLLERSVAIASEGK-R6-347	NELFB_HUMAN	76	-3.3	2	26.98
TLEGKLDPEK-D8-629	NELFE_HUMAN	98	2.5	1	26.96
RSRTLEGK-E6-323	NELFE_HUMAN	93	-2.9	2	26.95
NLTYSAPLYVDITK-L8-363	RPB2_HUMAN	124	4.1	3	26.93
NLSLTR-S3-347	NELFA_HUMAN	432	1.5	2	26.9
FEDEELQQILDDIQTQ-Q7-324	RPB4_HUMAN	128	-2.8	1	26.8
ISDEECFVLGXEPRL9-324	RPB1_HUMAN	236	-2.2	2	26.79
EEVTELLAR-A8-363	RPAB1_HUMAN	161	2.7	1	26.77
TPHYGSQTPLHDGSR-H3-324	SPT5H_HUMAN	801	-2.4	1	26.73
DMLEFPAQELR-E9-629	SPT5H_HUMAN	467	3.3	1	26.7
DKRTQIVYSDDVYK-Y13-324	NELFE_HUMAN	372	-2.6	1	26.7
KYFKMGDHSV-K1-324	SPT5H_HUMAN	470	2	1	26.67
XPYACKLLFQELMSMSIAPRMMSV-A4-324	RPB2_HUMAN	1154	-1.3	1	26.67
ECGYRIXK-I6-323	RPAB4_HUMAN	43	3.8	1	26.65
XHGGGPPSGDSACPLR-G5-630	RPB1_HUMAN	5	-0.9	1	26.64
INNQLRR-R6-347	RPB1_HUMAN	291	-1.5	1	26.63
GGFGSPGGGSGGMSRGRGR-G12-324	SPT5H_HUMAN	693	-3.1	1	26.62
YTPTSPSYSPSSPEYTPSPK-P17-363	RPB1_HUMAN	1855	5.7	1	26.6
SLSEYNNFKSXVVGAK-S3-323	RPB1_HUMAN	761	-1.7	3	26.59
EEEEPEEEEEEEEEYDEEEEEEDDRPPKPR-K30-669	SPT5H_HUMAN	68	-3.2	1	26.59
EGLIDTAVKTAETGYIQRR-R18-323	RPB1_HUMAN	862	0.7	1	26.59
IQAGDPVAR-V7-363	RPAB1_HUMAN	179	-2.7	1	26.58
ITIMPKHEDLK-H7-347	SPT5H_HUMAN	454	-0.9	1	26.57
RLTHVYDLCK-Y6-347	RPB1_HUMAN	145	3	1	26.57
KAYFLGYXVHR-K1-324	RPB2_HUMAN	361	4.9	1	26.57
SAAGSEKEEPEDEEEEEEEEEYDEEEEEEDDRPPKPR-E6-324	SPT5H_HUMAN	37	5.9	1	26.56
EEELGEYYMKKYAK-Y8-653	SPT5H_HUMAN	141	0.1	1	26.56
LVLIR-R4-653	RPB1_HUMAN	1258	1.7	3	26.56
IISDGLKYSLATGNWGDQK-I1-347	RPB2_HUMAN	439	5	2	26.55

