UNIVERSITY OF CRETE

FORTH

INSTITUTE OF MOLECULAR BIOLOGY & BIOTECHNOLOGY

The role of the nascent RNA in chromatin conformation and genome regulation

Ο ρόλος του νεοσυντιθέμενου RNA στη ρύθμιση της διαμόρφωσης της χρωματίνης και της γονιδιακής έκφρασης

Konstantinos Kydonakis

Heraklion 2021-2023

Abstract

Nascent RNA has for long time been considered as a passive molecule of transcription and co-transcriptional processes. However, recently published studies demonstrate and establish its role in transcription progression, splicing and chromatin architecture. In line with these studies, we tried to standardize and optimize two molecular methods to address the issue of RNA organization. Our aim was to elaborate on its implication in various nuclear function, such as 3D chromatin conformation maintenance and gene expression regulation. We apply Global RNA Interaction with DNA by deep sequencing (GRID-seq) is an "*all to all"* genomic approach to study RNA-chromatin interaction and CRISPR-dCas9 system-based chromatin precipitation (CRISPR pull down) as a molecular approach to targeting specific chromatin regions and co-precipitate every factor that interacts with it. Regarding GRID-seq, we managed to perform double fixation in our cells, but further optimization is necessary, particularly in improving the efficiency of nuclei isolation. In parallel, for CRISPR pull down, we worked with the *in vitro* approach, and attempted to isolate the target chromatin and evaluate if our assay was specific calculating fold enrichment ratio between expected vs non-specific loci via qPCR. Although we only reached low efficiency in our first attempts, we propose several optimizations that will improve our approach.

Περίληψη

Τα νεοσυντιθέμενα μόρια RNA θεωρούνταν για πολύ καιρό ως παθητικοί παίχτες της μεταγραφής και των συν-μεταγραφικών λειτουργιών. Ωστόσο πρόσφατα δημοσιευμένες μελέτες έχουν καταδείξει τον ρόλο των μορίων αυτών στην προαγωγή της μεταγραφής, του ματίσματος και στην αρχιτεκτονικής της χρωματίνης. Βασιζόμενοι σε αυτές τις μελέτες, προσπαθήσαμε να βελτιώσουμε και εντάξουμε στο πλαίσιο μελέτης της οργάνωσης και της χωρικής διευθέτησης του νεοσυντιθέμενου RNA, δύο νέες μοριακές μεθόδους. Μια εξ αυτών μελετάει ταυτόχρονα και καθολικά τις αλληλεπιδράσεις των μορίων RNA με την χρωματίνη και περιλαμβάνει σε τεχνολογίες αλληλούχισης νέας γενιάς (GRID-seq). Η επόμενη αξιοποιεί το σύστημα CRISPR-dCas9 και επιτυγχάνει την στόχευση συγκεκριμένων περιοχών χρωματίνης, επιτρέποντας την συν-κατακρήμνιση κάθε είδους αλληλεπιδρώντων παραγόντων. Όσον αφορά το GRID-seq καταφέραμε να πραγματοποιήσουμε διπλή χημική μονιμοποίηση, ωστόσο είναι απαραίτητο να γίνουν βελτιώσεις στη διαδικασία με σκοπό να αυξήσουμε τον αριθμό των πυρήνων που απομονώνουμε. Παράλληλα πραγματοποιήσαμε κατακρήμνιση χρωματίνης αξιοποιώντας το σύστημα CRISPR-dCas9, την *in vitro* προσέγγιση και αξιολογήσαμε τον εμπλουτισμό αξιοποιώντας την μέθοδο της ποσοτικής PCR. Παρόλο που στις περισσότερες περιπτώσεις φαίνεται το σύστημά μας να μπορεί να διακρίνει με υψηλή εξειδίκευση τους επιθυμητούς στόχους, περισσότερες βελτιώσεις είναι απαραίτητες ώστε να αυξήσουμε την απόδοση του εμπλουτισμού.

Ευχαριστίες

Καθώς η παρούσα εργασία πραγματοποιήθηκε στο εργαστήριο Μηχανισμών Ρύθμισης της γονιδιακής έκφρασης του Ινστιτούτου Μοριακής Βιολογίας και Βιοτεχνολογίας, αισθάνομαι την ανάγκη να ευχαριστήσω τον Δόκτωρ. Matthieu Lavigne που με δέχτηκε στο εργαστήριό του και μου έδωσε την ευκαιρία να αποκτήσω όλη αυτή την ανεκτίμητή εργαστηριακή εμπειρία. Εκτός αυτού, του οφείλω και ένα τεράστιο ευχαριστώ για την καλή διάθεση και προθυμία να προσφέρει την βοήθειά του τόσο με γνώσεις όσο και πράξεις κατά την διεξαγωγή των πειραμάτων, αλλά και που ασχολήθηκε να μου παρέχει κάθε δυνατή στήριξη για να μπορέσω να ανταπεξέλθω στις απαιτήσεις αυτού του προγράμματος. Με πλήρη ειλικρίνεια θέλω να καταγραφεί ότι με έναν τέτοιο ακαδημαϊκό υπεύθυνο δε φοβάμαι καμία πρόκληση!

Πλην των προηγούμενων, θέλω ακόμα να τον ευχαριστήσω που μου έδωσε την ευκαιρία να γνωρίσω και να συνεργαστώ με τον Βάιο Θεοδοσίου, έναν εξαιρετικό άνθρωπο που μου στάθηκε σε όλες της φάσεις και τις δυσκολίες. Ο Βάιος ήταν πάντα εκεί να με βοηθήσει τόσο με λύσεις όσο και με πράξεις, χωρίς ποτέ να παραπονεθεί ή να με αγνοήσει. Τον ευχαριστώ πολύ λοιπόν και ελπίζω να μπορέσω να του ανταποδώσω κάποια στιγμή.

Δε θα ήθελα όμως να παραλείψω και τα υπόλοιπα μέλη του εργαστηρίου: τη Μυρτώ Μίτλεττον που είναι συμφοιτήτριά μου και ήταν πάντα εκεί να με βοηθήσει με όποιον τρόπο μπορούσε, την Μαρία-Ηλέκτρα Κοντονίκου η οποία πραγματοποιεί την πτυχιακής της εργασία και έχει δουλέψει πάνω στο project, τον Νίκο Βουζουνεράκη, τον Ιωάννη Πετροσιάν, τον Νεκτάριο Βελμέζο, την Μαριάννα Σταγάκη και τον Στέργιο Μανάκα για την άψογη συνεργασία και συνύπαρξη μέσα στο εργαστήριο.

Ένα μεγάλο ευχαριστώ οφείλω στα μέλη της Τριμελούς επιτροπής μου, τους κυρίους Γιακουντή Αντώνιου και Σπηλιανάκη Χαράλαμπου, που δέχτηκαν να εποπτεύσουν την παρούσα εργασία.

Δεν μπορώ να παραλείψω την ανεκτίμητη βοήθεια που έλαβα από τα μέλη του εργαστηρίου Επιγενετικής του κυρίου Ταλιανίδη τόσο σε υλικά όσο και σε συμβουλές και καθοδήγηση.

Πέραν όμως των ανθρώπων από τον ακαδημαϊκό χώρο, οφείλω ένα τεράστιο ευχαριστώ στους γονείς μου, Εμμανουήλ Κυδωνάκη και Μαλαματένια Σαραντουλάκη, οι οποίοι παρόλες τις δυσκολίες πάντα ήταν και συνεχίζουν να είναι στο πλάι μου και στηρίζοντάς με όποιον τρόπο μπορούν. Δε μου αρνήθηκαν ποτέ τίποτα και πολλές φορές έχουν στερηθεί πράγματα για μπορώ εγώ να προοδεύω και να εξελίσσομαι. Επίσης δεν είναι λίγη η βοήθεια που έλαβα από τα αδέλφια μου σε ηθική ψυχολογική στήριξη.

Table of Contents

Introduction

RNA biology

It shouldn't be considered overstatement if we say that RNA molecules are, together with proteins, of most important, multifunctional, dynamic and versatile biological factors. They are linear oligo- or poly-meres that consists of 4 ribonucleotides (Adenosine, Guanosine, Cytidine and Uridine). Due to complementarity between their nitrogenous bases, RNA molecules form numerous folding patterns that result in a plethora of various 3D structures. In addition, they even participate in enzymatic functions including the most remarkable one: protein synthesis[1]. RNA molecules are found through all the life-associated forms from viruses to higher eukaryotes.

RNAPII-mediated transcription

In eukaryotes, RNA molecules are in single stranded form and are produced by a DNA-directed process known as transcription. Transcription is catalyzed by at least 3 RNA polymerase enzymes known as RNA polymerase I, II and III. Each one of them catalyze the production of distinct RNA classes; RNA polymerase I catalyze the production of ribosomal RNAs (rRNAs) and RNA polymerase III catalyze the production of transfer RNAs (tRNAs) and one of the 5S rRNA. The rest of the RNA molecules are transcribed by the RNA polymerase II (RNAPII). RNAPII transcribes all the protein coding RNAs, known as messenger RNAs (mRNAs) and most of the classes of non-protein coding RNAs (ncRNAs) such as small RNAs (sRNAs), microRNAs (miRNAs), small interference RNAs (siRNAs), circular RNAs (circRNAs), enhancer RNAs (eRNAs) and long non-coding RNAs (lncRNAs)[2].

Transcription process can be discriminated in 3 main distinct steps: *initiation*, *elongation,* and *termination*. All these steps are marked by well-known molecular events.

Initiation

The crucial first step for transcription *initiation* is the assembly of RNAPII on the promoter of the gene. General transcription factors for RNAPII (TFIIs) are recruited one by one on the core promoter binding sites in coordination with Mediator, a very large protein complex that coordinates and regulates transcription initiation. When RNAPII binds to the pre-initiation complex, transcription is ready to begin. The starting signal is the phosphorylation of RNAPII by CDK7 kinase, subunit of TFIIH. This phosphorylation occurs on ser5 residues of the repetitive heptapeptide C-terminal domain (CTD) of RNAPII. At this point RNAPII transcribes a few base pairs (bp) downstream of the promoter when it is paused, a step known as promoter proximal pause (PPP)[3].

Noteworthy, RNAPII molecules can bind to the promoter and start to transcribe bi-directionally. The anti-sense products, known as PROMPTs (Promoter-Upstream Transcripts) are 5′-capped transcripts of few hundreds bp[4]. Their role isn't clear yet, but a few studies indicate that PROMPTs interact with eRNAs to promote interaction of the TSS with specific distal enhancers[3,5].

Promoter proximal pause

In Metazoans, transcription initiation step is put "on hold" at the PPP site, that is located approximately 30-60bp downstream of the transcription start site (TSS). At PPP site RNAPII is stalled and the pause is maintained by binding of DRB Sensitivity-Inducing Factor (DSIF) and Negative Elongation Factor (NELF) on RNAPII[6]. Note that both DSIF and NELF bind to nRNA too. At this point of the process, RNAPII transition to elongating complex, known as PPP release, occurs when positive Transcription Elongation Factor beta (p-TEFb) phosphorylates DSIF, NELF and RNAPII CTD at ser2 via its CDK9 subunit. After that, transcription machinery keeps producing and elongating the RNA molecules[2].

PPP serves as another crucial step for fine-tunning and synchronizing gene expression. Many developmental genes exhibit temporal flexibility in their expression due to PPP [7]. In addition, via TC-NER mechanism (Transcription-coupled Nucleotide Excision Repair), transcription is engaged in detection and repairing of DNA lesions caused by genotoxic stress such as UV exposure. Under these conditions, a global PPP release from all expressing genes unleashes in the genic regions elongating RNAPII complexes, promoting uniform and DNA lesion recognition over all transcribed regions[8].

Elongation

PPP release marks the start of the elongation phase. During this phase, most of the RNA molecules is produced. While elongating RNAPII is considered to transcribe continuously in a forward direction, it undergoes many pausing/stalling and backtracking events.

RNAPII could pause when it encounters DNA bound factors. A remarkable example of RNAPII pausing is its encounter with a nucleosome. As RNAPII is incapable of transcribing histone-wrapped DNA, various transcriptional elongation factors, such as PAF1 complex, chromatin remodelers and histone chaperones such as FACT complex, are engaged to facilitate the pass-through of transcription machinery along nucleosomes[9]. Additionally, histone variants, such as H2A.Z, and histone modifications, such as H2B ubiquitylation (H2Bub), facilitates the process[10]. Beyond nucleosomes, other physiological barriers can also lead to RNAPII stalling. One well-characterized condition is the transcription-replication conflict, which occurs when DNA replication machinery clashes RNAPII in a transcribing locus. Resolving this conflict is crucial for DNA stability as defects in the machinery results in DNA damage[6].

RNAPII could be stalled when it encounters “non-scheduled” obstacles, such as DNA lesions or/and breaks (including nicks or double strand breaks). As mentioned earlier, RNAPII is part of the detection system of bulk DNA damage[8]. In cases of DNA breaks, RNAPII is stalled and is subsequently released from the chromatin either to be recycled or to be degraded[6].

Most of the time, along with RNAPII pausing/stalling, several backtracking events are observed. During backtracking, RNAPII undergoes backward translocation approximately 2bp. TFIIS factor has a fundamental role in the reactivation of backtracked RNAPII, as it mediates RNA cleavage to remove those 2bp resulting in a new 3′end in the RNAPII active site. Backtracking is an essential elongation process, since in TFIIS mutants present R-loop formation which is associated with widespread genomic instability and transcription stress[6]. This mechanism is also implicated in excision of mis-incorporated RNA bases[11].

Termination

Termination is the final phase of the transcription process and occurs when RNAPII has transcribed the entire genic region. Transcription termination has been extensively studied in protein-coding genes and the most accepted mechanism model is as follows: At the end of the gene, specific sequence features mark the termination site, such Poly-Adenylation Site or *PAS*, typically represented by 5′-AAUAAA-3′ sequence, that usually is flanked by U-rich sequences. When these regions have been transcribed, Cleavage and Polyadenylation machinery (CPA), via its catalytic subunit CPSF73, cleaves the transcript. This process is facilitated by several molecular events that affect the RNAPII’s conformation, processivity and velocity, such as de-phosphorylation of DSIF and phosphorylation of Thr4 on the CTD of RNAPII. When RNA molecules are cleaved, RNAPII continues to transcribe downstream of the gene. However, it produces a non-capped RNA molecule, that is recognized by XRN2, a 5′-3′ exonuclease and is degraded. As RNAPII transcription rate is reduced, XRN2 outpaces the RNAPII, resulting in its dissociation from DNA. This model combines elements of two prevalent models, Allosteric model (transcription termination occurs due to conformational changes in RNAPII) and torpedo model (where cleavage at PAS trigger 5′-3′ exonuclease degradation of RNA produced downstream of the gene, outpacing the RNAPII and result in its dissociation from chromatin)[12].

One well-known yet less understood phenomenon is the premature transcription termination which is defined as the transcription termination resulting in shorter RNA molecules[13]. The most well-characterized premature transcription takes place at the 5′end site of the gene, particularly at the PPP sites. At these sites most of the paused RNAPII are released from the chromatin instead of transitioning to elongation complexes to continue transcription of the underlying gene[12]. However, premature termination can occur at any location within the gene and at any point during transcription, resulting

in either stable or unstable shorter RNAs. Noteworthy is the fact that premature transcription could take place independently of the existence of PAS[14]. In metazoan, premature termination appears to be facilitated by Integrator complex. Integrator's catalytic subunit, Int11, cleaves the transcript at the premature termination site and XRN2 results in RNAPII dissociation, similar to combined model (*see the previous paragraph*). In addition, Integrator complex has a substantial role in transcription termination of non-coding RNAs such lncRNAs, snRNAs, snoRNAs, eRNAs and PROMPTs[15].

Co-transcriptional RNA processing

It is well-known that in eukaryotic cells, the majority of RNA molecules undergo extensive processing to ensure their functionality and stability. Classical processing events include 5′end capping, 3′end processing, with the most prominent being the poly-adenylation process, transcript maturation via splicing events and epitranscriptomic modification, such as methylation on the N6 of the adenine base (m6A). All of these processes, at least partially, commence and conclude during transcription. Therefore, they are referred to as '*co-transcriptional processes*' and they are significantly intertwined with the transcription process itself.

5′end capping

Capping is the first RNA modification process that happens 5′end of nRNAs when RNAPII is located 25-50bp downstream of the TSS. The cap structure is defined as the guanosine dinucleotide linked by a 5′-5′ triphosphate bond, with a methyl group attached to N7 position of the upstream guanosine (m7G). In eukaryotes, one more methylation event takes place co-transcriptionally on the 2-′OH group of the +1 nucleotide.

Capping has a crucial role in ensuring transcript stability by protecting it from 5′-3′exonuclease degradation both inside and outside of the nucleus. In addition, capping is essential for subsequent RNA processing events, such as splicing and 3′end processing, facilitating by cap-binding complex (CBC).

Splicing

Splicing is the most well-known and well established co-transcriptional process. It facilitates the removal of non-coding introns from the newly transcribed RNA molecules, contributing to their maturation process. This process is conducted by the large U1, U2, U4, U5, U6 ribonucleoprotein complexes (snRNPs) which carry snRNA U1, U2, U4, U5 and U6 molecules, respectively. All of them are assembled at the transcription site to form the splicing machinery, known as spliceosome[16].

Splicing sites have specific sequences that can be recognized by the spliceosome complex snRNPs. At 5′splice site (5′SS), there exists a highly conserved GU dinucleotide that either is flanked

upstream by AG dinucleotide or downstream from RAG, where R represents a purine. At the 3′end (3′SS), there exists a highly conserved AG dinucleotide flanked upstream by a pyrimidine nucleotide. Between 5′SS and 3′SS it is necessary to exist a branch point or BP (YNYUR**A**Y), where N represents all nucleotides. Between BP and 3′SS, there exist a polypyrimidine tract[17]. Additional *cis*-regulatory elements that contribute to splicing are the exonic or intronic splicing enhancers or silencers. These elements are sequences, in exonic or intronic regions of the nRNA, respectively, that are recognized by splicing regulatory proteins (hnRNPs and SR proteins) and contribute to the selection of the appropriate 5′SS and 3′SS pair[18].

Spliceosome assembly on the splicing elements is a strictly coordinated procedure. Initially, different snRNPs binds to the splicing elements, with U1 snRNP binding to 5′SS and U2 snRNP binding to the BP[17]. Additional splicing factors, such as splicing factor 1 (SF1) and U2 snRNP Auxiliary Factor (U2AF), contribute to the primary spliceosome assembly. SF1 binds to the BP and promotes U2 snRNP binding, while U2AF binds to both polypyrimidine tract and 3′SS[16]. U4, U5 and U6 snRNPs are assembled in a tripartite complex and join the other two snRNPs. Initially, 5′SS and BP come into proximity and the 2′OH of the adenosine nucleotide at the BP attacks the 5′-phosphate group of the guanosine nucleotide at the GU site, resulting in a 2′-5′ phosphodiester bond and an intermediate lariat structure. Subsequently, with assistance from other transcription factors, the 3′SS comes in proximity and the free 3′OH of the 5′ exon attacks the 3′SS. In this manner, the spliced transcript is released by the spliceosome[17].

The majority of human genes undergo alternative splicing, which is defined as the production of a transcript that either contains or omits a sequence part. The most prominent types of alternative splicing are the exon skipping or inclusion, intron retention and alternative 5′SS or 3′SS usage. Consequently, alternative splicing contributes to the variety of the RNA isoforms, providing a large number of protein coding mRNA as well as many gene expression regulatory effectors[18]. The speed of RNAPII has great impact on the combination of exons/introns generated during co-transcriptional splicing and the robust patterns observed in different cell types or upon stress suggest that deterministic associations of snRNPs and conformation of nascent RNAs must occur[19]. Chromatin state at these loci is thought to be a great influencer and regulator of these alternative processes[20]

3′ end processing

The 3′end processing of RNA transcripts is spatially and temporally coupled with the transcription termination process[12]. The most prominent 3′end processing event is the polyadenylation process, which occurs in the majority of the protein coding pre-mRNAs and in many lncRNAs[21]. Polyadenylation is essential for RNA molecule stability, both in nucleus, preventing degradation by

RNA exosome 3′-5′exonucleolytic complex and in the cytoplasm, preventing degradation by other nucleases[12].

Two crucial factors for poly(A) tail formation are polyadenylation polymerase (PAP) and polyadenylation binding protein (PABPN1), both of which are subunits of the CPA termination complex. PABPN1 binds to the cleaved RNA transcript and regulates the function of PAP, which adds adenosine nucleotides to the 3′end of the transcript[22].

The majority of metazoan genes have two or more alternative polyadenylation sites, which can be positioned in intronic regions, exonic regions or downstream of the terminal exon in the 3′ untranslated region (UTRs). The selection of an intronic PAS result in the elimination of the downstream exons and retention of the intronic sequence in the mature transcript. In contrast, the selection of an exonic PAS or a PAS located in the 3′UTR region affects the representation of the 3′UTR itself in the mature transcript. 3′UTR play a significant role in the regulation of the translation. Consequently, the choice of a PAS is crucial for the regulation and function of the newly transcribed RNA molecule[22].

nRNA

Transcription and co-transcriptional processes have been studying for many years, alongside the research on genome 3D conformation, DNA replication and integrity. Most evidence demonstrating the interplay among these nuclear processes has emerged through the development and advancement of next generation sequencing approaches[23]. However, in all these studies, nRNA has been considered solely as the transcription product, that undergoes many processing event and maturation. Although, it has been neglected the role of the nRNA on chromatin functions and gene regulation, in recent years an increasing number of studies have begun to acknowledge the role of these molecules in the complex "*nucleus reality*".

nRNA and chromatin modifications

An increasing number of studies have accumulated data that nRNA molecules are implicated in chromatin modification and remodeling processes. H3K36me3 is highly deposited at exonic regions of expressed genes. It has already been shown that RNA binding factors, such as U2 snRNP complex and hnRNP L complex, mediate the accumulation of H3K36me3 via SETD2. Similarly, Polycomb Repressor Complex 2 (PRC2) chromatin binding is reversely correlated with the levels of nRNA, indicating that nRNAs could positively feedback their own production by reducing H3K27me3 and its repressive function. WDR5 subunit of MLL histone methyltransferase seems to catalyze the H2K4me2/3 upon interaction with nRNA molecules[24].

nRNA and 3D chromatin conformation

The implication of nRNA in higher order chromatin organization has become apparent in recent years. RNA immunoprecipitation (RIP)-seq experiments have identified the interaction of STAG2, subunit of cohesin - a known higher order chromatin conformation factor - with RNA molecules[24,25]. However, further investigation is necessary to understand the interaction between cohesin and RNA since this study concludes that it is restricted to a low number of RNA molecules[25]. In contrast to cohesin, the functional interaction of CCCTC-binding factor (CTCF), a crucial 3D chromatin conformation factor, is more well-established with RNA molecules,[26,27] even extending to nRNAs[26].