CFTCGKIVGNK-I7-347	RPAB5_HUMAN	13	1.4	1	26.54
ETCQGXRHAIYDK-C3-363	RPB2_HUMAN	837	5.9	1	26.54
ACQALGAMLSK-K11-629	NELFD_HUMAN	302	3.2	1	26.54
FSTRDYIXEPSIFNTLK-T15-323	NELFD_HUMAN	65	2.6	1	26.53
KNLSLTR-L3-629	NELFA_HUMAN	431	-2.5	1	26.5
ETCQGXRHAIYDKLDDGLIAPGVR-A9-630	RPB2_HUMAN	843	-0.2	1	26.45
SAAGSEKEEPEDEEEEEEEYDEEEEEEDD	SPT5H_HUMAN	38			
DRPPKKPR-K7-324			2.6	1	26.44
DKRTQIVYSDDVYK-Y8-324	NELFE_HUMAN	367	-1.1	1	26.44
TLEGKLDPEK-L6-629	NELFE_HUMAN	96	-1.2	2	26.43
LLALQGAWDXIDSQVFK-Q15-669	NELFB_HUMAN	358	2.9	1	26.43
KYFKXGDHVK-Y2-324	SPT5H_HUMAN	471	2.4	8	26.38
GPIQILNR-I3-630	RPB2_HUMAN	1067	-3.5	1	26.38
XHGGGPPSGDSACPLR-H2-630	RPB1_HUMAN	2	-0.5	3	26.36
LGYNWQQMVPIK-Q6-363	SPT5H_HUMAN	252	3.6	1	26.35
MHGGGPPSGDSACPLR-G4-324	RPB1_HUMAN	4	-2.7	1	26.33
SLSEYNNFKSXVVSQAK-S3-324	RPB1_HUMAN	761	4.2	4	26.32
DSKYAYTGECE-E9-324	RPB2_HUMAN	220	4.4	1	26.3
IPFGFKHR-G4-347	RPB1_HUMAN	801	5.7	1	26.28
DQREEELGEYXKK-R3-323	SPT5H_HUMAN	133	6	1	26.26
EMLPHVGVSDFCETKK-P4-324	RPB2_HUMAN	349	2.3	2	26.25
DKRTQIVYSDDVYK-K14-324	NELFE_HUMAN	373	-3.5	1	26.24
YIIRDNGDRIDLR-R13-347	RPB1_HUMAN	430	3.9	2	26.24
DIVKVIDGPHSGR-P9-324	SPT5H_HUMAN	606	-3.5	1	26.19
GEVVDVAVTQVVK-V4-630	RPB7_HUMAN	85	-0.8	1	26.18
NDIFAIGSLXDDYLGLVS-S8-669	RPB7_HUMAN	162	-0.4	1	26.16
DTYLDTQVVGQTVIR-Q11-653	SPT5H_HUMAN	1000	5.2	2	26.16
ETFQVLNXYGKVVTVR-Y9-669	SPT5H_HUMAN	564	1.9	1	26.15
XHGGGPPSGDSACPLR-H2-324	RPB1_HUMAN	2	-0.5	1	26.15
EGLIDTAVKTAETGYIQR-R18-323	RPB1_HUMAN	862	-1.7	1	26.14
NLTYSAPLYVDITK-Y9-363	RPB2_HUMAN	125	3.6	3	26.11
QRNKNALLR-Q1-363	NELFB_HUMAN	435	4.3	1	26.1
ECGYRIMYKK-K10-324	RPAB4_HUMAN	47	-2	1	26.08
LLFQELMSMSIAPR-E5-629	RPB2_HUMAN	1161	-0.7	4	26.05
RSLSEQPVMDDTATATEQAK-A12-630	NELFE_HUMAN	59	2.1	2	26.03
VQFGVLSPEL-L11-324	RPB1_HUMAN	31	-2.6	1	26.02
AELLQKSTETAQQLK-T10-630	NELFA_HUMAN	185	5	1	26.01
SVRIPQIGDKFASR-Q6-347	RPB2_HUMAN	930	5.3	1	26
VERHMCDDGDIVFNQPTLHK-K21-347	RPB1_HUMAN	466	3.6	1	25.97
MEPDGTYEPGFVGIRFCQECNNXLYPK-R15-347	RPB9_HUMAN	15			
HLALLCDTXXTCR-H1-669	RPB1_HUMAN	1397	-3.9	1	25.96
VYMHLPTQDNK-T8-669	RPB1_HUMAN	1314	3.4	1	25.96
ECGYRIMYKK-K10-324	RPB1_HUMAN	47	3.7	1	25.94
ECGYRIMYKK-K10-324	RPAB4_HUMAN	47	-1.6	1	25.92