The vast majority of data on nRNA and chromatin 3D organization have emerged from the recently developed "*all to all*" RNA versus chromatin interaction approaches, which can map the nuclear RNA against their underlying chromatin genome wide. In principle, based on a bivalent artificial linker, all these approaches capture RNAs in proximity to chemically crosslinked chromatin, ligating the RNA at one end and DNA on the other. Following pull-down, hybrid nucleic acid fragments are incorporated into libraries and analyzed by deep sequencing[28]. Except the first approach, Mapping RNA-genome interactions (MARGI)[29], all the other approaches conduct linker ligation *in situ* within intact nuclei[28]. This optimization decreases the background noise from randomly diffused RNA molecules. However, despite this optimization, background noise still remains at high levels, necessitating sophisticated methods for background deduction have been used[30]. While these methods revolutionize the field of RNA-chromatin interaction, it is important to exercise caution in interpreting the biological implications of these data. Although these experiments provide a holistic view regarding the spatial interaction between RNAs and chromatin, establishing functional interaction between a single RNA and its respective chromatin site requires more specific functional experiments.

These global RNA-chromatin interaction approaches have highlighted the extensive interactions of nRNA molecules with many genome regions, excluding their transcription site. While there is good understanding and consensus from researchers on how nRNA interacts with distal *cis* chromatin regions, there are conflicting evidence regarding inter-chromosomal interactions[29–31]. In addition, exploiting these interaction maps between promoters and their associated enhancers could be generated, similar to Hi-C experiments, even in a condition- or cell-specific context[30,32].

nRNA folding

Examining the all the above information and considering that nRNA is a long polymer similar to DNA, it is natural to wonder how these molecules are spatially organized within nuclei as well as how they affect or are affected by nuclear processes. Many studies have addressed the issue of nRNA folding.

It is apparent that nRNA molecules form intramolecular secondary structures as they emerge from the RNAPII, due to base complementarity[33]. nRNA folding seems to be a significant regulator of RNAPII speed, as more folded molecules increase the RNAPII rate, decreasing backtracking events[14]. Additionally, nRNA folding can interfere with splicing machinery. Certain folded structures can either present the splicing sites to the splicing effectors in loops or can mask and prevent their accessibility due to a stem structure, thus regulating the alternative splicing[34].

In agreement with the previously mentioned studies, imaging approaches via fluorescent *in situ* hybridization (FISH) have visualized the nRNA molecules to be organized and localized in near vicinity to their transcription site[35]. However, all of these studies focus on specific genes and their transcription and co-transcriptional processes. As far as we know, there are no studies that investigate the conformation and spatial organization of nRNA and if nRNA is important for feedback on gene regulation.

Master Thesis Aim

In this study, we attempted to develop and optimize techniques and tools for studying nRNA conformation and its interaction with other factors, such as chromatin. In this way, we can address many scientific questions regarding nRNAs and how they affect chromatin conformation as well as processes, such as transcription, RNA processing etc. Specifically, we attempted to standardize GRID-seq and CRISPR-Cas9 pull-down.

Material and Methods

Cell line culture

Huh.7 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (*Thermo Scientific* # 11965) supplemented with 10% FBS (*Gibco* 10437-028) and 1% Gentamycin (*AppliChem*, 1405-41-0) in a sterile, humidified chamber with 5% CO_2 at 37ºC. The cells were grown in various cell culture formats, such as 25cm^2- and 75cm^2- flasks (*ThermoScientific*, 169900 and 178905) and 15cm-dishes (*Falcon A corning brand,* 353025). All treatments were conducted in a cell culture laminar flow cabinet sterilized by UV exposure. For cell passaging, when culture was at 70% to 90% confluency, a trypsinization protocol that use Trypsin/EDTA (*Gibco*, 15400-054) solution diluted 10 times in Phosphate Buffer Saline (PBS; 137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 2mM KH_2PO_4) was applied. The cells were incubated for at least 4min at 37ºC. The trypsinization reaction was inhibited by adding double volume of culture medium (DMEM/10%FBS/1%Gentamycin). The cells were precipitated by centrifugation at 500g for 5min at Room Temperature (RT) and the supernatant was discarded. The pellets were re-suspended in fresh culture medium and split into the cell culture vessels.

Cell fixation

1% formaldehyde fixation

For 1% formaldehyde fixation (FA, *Sigma Aldrich*, F8775-25ML), cells cultured in 15cm dishes on 80% to 90% confluency were used. The appropriate amount of FA (36%) was added to the culture medium. The cells were incubated for 10min at RT with gentle rocking. The fixation reaction was inhibited by adding 126mM glycine (Glycine for molecular biology, *AppliChem*, A1067) and incubation at RT with gentle rocking. The cell medium was discarded and, the cells were washed 3 times with 1X PBS with each wash taking 5min. Cells were harvested by scraping (Cell lifter, *Corning*, 3008) into 1X PBS. The collected cells were then precipitated by centrifugation (Tabletop Micro Refrigerated Centrifuge Model 3520, *KUBOTA*) at 1000g for 5min at 4ºC.They were either used for downstream procedures or stored at -80ºC.

Double Fixation

For double fixation, we used both 3% FA and 2mM Disuccinimidyl Glutarate (DSG). 10mg DSG reagent (*Cayman Chemical*, 20646) were resuspended in Dimethyl sulfoxide (DMSO, *AppliChem*, A3672) to achieve a final concentration of 500mM. This solution was stored at -20ºC.

For double fixation, cells cultured in 15cm dish at 70% to 80% confluency were used. The cells were harvested in their culture medium by scraping, were precipitated, and were resuspended in 1X PBS. After cell counting using Neubauer plate, DSG at final concentration 2mM in 5-6 million cells resuspended in 1X PBS was added and incubated at RT for 45min with gentle shaking. Immediately after that, the cells were precipitated by centrifugation at 1000g for 5min at 4ºC and washed once with equal volume of 1X PBS. The cell pellets were then resuspended in 3% FA/PBS 1X buffer and incubated at RT for 10min with agitation. The reaction was inhibited by adding glycine (381mM), and mixture was incubated for 5min at RT. The cells were centrifuged at 1000g for 5min at 4ºC and washed twice by PBS 1X.

Cells that were fixed only with 3% FA were treated as above, except the DSG fixation step.

Nuclei isolation

Cells fixed by 1% FA were resuspended and incubated in cell lysis buffer (50mM HEPES (HEPES sodium salt, *Sigma Aldrich*, RES6007H) (pH 7,5), 140mM NaCl (*MERCK*, 1.06404), 1mM EDTA (Ethylenediaminetetraacetic acid disodium salt dihydrate, *Sigma Aldrich*, E5134-1KG) (pH=8), 10 % glycerol (Glycerol anhydrous for molecular biology, *AppliChem*, A2926), 0,5% IGEPAL (MERCK, 18896), 0,25% Triton X-100 (*MERCK*, T8787), 1mM PMSF (*AppliChem*, A0999), Protease Inhibitor 1X (*Sigma Aldrich*, S8820)) for 10min at RT and 10min on the ice. The cell lysate was

centrifuged at 1000g for 5min at 4ºC and resuspended in washing buffer (10mM Tris-Cl (pH=8,1) (*AppliChem*, A1086), 200mM NaCl, 1mM EDTA (pH=8), 0,5mM EGTA (pH=8), 1mM PMSF, Protease Inhibitor 1X). It was centrifuged as above, resuspended in 'sonication' buffer (10mM Tris-Cl (pH=8,1) , 100mM NaCl, 1mM EDTA (pH=8), 0,5mM EGTA (pH=8) (*Sigma Aldrich*, E3889), 0,1% Sodium Deoxycholate (NaDoc), 0,5% N-Lauroyl sarcosine (*Sigma Aldrich*, 61747), 1mM PMSF, 1X Protease Inhibitor) and incubated overnight at 4ºC.

Cells fixed by both 3%FA and DSG were adjusted in 1X Tango buffer (33mM Tris-acetate (pH 7.9 at 37°C), 10mM magnesium acetate, 66mM potassium acetate, 0.1mg/ml BSA, *ThermoFischer*, BY5) and 0.1% Tween20. After centrifugation at 1500g for 5min at 4ºC, the pellet was resuspended in nuclei lysis buffer (1X Tango buffer and 0.2% SDS) and incubated at 62ºC for 10min. SDS activity was inhibited by 1.5% TritonX100 at RT.

Sonication and input reverse crosslinking

All samples were sonicated in a final volume 2ml sonication buffer (*Covaris*, S220 Focused-ultrasonicator). Sonication settings: Peak power = 500, Duty Factor = 20 and Cycles per burst = 1000 for 5-12min at 6ºC.

A 5% fraction of the sonicated chromatin was collected and was incubated with 200mM NaCl and 0.5µg/µl Proteinase K at 65ºC overnight. The DNA amount was assessed with nanodrop and gel electrophoresis.

Classical PCR and analysis by agarose gel electrophoresis

All primers, except primers for cloning experiment (see below), were designed using *PrimerQuest Tool* (Integrated DNA Technologies, IDT; https://www.idtdna.com/pages/tools/primerquest) and Primer Blast (NCBI, https://www.ncbi.nlm.nih.gov/tools/primer-blast/). All primers were purchased by (*macrogen*). All primer sequences are presented on the *table 1, 2 and 3*.

For all diagnostic PCRs, Taq Polymerase (*Minotech*, 203) and its corresponding buffer (50 mM KCl, 10 mM Tris-HCl (pH=8.5) at 25ºC, 1.5 mM $MgCl_2$, 0.1% Triton X-100) were used. In brief, 0.04 to 0.05unit/µl enzyme, 200nM dNTPs (*ThermoFischer*, R0181), 500nM for each primer and 1-30ng DNA template were used for each reaction. Template was amplified using the following settings: At first, denaturation at 94ºC for 2min, (94ºC for 45sec, 57ºC-67ºC for 30sec and 72ºC for 20sec-1min) for 33-40 cycles and final extension at 72ºC for 5min. Annealing temperature was determined based on Tm of primer pair and time of extension at 72ºC based on the expecting amplicon length.

For all PCR-based DNA synthesis, high fidelity Taq polymerase (Q5® High-Fidelity 2X Master Mix, *NEB* M0492) was used. In brief, in each reaction, 500nM per primer and 0.1ng DNA template was included. Template was amplified according to the following settings: At first, denaturation at 98ºC for 30sec, (98ºC for 10sec, 57ºC-65ºC for 10sec and 72ºC for 5sec-50sec) for 28-33 cycles and final extension at 72ºC for 2min.