XAAGGSDPR-P8-363	RPB4_HUMAN	8	0.1	1	25.9
QPTLHKMSMMGHR-H12-347	RPB1_HUMAN	472	3	3	25.9
EEVTELLAR-V3-347	RPAB1_HUMAN	156	-3	2	25.87
EEVTELLAR-L6-363	RPAB1_HUMAN	159	4.3	4	25.87
SVWGSLAVQNSPK-N10-324	NELFE_HUMAN	352	0.3	1	25.87
EEVTELLAR-V3-363	RPAB1_HUMAN	156	3.7	2	25.87
XIIPVRCFTCGKIVGNK-I2-669	RPAB5_HUMAN	2	2.7	2	25.84
LVLIRIMNSDENK-I5-653	RPB1_HUMAN	1259	1.5	2	25.82
MEPDGTYEPGFVGIR-P9-363	RPB9_HUMAN	9	2.3	1	25.81
VVVENGELIXGILCK-L8-324	RPB1_HUMAN	635	-1.4	1	25.8
RQGEVVQRLTR-Q7-630	NELFB_HUMAN	202	-3.7	3	25.79
LGYNWQQMVPIK-N5-363	SPT5H_HUMAN	251	5	1	25.78
TEGEHDPVTEFIAHCKSNFIMVN-I20-324	NELFD_HUMAN	587	-2.9	1	25.75
IAQEVQR-E4-363	NELFD_HUMAN	264	1.6	1	25.74
FEDEELQQILDDIQTQR-E5-323	RPB4_HUMAN	126	3.8	1	25.71
ITSQIFIGPTYQRLK-P9-363	RPB2_HUMAN	1045	-2.1	1	25.71
TLEGKLDPEK-L6-629	NELFE_HUMAN	96	0	1	25.7
DIVKVIDGPHSGR-V5-324	SPT5H_HUMAN	602	-3.7	1	25.7
DFNLELAIKTR-E5-347	RPB2_HUMAN	432	3.1	2	25.69
KQSSSSTTSQGGVKR-Q10-363	NELFE_HUMAN	43	-2.8	1	25.65
AHNNELEPTPGNTR-N12-347	RPB1_HUMAN	731	1.5	1	25.64
TLEGKLDPEK-L6-629	NELFE_HUMAN	96	1.9	1	25.63
AXQKSGRPLKSLK-G6-363	RPB1_HUMAN	333	1	1	25.62
QPXEGRSRDGLR-R8-669	RPB2_HUMAN	1080	-1.7	2	25.62
MPYACKLLFQELMSXSIAPRMMSV-K6-324	RPB2_HUMAN	1156	-0.9	1	25.61
LLALGQGAWDXIDSQVFK-V16-669	NELFB_HUMAN	359	0.2	1	25.6
ISDEECFVLGXEPV-V8-324	RPB1_HUMAN	235	-2.4	1	25.58
XPYACKLLFQELMSXSIAPRMMSV-A4-324	RPB2_HUMAN	1154	-2.9	1	25.57
ETCQGXRHAIYDKLDDGLIAPGVR-D16-630	RPB2_HUMAN	850	-2.4	1	25.55
TKRLVFDAR-V6-653	RPAB4_HUMAN	54	3	1	25.55
SLSEYNNFKSXVVGAK-S3-324	RPB1_HUMAN	761	3.8	2	25.54
AXQKSGRPLKSLK-P8-363	RPB1_HUMAN	335	0.7	2	25.53
RSRTLEGK-E6-323	NELFE_HUMAN	93	-3.3	2	25.52
FAVALDSEQNNIHVK-V14-323	SPT5H_HUMAN	596	-3.4	1	25.52
IISDGLKYSLATGNWGDQK-D4-363	RPB2_HUMAN	442	-2.2	1	25.47
FEQIYLSKPTHWER-S7-347	RPB2_HUMAN	94	2.4	1	25.46
ELQGFLDGVK-G8-653	NELFB_HUMAN	289	5.1	1	25.46
AELLQKSTETAQQLK-Q5-323	NELFA_HUMAN	180	-3.2	1	25.46
KKIIITEDGEFK-T6-629	RPB1_HUMAN	1323	-1	1	25.45
TLEGKLDPEK-K7-629	NELFE_HUMAN	97	0.3	1	25.44
KAKQDVIEVIEK-K3-347	RPB1_HUMAN	710	-2	1	25.44
EEELGEYMKKYAK-L4-630	SPT5H_HUMAN	137	2.4	3	25.4
GFDQEEVFEKPTR-P11-347	RPB2_HUMAN	832	0.9	3	25.4