Every PCR product was analyzed using agarose gel electrophoresis. In brief, agarose (UltraPure™ Agarose, *Invitrogen*, 16500500) diluted in 1X TAE buffer (40mM Tris, 20mM Acetic acid and 1mM EDTA) at final concentration 0.8% w/v to 2.2% w/v and Ethidium Bromide (0.5μg/ml). 2-5μl PCR product, mixed with loading buffer (6X Loading Dye Solution, *Minotech*, K12) were electrophorized at 4-6V/cm for at least 30. Gels were visualized under UV radiation (*Bio-Rad* Gel Doc XR System w/ Universal Hood II).

Gel extraction and quantification

DNA samples were electrophorized as above, the bands on the expecting length were isolated and were weighed. Gel extraction was conducted following the Kit manufacturer's instructions. In brief, gel bands were resuspended in lysis buffer (the volume depends on band's weight) and incubated at 60ºC for at least 10min. An equal volume of isopropanol was added, and the samples passed through a spin column by centrifugation at 15000g for 1min at RT. Columns were washed with washing buffer twice and eluted using elution buffer pre-heated at 50ºC. All samples were further cleaned up by Ethanol precipitation; 2,5 volumes 100% Ethanol (EtOH) 0.1 volume NaAC 3M (pH=5.2) were incubated for 1h at -80ºC, centrifuging at 15000g for 33min at 4ºC. Pellets were washed by 70% EtOH once and re-centrifuged at 15000g for 5min at 4ºC. Dry pellets were resuspended in ddH_2O.

All samples were quantified by spectroscopy at 260nm and 280nm (Nanodrop, *DNA Technologies Core*).

RNA extraction and cDNA synthesis

For RNA extraction and isolation from cells NucleoZOL Reagent (*Macherey-Nagel*, 740404.200) was used following manufacturer's instructions. In brief, 10^6 cells were diluted in 500μl Nucleazol reagent and homogenized them. 200μl ddH_2O were added, were vortexed harshly and were incubated at RT for 5min. Afterwards, they were centrifuged at 12000g for 15min at 4ºC, were collected the supernatant and were added an equal volume 100% isopropanol. They were incubated for 10min at RT and were centrifuged at 12000g for 10min at 4ºC, discarding the supernatant. The pellets were washed with 75% EtOH twice and resuspended dried pellets in RNase-free ddH_2O.

All RNA samples were treated with DNase (TURBO™ DNase, *Invitrogen*, AM2238). In brief, total RNA (less than 10µg) was incubated with 2units DNase in its compatible buffer at 37ºC for 30min. Reaction volume was increased to 200µl and mixed with an equal volume of Phenol-Chloroform-Isoamyl alcohol (25:24:1). After centrifugation at 15000g for 5min at RT and collection aqueous phase, samples were mixed with an equal volume chloroform-isoamyl alcohol (24:1) (Chloroform: Isoamyl Alcohol 24 : 1 *BioChemica*, A1935,0500) and were centrifuged as above. Aqueous phase was collected and was incubated with 2.5X volumes of 100% EtOH, 0.1 volumes of NaAc 3M (pH=5.2) at -20ºC overnight. After centrifugation at 15000g for 33min at 4ºC, pellets were washed with 75% EtOH once and dried pellets were resuspended in RNase-free ddH_2O. RNA purification was validated by nanodrop measurement and agarose gel electrophoresis (1-1.5% w/v).

For cDNA synthesis, M-MuLV Transcriptase (*Minotech*, 801-1) was used. In brief, 1µg of RNA sample was incubated with 200ng Random Hexamers (*ThermoFischer*, SO142) and 1mM dNTPs at 50ºC for 5min. Following that, 5mM DTT, RNase inhibitor (RiboLock, *ThermoFischer*, EO0382), 1X RT assay buffer (Minotech, 801-1) and 200units M-MuLV Transcriptase were added. Samples were incubated at RT for 5min and at 42ºC for 1h. The resulting cDNA was stored at -20ºC.

Cloning

Vector digestions

We tried to construct a plasmid vector containing dCas9 gene and two gRNA cloning sites. For this purpose, we used the All-in-One (AIO Puro, addgene's number: 74630) plasmid, which expresses Cas9 D10A (Nickase or nCas9) tagged with 3xFLAG and features two sites for gRNA cloning. In addition, we utilized the pRP-Puro-EF1A>FLAG/dCas9*/3xNLS, which expresses dCas9 tagged with 1xFLAG peptide. Our objective was to convert nCas9 to dCas9 by substituting Histidine 840 for Alanine. To achieve this, 2µg AIO Puro plasmid were digested with 20unit EcoRV (EcoRV-HF®, *NEB*, R3195S) in Cutsmart buffer 1X (*NEB*, B6004S) at 37ºC for 2h and purified by the phenol-chloroform-isoamyl alcohol protocol as 'RNA extraction' paragraph. The linearity of DNA was assessed by electrophoresis on agarose gel (0.8%). Linear DNA was further digested with 25units BsmI (*NEB*, R0134S) in Cutsmart buffer 1X at 65ºC for 2h. The reaction product was electrophorized in an agarose gel (0.8%) and the bands of interest were extracted with NZY tech kit (see above).

Insert preparation

The corresponding DNA segment that has the H840A mutation was amplified (Q5 PCR-mediated DNA synthesis, as described above) from the pRP-Puro-EF1A>FLAG/dCas9*/3xNLS with primers that span the EcoRV recognition site and ~100bp downstream of the BsmI recognition site

(**table 1**). The PCR products were purified by phenol-chloroform-isoamyl alcohol protocol and then were digested with 25units BsmI in Cutsmart 1X buffer. The digested PCR products were subjected to end-repairing by incubating them with 5units Polynucleotide Kinase (PNK, T4 Polynucleotide Kinase, *NEB*, M0201S) in Ligase reaction buffer 1X (T4 DNA Ligase Reaction Buffer, *NEB*, B0202S) for 30min at 37ºC. AIO Puro-derived fragment and digested PCR product were mixed in 1:3 molar ratio and incubated with Ligase 20units (T4 DNA Ligase, *NEB*, M0202S).

Transformation

2-5ng plasmid DNA (pDNA) or 5µl ligation reaction were mixed with competent E. coli (DH5a) and were subjected to heat shock at 42ºC for 45sec. Immediately after they were incubated on ice for 2min. The shocked bacteria were incubated in Luria Bertani broth (LB, 1% w/v tryptone (*Sigma Aldrich*, 91079-40-2) and NaCl and 0.5% w/v Yeast Extract (*Sigma Aldrich*, 70161-500G) in absence of antibiotics at 37ºC for 1h. Subsequently, bacteria were streaked on an LB-Agar-Ampicillin-containing petri dish (LB, 1.5% Agar (*Sigma Aldrich*, 05040) and 100µg/ml Ampicillin (*Sigma Aldrich*, A9393-5G)) and were incubated at 37ºC for 18h. All treatments with bacteria were conducted in close proximity to flame, using clear and sterile equipment.

Plasmid isolation and sequencing

Single colonies without satellites were selected and cultured in 5ml LB broth with 100µg/ml Ampicillin at 37ºC overnight with agitation. 1ml of the culture was used to isolate pDNA by homemade Miniprep protocol. In brief, bacteria cells were precipitated and then were diluted in solution I (50mM Glucose, 10mM Tris and 25mM EDTA). Solution II (0.2M NaOH, 1%SDS) was added and was incubated for 2-3min at RT. Solution III (5M potassium acetate) was added and was incubated for 5min at RT. Chloroform was added and was centrifuged at 15000g for 5min at 4ºC. The upper phase was collected and incubated with 2.5xVolumes of 100% EtOH in -20ºC for at least 20min. Samples were centrifuged at 15000g for 5min at 4ºC and pellets were washed with 70% EtOH. After centrifugation as above, pellets were dried and resuspended in ddH_2O. pDNA was 100µg/µl RNase A treated at 37ºC for 1h. pDNAs were electrophorized on agarose gel (0.8% w/v). Diagnostic digestions with EcoRV and BsmI validated the clone carrying the recombined plasmid.

The pDNAs that were sequenced and used for other downstream assays were cultured in greater medium volumes and isolated by Midi prep (ZymoPURE II™ Plasmid Midi prep, *Zymo Research*, D4201). In brief, 100ml of the bacteria culture carrying the plasmid of interest was pelleted and well-resuspended (no harsh vortex) in P1 buffer. P2 buffer was added and incubated for 2min at RT. P3 buffer was added, was incubated for 1min and all bacteria lysates were loaded on a Luer-lock syringe

filter and were incubated for 5-8min. Lysates were filtered and loaded on the kit's column. The column was washed 3 times with washing buffers, was dried and was incubated with elution buffer at 50ºC for 10min. All plasmids quantity and quality were evaluated by Nanodrop and gel electrophoresis (0.8% w/v).

70ng/µl recombinant plasmid (dCas9-AIO-puro) was sequenced by sanger sequencing (*Azenta Life Science*) using 2.5nM sequencing primer R_Paio_Seq_Gib (**Table 1**).

qPCR

For qRCR reactions we utilized Kappa Mix (*Roche*, KK4652). In each reaction we used 200nM from each primer, 4% of the volume from RT reaction for cDNA and 1ng for gDNA. Template was amplified using the following settings: Initially, denaturation at 95ºC for 2min, (95ºC for 10sec, 57ºC-65ºC for 20-30sec) for 35-40 cycles. For melting curve analysis, fluorescence was measured in the temperature range 55ºC to 95ºC with a step of 0.5ºC.

For qPCR data analysis we employed ΔΔCt method. This type of analysis utilizes Ct values to calculate the Fold Change or FC (2^(-ΔΔCt)) value for a specific feature between two different conditions normalized against a reference feature. For cDNA: $\Delta\Delta Ct = [(Ct_{target} - Ct_{GAPDH})_{Spliced} - (Ct_{target} - Ct_{GAPDH})_{Non\text{-}spliced}]$. For Pull-down: $\Delta\Delta Ct = [(Ct_{enriched} - Ct_{input})_{target} - (Ct_{target} - Ct_{GAPDH})_{other\ locus}]$.

sgRNA design and IVT

gRNAs' sequences were designed using benchling, a free online tool (https://benchling.com/signin/welcome). gRNAs were selected to have the following features: GC content over 45%, off target score over 60% and length equal to 20nt. For gRNAs which didn't begin with GG dinucleotide we added one or two at the start. In addition, distance between upstream and downstream viewpoint was more than 2kb and the spacing between gRNAs that target the same viewpoint was at least 150bp. Sequences of gRNAs are presented on the **Table 4**.

The template for sgRNA transcription was synthesized by PCR process (Q5 Hi Fidelity PCR) using a specific oligos that include the T7 polymerase binding site, the target sequence, and a portion that complements a commonly used reverse primer encompassing the remaining sgRNA sequence (**Table 4**). All PCR products were electrophorized and were subjected gel extraction. For (IVT) reactions we used HI Scribe T7 Transcription Kit (*NEB*, E2040S). In brief, 250ng dsDNA was mixed with 7.5mM for each NTP, 0.75X from reaction buffer and enzyme mix equal to 10% of the final volume of the reaction. Samples were incubated overnight at 37ºC. 2 units of DNase was added and

incubated at 37ºC for 30min. The DNase-treated IVT product was then purified as is described on the 'RNA extraction' paragraph.

Pull-down

The complex of SNAP-tagged dCas9 (EnGen® Spy dCas9 (SNAP-tag®) *NEB*, M0652T) with sgRNAs was formed in dCas9 1X reaction buffer (NEBuffer™ r3.1, *NEB*, B6003) by incubating them in 1:5 molar ratio in presence of 1mM DTT at 37ºC for 1h. RNP complex immobilized on SNAP tag beads (SNAP-Capture Magnetic Beads *NEB*, S9145S) following the manufacturer' instructions. In brief, a 20µl bed volume of beads was equilibrated in dCas9 reaction buffer and then RNP complex was incubated with them for 1h at RT. Fragmented chromatin (or gDNA) was incubated with RNP-bead complex at 4ºC overnight under rotation. The beads were washed 4 times with 1X reaction buffer containing 1mM DTT and 0.05% Triton X-100. Chromatin/gDNA bound to RNP complex was eluted in modified dCas9 reaction buffer (200mM NaCl) and 0.5µg/µl Proteinase K at 65ºC for 4h.

TABLE 1

FOR CLONING EXPERIMENT

Primer name	**Sequence**	**GC content %**	**n-bases**
87_F_AIO_PURO_EcoRV_del3112	GATATAGTGCTGACCCTGAC	50	20
88_R_AIO_PURO_BsmI_4979	CTTCAGCTTCTCATAGTGGC	50	20
166_F_AIO_PURO_EcoRV_del3112_OK	ATAGTGCTGACCCTGACACTG	52	21
95_R_pAIO_seq_Gib	GTGCTTTGTGATCTGCCG	56	18

TABLE 2

FOR SPLICED VS UNSPLICED QUANTIFICATION

Primer name	**Sequence**	**GC content %**	**n-bases**
96_OSMR_e1_F1	TTGCCCCGCAGCTGATTCATA	61.3	21
97_OSMR_i1_F1	CCTCATCTACCACAATTGGCTC	62.1	22
98_OSMR_e2_R1	CTGGTAAGTCCTCAAGGACAGC	64	22
99_OSMR_i2_R1	AATGACTTCATTTGTTCCCGACG	61.1	23
100_RAD23B_e5_F1	TCGGGTGATTCTTCTCGGTC	60.5	20
101_RAD23B_i5_F1	AGCATAGTAGTTCCTGAAATGTTG	60.1	24
102_RAD23B_e6_R1	TTCTCGTAAGACTGACCCGT	58.4	20
103_RAD23B_i6_R1	AAGTGAACTGGTCCTTTAAAATCTG	60.9	25
104_HIF1A_e1_F1	TAGTCTCACGAGGGGTTTCC	60.5	20

105_HIF1A_i1_F1	CTTGTATACACTTTCCATCTCGTG	61.8	24
106_HIF1A_e2_R1	TGGAAGTGGCAACTGATGAGC	61.3	21
107_HIF1A_i2_R1	TCAAAACATTGCGACCACCTTC	60.3	22
108_NUP98_e1_R1	CGCGTTGCCCAATGAATAACA	59.4	21
109_NUP98_i1_R1	ACAGTTTGAGGATTTTGAATGAGAG	60.9	25
110_NUP98_e2_F1	GTCCAAATGTTGAAGTTGTGCC	60.3	22
111_NUP98_i2_F1	AAGGTTTCTCCCTCTTTTGTCCT	61.1	23
112_KMT5B_e1e2_R1	CGCAGGCGGAGAGAACA	57.2	17
113_KMT5B_i1_R1	TCAAGGGTCTGTTTAGCATGATCT	61.8	24
114_KMT5B_e2_F1	CCTCTCGACTGCATTTTTGCC	61.3	21
115_KMT5B_i2_F1	ACTACCCCATATCACAATCAGAAC	61.8	24

TABLE 3			
PULL-DOWN EXPERIMENT			
Primer name	**Sequence**	**GC content %**	**n-bases**
147_HIF1A_UG1_F	TGGATATTGTGTGTGCCACAGC	50	22
148_HIF1A_UG1_R	AGCAGCTATAAGCCAGTGCAGG	54.55	22
149_HIF1A_UG2_F	GGCACCTAGTACAGGGTAATG	52.38	21
150_HIF1A_UG2_R	GCTGTCTTCCTTTGCTTAACTG	45.45	22
151_HIF1A_DG1_F	CTTGACCGCTGGGACTC	64.71	17
152_HIF1A_DG1_R	ATGGTAGCACGCGTCTGTG	57.89	19
153_HIF1A _DG2_F	ACGCCCAGGCTACCTTT	58.82	17
154_HIF1A _DG2_R	TCTCCTTGCCTTAATGGTGT	45	20
156_OSMR _UG1_F	CTTCTGGTTCCAGTAGGTGATAC	47.83	23
157_OSMR _UG1_R	AGACCGGCCCTGATGATTTA	50	20
158_OSMR _UG2_F	CAGTGGCACACAGCAATAAG	50	20
159_OSMR _UG2_R	TAGCTGAGACCAGCCTAGAT	50	20
160_OSMR _DG1_F	GCAGCTTGCAGAGTCCAATTA	47.62	21
161_OSMR _DG1_R	CAGTCCCTTTAAGGAGGAAAGC	50	22
162_OSMR _DG2_F	TATGCAGTGGCAGAAGTGAG	50	20
163_OSMR _DG2_R	CAGCCACCTTACAGCCTATTT	47.62	21
164_OSMR_I1_F	GGGACTCTAAAGTACCATGAC	47.62	21
165_OSMR_I2_R	GGGAGTAGCTAAGTGACAAATA	40.91	22

TABLE 4			
PULL-DOWN EXPERIMENT			
sgRNA name	**Sequence**	**GC content %**	**n-bases**

167_HIF1A_E2_Up_gRNA_1	GAAATTAATACGACTCACTATAGGTAGCAGCTATAAGC CAGTGCGTTTTAGAGCTAGAAATAGC	50	55
168_HIF1A_E2_Up_gRNA_ 2	GAAATTAATACGACTCACTATAGGCACCTAGTACAGG GTAATGTGTTTTAGAGCTAGAAATAGC	45	55
169_HIF1A E2 Down_gRNA_1	GAAATTAATACGACTCACTATAGGCATGGTAGCACGC GTCTGGTTTTAGAGCTAGAAATAGC	65	55
170_HIF1A_E2Down _gRNA_2	GAAATTAATACGACTCACTATAGGTTTAACTCCACCAG GTGTGGGTTTTAGAGCTAGAAATAGC	50	55
171_OSMR E2_Up_gRNA_1	GAAATTAATACGACTCACTATAGGTAATTAAATCATCA GGGCGTTTTAGAGCTAGAAATAGC	40	54
172_OSMR E2_Up_gRNA_2	GAAATTAATACGACTCACTATAGGGTCAGCTGCTCTTG TCTGGTTTTAGAGCTAGAAATAGC	60	54
173_OSMR E2_ Down_gRNA_1	GAAATTAATACGACTCACTATAGGACCTGCATTTCGGT TACACCGTTTTAGAGCTAGAAATAGC	50	55
174_OSMR E2_Down_gRNA_ 2	GAAATTAATACGACTCACTATAGGAAAGCAGGGGGCC TATGCAGGTTTTAGAGCTAGAAATAGC	60	55

Results

3XFLAG-tagged dCas9 cloning

The dCas9-based pull-down system is a known method for determining and studying interactions of specific genomic loci with other macromolecules, such as RNA, proteins, and other genomic loci[1]. We aim to utilize this method to precipitate chromatin both upstream and downstream of an exon sequence of an active gene and to quantify its associated nRNAs. By quantifying the downstream nRNA sequence levels in the upstream locus and vice versa, we could indirectly deduce how nRNA is arranged in the space. If nRNA levels are similar between these two viewpoints, it would indicate that nRNA isn't highly structured. However, if there is a significant difference in nRNA level, it suggests a highly structured and well-arranged nRNA molecule compatible with the idea that co-transcriptional splicing of the nRNA particle will affect the interactions of nRNA with the underlying genomic locus. Two different approaches can be utilized: *in vivo* and *in vitro*.

For *in vivo* experiments, it is necessary the dCas9 and sgRNAs to reach the chromatin of the cell and, as a complex, to bind to the target sequence. The most common method involves transfecting cells with a vector that expresses dCas9 and sgRNAs. For this reason, we attempted to create a plasmid vector that carries 3XFLAG-tagged dCas9 and the sgRNA that target our locus of interest. Our goal is to pull-down the nRNAs associated with the target chromatin.

For the generation of the vector that expresses both dCas9 and sgRNAs, we employed classical genetic engineering approach involving restriction enzymes and ligation (as described in Material and Methods, session 'Cloning'). The recombinant plasmid (**Figure 1A**) was then transformed into *E.* coli bacterial cells. The appropriate clone was assessed through diagnostic digestions (**Figure 1B**). To distinguish between the recombinant plasmid (dCas9-AIO puro) and the initial AIO Puro, we

introduced a single transversion mutation (C-G to A-T) at the sixth position of the primer that targets the EcoRV restriction site (see 87_F_AIO_PURO_EcoRV_del3112 **table 1**). This modification abolishes the digestion ability of EcoRV at that site of the dCas9-AIO Puro. Therefore, after EcoRV treatment, the recombinant plasmid remains circular and intact, whereas AIO Puro becomes linear (**Figure 1B**; *sixth lane*). pRP-Puro-EF1A > FLAG/dCas9*/3Xnls plasmid carries two BsmI restriction sites. The first one is identical to the AIO Puro, and the second one is located between the first site and EcoRV. Consequently, the recombinant dCas9-AIO Puro plasmid carries two BsmI restriction sites (**Figure 1B**; *seventh lane*). In addition to the digestions, dCas9-AIO Puro was sequenced (**Figure C**). Through sequencing we determined that an extra triplet had been added upstream from the EcoRV restriction site (**Figure C**; *lower panel*). This occurred due to an error in primer design. The first 5'-triplet of the primer should have been ignored as it is already contained in the cut AIO Puro vector. By incorporating it in our primer sequence, the new recombinant plasmid has one more triplet that codes for an extra glutamate residue on the position 617. Based on InterPro database we realised that this addition lays on the recognition (Rec) domain of the Cas9, responsible for nucleic acid binding (**Figure D**). Considering that glutamate is a negatively charged residue, we thought that this insertion could significantly modify the binding activity of dCas9 to DNA and decided not to proceed further with this approach.