GAFSPFGNIIDLSXDPPR-D11-324	NELFE_HUMAN	288	5	1	25.4
EMLPHVGVSDFCETKK-P4-347	RPB2_HUMAN	349	1.6	3	25.39
KSLYESFVSSSDR-E5-630	NELFE_HUMAN	134	4.9	2	25.39
DMLEFPAQELRKYFK-E9-324	SPT5H_HUMAN	467	-3.3	1	25.38
AXQKSGRPLKSLK-G6-363	RPB1_HUMAN	333	1.6	3	25.37
SLSEYNNFKSXVVGAK-N7-363	RPB1_HUMAN	765	-2.5	2	25.37
SLKQRLK-R5-630	RPB1_HUMAN	342	4.7	1	25.37
VPEKKLK-K4-363	NELFB_HUMAN	126	-3.1	2	25.37
GAFSPFGNIIDLSXDPPR-L12-324	NELFE_HUMAN	289	5	1	25.34
DFNLELAIKTR-A7-347	RPB2_HUMAN	434	3.4	3	25.3
ESDTGLWLHMK-E1-324	NELFA_HUMAN	6	4	1	25.3
VYLLMKKLAF-M5-347	RPAB3_HUMAN	145	3.4	1	25.29
SLSEYNNFKSXVVGAK-S3-324	RPB1_HUMAN	761	5.5	3	25.28
DXLEFPAQELRK-Q8-363	SPT5H_HUMAN	466	2.7	1	25.27
FEQIYLSKPTHWER-Q3-324	RPB2_HUMAN	90	2.8	2	25.25
QGEVVQRLTRMVGK-G13-630	NELFB_HUMAN	209	4.9	1	25.25
EEELGEYMKKYAK-K10-653	SPT5H_HUMAN	143	-0.6	1	25.24
GHLMAITR-I6-669	RPB1_HUMAN	1414	5.8	1	25.24
GALNPADITVLFK-T9-653	NELFD_HUMAN	311	-3.6	1	25.22
AQSGDKPSEGRPR-E9-363	RPAB1_HUMAN	50	-3.2	1	25.19
LAEHPDCLXLNFTVK-F13-324	NELFD_HUMAN	174	-2.1	3	25.18
ISDEECFVLGXEPV-V8-324	RPB1_HUMAN	235	-2.2	1	25.14
SCLNSSRPTSTIWWMLAR-M17-653	RPB2_HUMAN	239	-2.9	1	25.13
DKTGSSAQKSLSEYNNFK-S10-323	RPB1_HUMAN	759	4.8	1	25.13
XHGGGPPSGDSACPLR-L15-363	RPB1_HUMAN	15	-0.1	1	25.12
GAKPGVTKEKR-R11-324	RPB2_HUMAN	335	1.4	1	25.1
MSDSEDSNFSEEDSERSSDGEEAEVDEER-V26-347	SPT5H_HUMAN	26	-0.4	1	25.09
IYAQKFIDR-Y2-324	RPB2_HUMAN	418	-0.8	1	25.09
RQMDIIVSEVSMIR-Q2-630	RPB2_HUMAN	593	5	1	25.09
TGSSAQKSLSEYNNFK-A5-324	RPB1_HUMAN	756	-3.5	1	25.08
HLALLCDTXXTCR-H1-669	RPB1_HUMAN	1397	3.7	1	25.06
LNSPIGRDGK-S3-363	RPB2_HUMAN	487	-1.1	1	25.05
ISDEECFVLGXEPV-E12-324	RPB1_HUMAN	239	-1.9	2	25.05
QGVIERTRG-G8-630	RPB1_HUMAN	69	-2.2	1	25.04
GAFSPFGNIIDLSMDPPR-D15-629	NELFE_HUMAN	292	-1.9	1	25.03
KYKPMTNVS-K1-324	NELFA_HUMAN	520	1.6	1	25.03
TLPHFIKDDYGPEER-F5-347	RPB1_HUMAN	810	-3.6	1	25.02
VERHMCDDIVIFNRQPTLHK-P17-347	RPB1_HUMAN	462	5.3	2	25.02
IQAGDPVARYFGIKR-I13-669	RPAB1_HUMAN	185	3.6	1	25.01

# Acknowledgements

First of all, I would like to thank Dr. Claus-D. Kuhn, who gave me the opportunity to start this work as a half-baked master student and to progress with this exciting topic during my PhD studies. I highly appreciate his support, tireless encouragement, and belief in me and my work. And looking back at the last years, I learned to value his constant pushing and to reflect on things more from different perspectives.

Next, I would like to thank my dear master student Filiz Kuybu, who worked adamantly on our topic and pushed this project forward by establishing and performing highly relevant EMSA experiments. Filiz is an excellent dialog partner, avid tea drinker and provided me regularly with food and drinks when I forgot to do so myself, thereby helping me to survive and proceed.

Further, I would like to thank Andreas Pittroff for helping me analyze our sequencing data, his bioinformatics support in general, and finally, for being such a pleasant and calm character (providing a balance to the hectic rest of the group). I would also like to acknowledge Dr. Seung-Kyoon Kim, Dr. Katie Schaukowitch, and Dr. Taekyung Kim for providing me the basis for the start of this work and the starting material for my experiments. I'm very grateful to Dr. Alexander Leitner for his expertise on and performance of the RNA-protein crosslinking and mass spec experiments. Moreover, I want to acknowledge Elena Katzowitsch and especially Dr. Kristina Döring for making our sequencing runs smooth and for fantastic communication.

A big thank goes to my friend and colleague Dr. Felix Klatt, who gave me support, made me laugh on long working days, and just for being on the same wavelength (in most things but not all, which made our discussions even more valuable). I would also like to thank all our former and recent group members, irrespective of whether we got along with each other well or not. Thanks for being as you are, so unbelievably diverse and individual characters. All the experiences and relationships, good or bad, are shaping who we are and let us grow.

Eventually, special thanks go to Thomas Forster for his endless patience, love, and being there. Last but not least, I would like to thank my parents, especially my mom, for her unconditional love, her story of unbelievable rehabilitation after her accident, and her will to live, which made me believe in miracles and taught me the lesson of "mind over matter."

# (Eidesstattliche) Versicherungen und Erklärungen

(§ 9 Satz 2 Nr. 3 PromO BayNAT)

*Hiermit versichere ich eidesstattlich, dass ich die Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe (vgl. Art. 64 Abs. 1 Satz 6 BayHSchG).*

(§ 9 Satz 2 Nr. 3 PromO BayNAT)

*Hiermit erkläre ich, dass ich die Dissertation nicht bereits zur Erlangung eines akademischen Grades eingereicht habe und dass ich nicht bereits diese oder eine gleichartige Doktorprüfung endgültig nicht bestanden habe.*

(§ 9 Satz 2 Nr. 4 PromO BayNAT)

*Hiermit erkläre ich, dass ich Hilfe von gewerblichen Promotionsberatern bzw. –vermittlern oder ähnlichen Dienstleistern weder bisher in Anspruch genommen habe noch künftig in Anspruch nehmen werde.*

(§ 9 Satz 2 Nr. 7 PromO BayNAT)

*Hiermit erkläre ich mein Einverständnis, dass die elektronische Fassung der Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Überprüfung unterzogen werden kann.*

(§ 9 Satz 2 Nr. 8 PromO BayNAT)

*Hiermit erkläre ich mein Einverständnis, dass bei Verdacht wissenschaftlichen Fehlverhaltens Ermittlungen durch universitätsinterne Organe der wissenschaftlichen Selbstkontrolle stattfinden können.*

.....  
Ort, Datum, Unterschrift