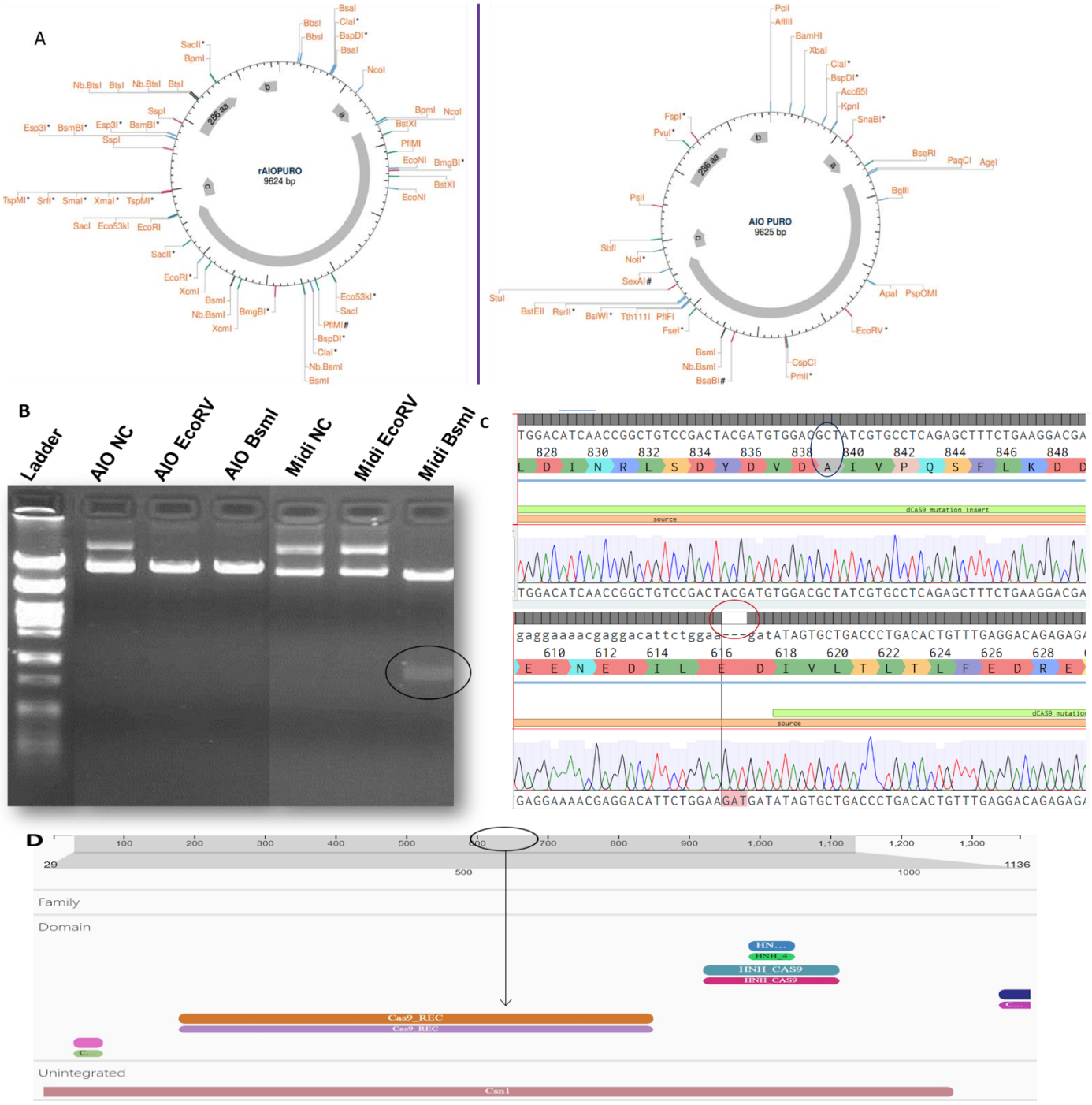


Figure 1: dCas9 vector construction. (**A**) Left panel the recombinant vector map; right panel AIO Puro vector map. (**B**) Digestion with EcoRV and BsmI. Lane 1 ladder, Lane 2-4: AIO Puro vector, Lane 5-7: recombinant dCas9 vector. NC: Negative control. (**C**) Sequencing of plasmid. Upper panel determines the that previous residue was substituted by alanine; lower panel determines the addition of a triplet in the EcoRV restriction site. (**D**) Cas9 domain map shows that the extra triplet lies on the Rec domain.

In vitro dCas9 pull-down

The *in vivo* approach is a very convenient method to indicate putative functional interactions between target chromatin and other macromolecules. However, our aim is to target active gene bodies and quantify the chromatin associated nRNAs (and if needed the proteins or DNA content at these loci). We thought that the binding of dCas9 on the target locus could halt or interfere with the transcription machinery in living cells. Consequently, if we proceed with transient or stable expression of sgRNAs targeted to gene bodies, our results will most likely not accurately represent the physiological expression status. For this reason, we have decided to employ an *in vitro* approach that circumvents the problem by enabling sgRNA and dCas9 binding after crosslinking and chromatin isolation.

In brief, cells were fixed with FA (1%), chromatin was isolated and was fragmented by sonication. Fragmented chromatin was incubated with SNAP-tagged dCas9-sgRNA RNP complex and pulled down using SNAP tag-specific magnetic beads. In brief, SNAP is a polypeptide based on the mammalian O^6-alkyltransferase. It targets O^6-benzylguanine and acts as a suicide enzyme; this means it binds irreversibly to the substrate through covalent bond (NEB SNAP-tag beads manual). After the pull-down and repeated washes, we eluted by incubating with a moderate salt concentration and Proteinase K (refer on 'Material and Methods'). We utilized qPCR to assess the enrichment of our target chromatin locus against a non-target chromatin sequence on another chromosome and normalized our data against input chromatin. Results from the first attempt indicate that both the target and non-target regions were pulled down at similar levels (**Figure 2A**). We conclude that we did not achieve our goal to enrich our chromatin for sgRNA targeted sequences. However, we noticed that melting curves obtained during qPCR for the downstream viewpoint (HIF1A DG) were not reliable because the amplification isn't specific and produced multiple products (**Figure 2C-D**; *first gel and second melting curve plot respectively*). We used the ΔΔCq method (C_q difference between the pull-down sample and input at target loci against control loci) to normalize for the possible disparity in the amount of DNA used in the reaction.

To troubleshoot this failed attempt, we tried to repeat the pull-down procedure using fragmented, reverse-crosslinked and purified input DNA. Following the same procedure but being able to use recommended binding buffer we were surprised not to be able to get better results and could not isolate efficiently the target loci (HF1A exon 2 UG or DG) vs control locus (non-specific target). (**Figure 2B**).

We therefore questioned the efficiency of sgRNA loading onto dCas9. To investigate this, we repeated the entire procedure from the start. One sample contained both dCas9 and sgRNA, while the second sample contained only sgRNA, serving as negative control. Both samples were incubated overnight with SNAP-tag binding beads, without chromatin or gDNA. We followed similar washing and elution protocol and then we analyzed the eluted samples by agarose gel electrophoresis. If dCas9 and sgRNA were successfully complexed, we expect to identify sgRNAs in the gel at the appropriate molecular weight/ size (120bp). However, neither the negative control sample nor the one with dCas9 showed RNA signal on the gel. We are not sure if simply the sgRNA-dCas9 sample didn't form or if most likely the sgRNA we have used could be degraded, or if the sgRNA was not prepared properly by omitting a melting/cooling step after IVT that will ensure that it folds properly.

Figure 2: Pull-down qPCR data for HIF1A gene. Viewpoints lay upstream (UG) and downstream (DG) of exon 2. (**A-B**) Log_2 FE of chromatin pull-down (**A**) and gDNA pull-down (**B**) vs a non-specific target. (**C**) Agarose gel electrophoresis of qPCR products. (**D**) Melting curves of each primer pair. (**E**) Linear regression analysis of sequentially diluted gDNA.

Co-transcriptional splicing

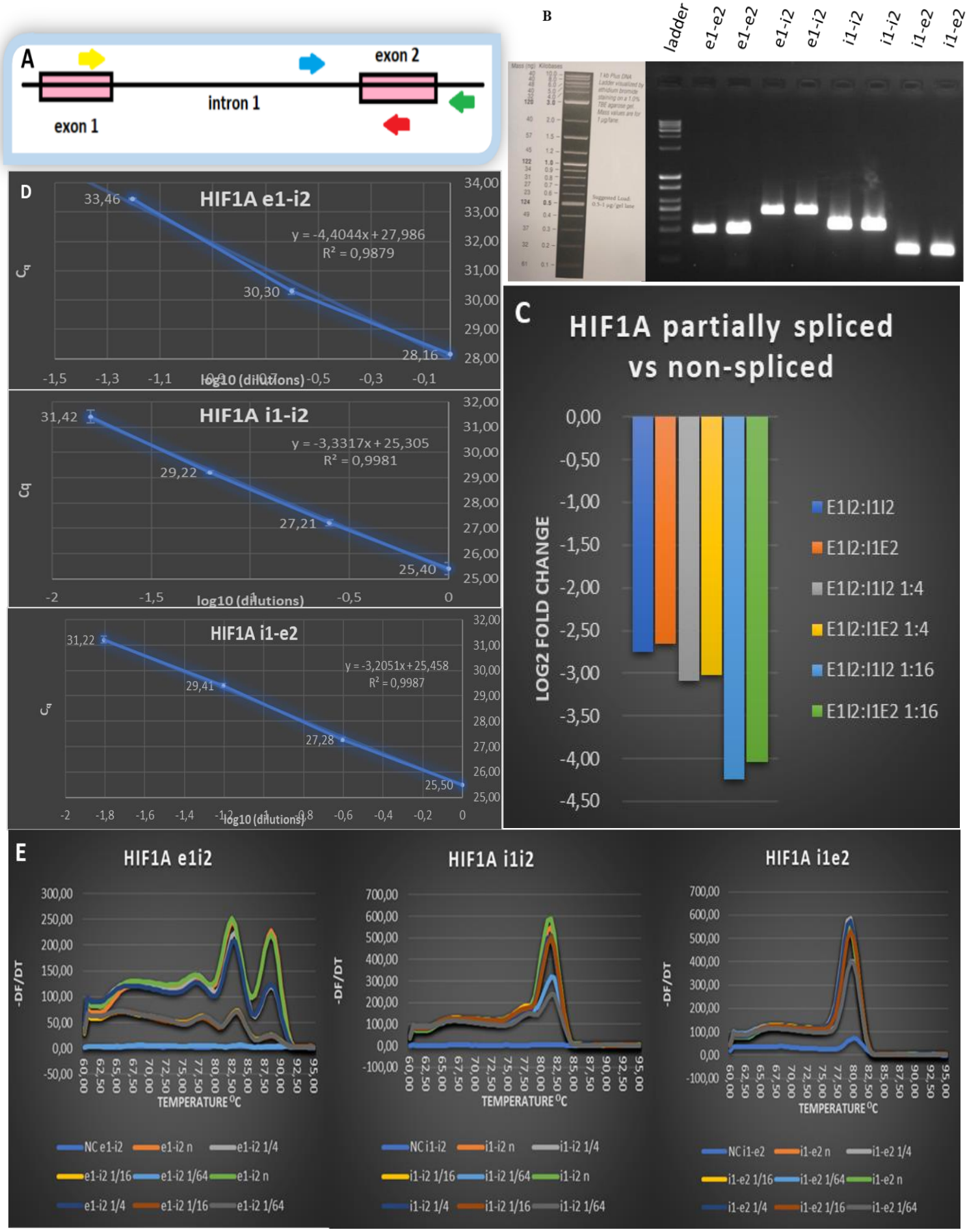
Regardless of the future success for the experiment above, for our analysis to be valid, it is necessary to know *a priori* that the upstream intron sequence of our target genes are spliced during the elongation of Pol 2 across the junction (co-transcriptional splicing). We know from a previous study that co-transcriptional splicing of introns often takes place when RNAPII is located approximately 300bp downstream of the start of the previous intron[36]. We attempted to test if our genes of interest are spliced in the downstream viewpoint on total RNAs that include nRNAs. For this reason, we designed and applied qPCR reactions with primers that span either partially spliced or non-spliced products depending on the intron content (**Figure 3A and M**). Our aim was to quantify the abundance of these two intermediates (*see Material and Methods "qPCR"*).

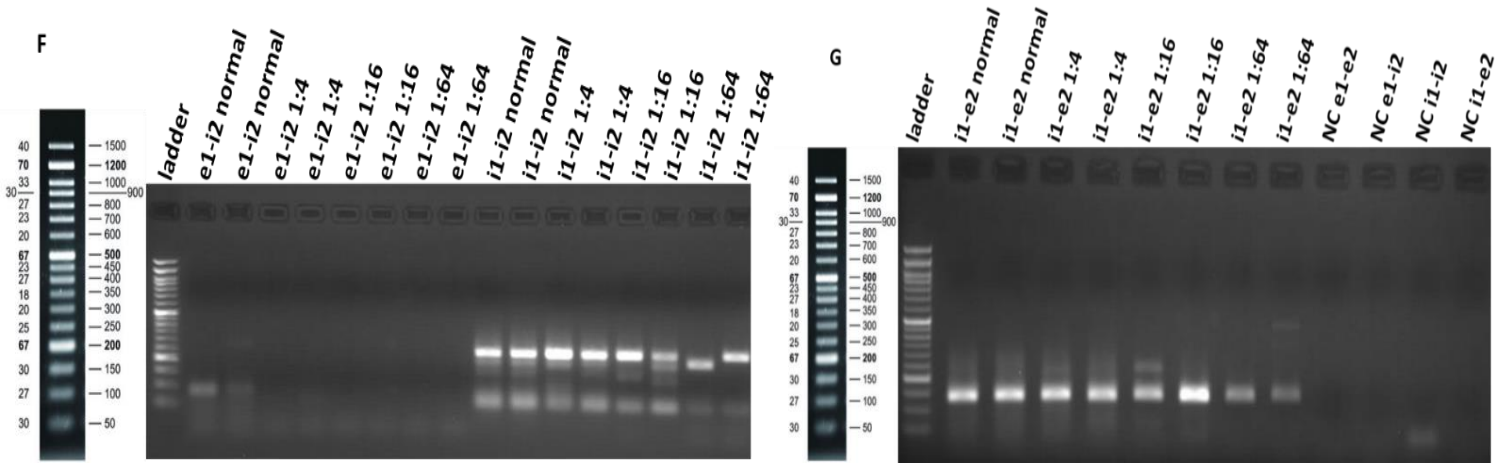
We chose to test HIF1A (**Figure 3A-E**), OSMR (**Figure 3F-J**) and KMT5B (**Figure 3K-P**) genes. In all three genes, our analysis reveals lower signal for partially spliced transcript vs non spliced ones. This indicates that co-transcriptional splicing for the above genes doesn't take place a few bp downstream of the exon's start (e.g., for HIF1A Log_2 FC is almost 3, that means partially spliced transcript is 8-fold less than non-spliced one).

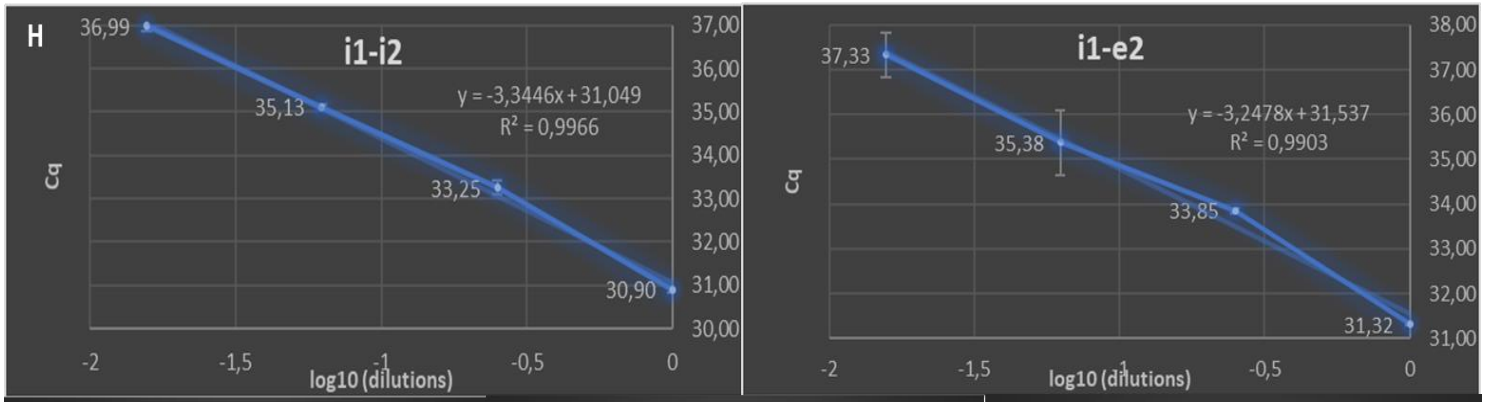
As it is well-known that $\Delta\Delta C_q$ method is prone to miscalculation when the PCR efficiencies are not equal or very similar[37], we calculated the efficiency of each primer pair. The efficiency values of each primer pairs displayed significant variation (**table 5**). This could explain why when we used diluted samples for efficiency calculation, we found variable FCs (**Figure 3C, 3J** and **3N**). Additionally, for OSMR gene our PCR failed to discriminate between two isoforms (**Figure 3K**; *left panel*) and primer pair for partially spliced transcript didn't work so we could not accurately estimate the FC value for that one. In conclusion, this analysis needs to be refined to better infer co-transcriptional splicing rules and status in these genes.

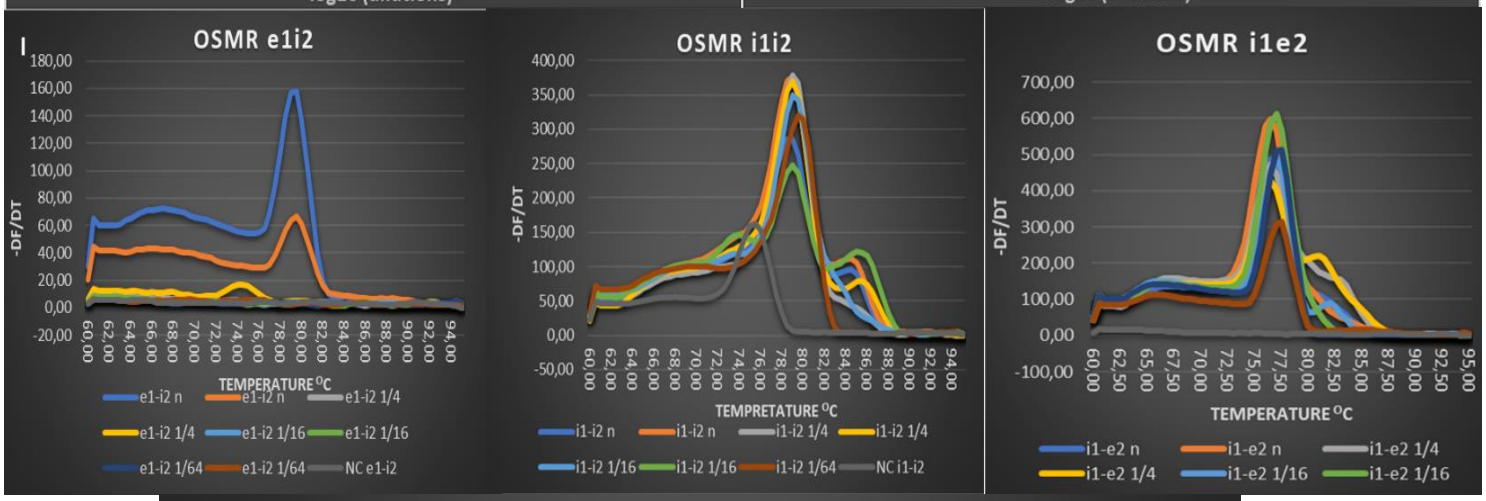
TABLE 5

Primer pair	HIF1A E1I2	HIF1A I1I2	HIF1A I1E2	OSMR E1I2	OSMR I1I2	OSMR I1E2	KMT5B E1E2I2	KMT5B I1I2	KMT5B I1E2
Efficiency (%)	68.68	99.92	105.12	-	99.08	103.22	132.43	100.66	78.49


A
exon 1
intron 1
exon 2
B
ladder
e1-e2
e1-e2
e1-i2
e1-i2
i1-i2
i1-i2
i1-e2
i1-e2
D
HIF1A e1-i2
y = -4,4044x + 27,986
R² = 0,9879
33,46
30,30
28,16
Cq
log10 (dilutions)
HIF1A i1-i2
y = -3,3317x + 25,305
R² = 0,9981
31,42
29,22
27,21
25,40
HIF1A i1-e2
y = -3,2051x + 25,458
R² = 0,9987
31,22
29,41
27,28
25,50
C
HIF1A partially spliced vs non-spliced
LOG2 FOLD CHANGE
E1I2:I1I2
E1I2:I1E2
E1I2:I1I2 1:4
E1I2:I1E2 1:4
E1I2:I1I2 1:16
E1I2:I1E2 1:16
E
HIF1A e1i2
HIF1A i1i2
HIF1A i1e2
-DF/DT
TEMPERATURE °C
NC e1-i2
e1-i2 n
e1-i2 1/4
e1-i2 1/16
e1-i2 1/64
e1-i2 n
e1-i2 1/4
e1-i2 1/16
e1-i2 1/64
NC i1-i2
i1-i2 n
i1-i2 1/4
i1-i2 1/16
i1-i2 1/64
i1-i2 n
i1-i2 1/4
i1-i2 1/16
i1-i2 1/64
NC i1-e2
i1-e2 n
i1-e2 1/4
i1-e2 1/16
i1-e2 1/64
i1-e2 n
i1-e2 1/4
i1-e2 1/16
i1-e2 1/64


F
ladder
e1-i2 normal
e1-i2 normal
e1-i2 1:4
e1-i2 1:4
e1-i2 1:16
e1-i2 1:16
e1-i2 1:64
e1-i2 1:64
i1-i2 normal
i1-i2 normal
i1-i2 1:4
i1-i2 1:4
i1-i2 1:16
i1-i2 1:16
i1-i2 1:64
i1-i2 1:64
G
ladder
i1-e2 normal
i1-e2 normal
i1-e2 1:4
i1-e2 1:4
i1-e2 1:16
i1-e2 1:16
i1-e2 1:64
i1-e2 1:64
NC e1-e2
NC e1-i2
NC i1-i2
NC i1-e2


H
i1-i2
y = -3,3446x + 31,049
R² = 0,9966
36,99
35,13
33,25
30,90
Cq
log10 (dilutions)
i1-e2
y = -3,2478x + 31,537
R² = 0,9903
37,33
35,38
33,85
31,32
Cq
log10 (dilutions)


I
OSMR e1i2
-DF/DT
TEMPERATURE °C
e1-i2 n
e1-i2 n
e1-i2 1/4
e1-i2 1/4
e1-i2 1/16
e1-i2 1/16
e1-i2 1/64
e1-i2 1/64
NC e1-i2
OSMR i1i2
-DF/DT
TEMPRETATURE °C
i1-i2 n
i1-i2 n
i1-i2 1/4
i1-i2 1/4
i1-i2 1/16
i1-i2 1/16
i1-i2 1/64
NC i1-i2
OSMR i1e2
-DF/DT
TEMPERATURE °C
i1-e2 n
i1-e2 n
i1-e2 1/4
i1-e2 1/4
i1-e2 1/16
i1-e2 1/16


J
OSMR
1
e1i2:i1e2
e1i2:i1i2
-8,20
-8,00
-7,80
-7,60
-7,40
LOG2 FC

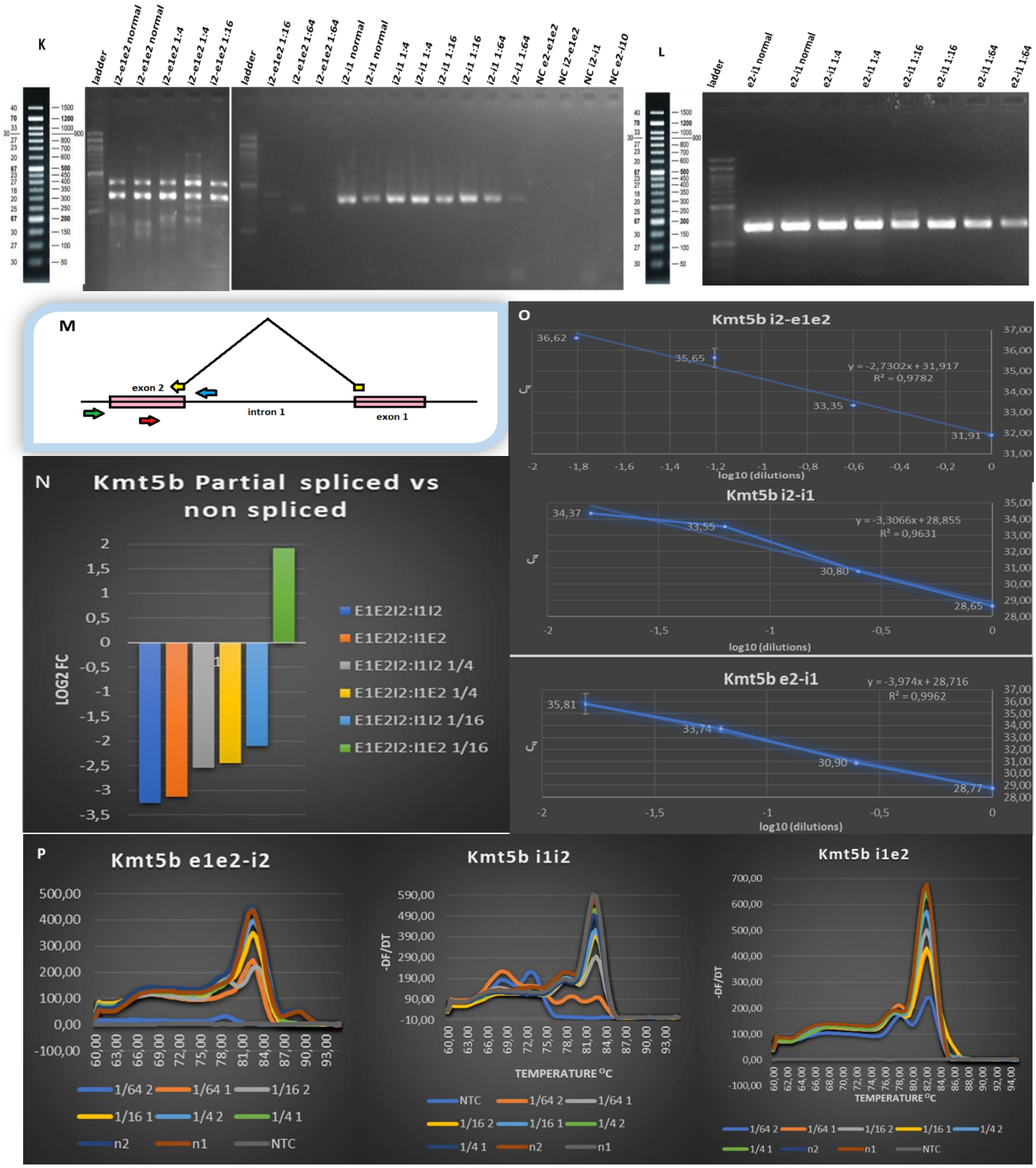


***Figure 3*:** Co-transcriptional splicing qPCR data for HIF1A (**A-E**), OSMR (**A, F-J**) and KMT5B genes (**L-P**). (**A and M**): Primer binding on target gene sequence for HIF1A, OSMR (**A**) and KMT2B (**M**). (**B, F, G, K, L**) Agarose gel electrophoresis of qPCR products to test the specificity of the assay. Note that in OSMR (**K**) we there are double bands that correspond to 2 OSMR isoforms. (**C, J and N**) Bar charts that plot Log_2 FC of partially spliced (e1i2 and e1e2i2) vs non spliced (i1i2 and i1e2) transcripts. (**D, H and O**). Scatter plots that were used to calculate the efficiency of qPCR from sequentially diluted samples. Equation R^2 from linear regression analysis are presented on the graph. (**E, I and P**) Melting curve analysis plots for each qPCR reaction. e1: exon1, e2: exon2, i1: intron1, i2: intron2, e1e2: the primer that spans both the end of exon1 and the start of the exon 2 of KMT5B gene. NC or NTC stands for negative control and non-template control respectively. For our qPCR are identical.

GRID-seq

Global RNA Interaction with DNA by sequencing (GRID-seq) assay is an 'all to all' genomic approach to determine RNA-chromatin interaction pattern[30]. We want to employ this assay to address the issue of the nRNA folding and conformation. We aim to map the interaction of nRNA with adjacent chromatin. From this mapping we can deduce about nRNA conformation, as highly condensed and organized RNA molecules will be more restricted near to their transcription site. In contrast, less condensed molecules will be found to be in proximity with more distal regions.

At first, we tried to test the efficacy of dual fixation and compare that we with single fixed samples. For this reason, we fixed cells with both DSG and FA and cells only with FA. Nuclei were extracted by mild SDS protocol (see *Material and Methods*). After that, we attempted to measure nuclei by using a Neubauer plate. Note that dually fixed nuclei were fewer than single fixed. I calculated approximately ~350000 nuclei/ml in double fixed sample and ~2*10^6 in single fixed one. However, I don't consider these measurements reliable, as there were many nuclei aggregates, especially in double fixed sample. After that, nuclei were centrifuged no pellet was apparent in double fixed samples. Both samples were sonicated and precipitated. Half of the volume were de-crosslinked. Crosslinked and de-crosslinked samples were analyzed by electrophoresis (**Figure 4**). As we expected, crosslinked samples were stuck on the upper part of the gel. However, the signal from the double fixed samples was very low. It was expected, as nuclei number was very low.

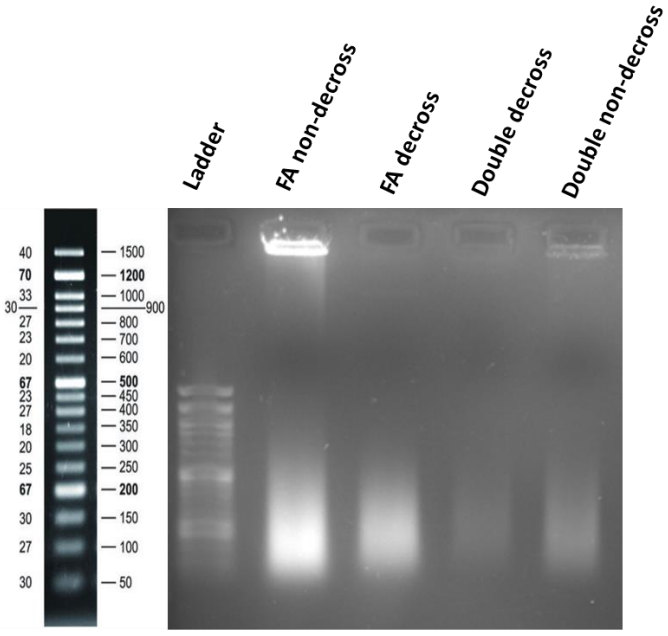


Figure 4: decross: de-crosslinked, Double: double fixed, FA: Fixed with formaldehyde only

Discussion

CRISPR pull-down system

In vivo CRISPR pull-down assay is preferred and has been applied more frequently than *in vitro* one. However, many concerns have been raised about its biological interpretation. After transfection of plasmids expressing dcas9 and sgRNAs or of recombinant RNP complexes, they need reach the chromatin and bind to its target sites. however, it is difficult to assess the efficiency of the process and estimate the potential consequences for the nuclear functions, such as chromatin remodeling or transcription. Although the *in vitro* approach is less efficient, it does not interfere with cellular processes and will allow to avoid transfection and chromatin regulation-related disturbances[38,39].

dCas9-sgRNA RNP complex formation is supposed to be a rapid and effective process. Theoretically, one just need to incubate the components in an appropriate buffer to enable the stable interaction. However, to ensure functionality, it is necessary to make sure that their 3´ terminal stem loops are well-formed.[40] For this reason, in most studies, sgRNAs undergo a denaturation and re-annealing step before they are used to assemble RNP complex[40–42], a step that we had omitted in our attempt.

Beyond the issue of RNP complex formation, during the chromatin pull-down, we believe that a pre-clearing step of large molecular weight chromatin could enhance the precipitation efficiency. Pre-clearing can be performed either using uncoated beads[43] or by high speed centrifugation in a high salt concentration[44]. All of these, along with more washing steps could increase the target to noise ratio.

One more issue we need to address is how to transition from the sonication process to the dCas9 pull-down. In classical ChIP experiments, after fragmentation, most studies continue by incubating with their antibody without any buffer exchange, as the conditions (detergent concentration) in the diluted sonication buffer doesn’t significantly affect antibody-antigen interactons[8]. However, in the case of dCas9, we lack confidence that DNA-sgRNA-dCas9 tripartite complex is stable in such conditions. To overcome this obstacle, we aim to either test if a buffer that is more compatible with dCas9 can be used as a sonication buffer or to transition from sonication buffer to dCas9 reaction buffer by applying dialysis.

Regarding our attempts to quantify the chromatin, we have two aims: 1) to confirm that we have enriched our sample with the viewpoints of interest, and 2) to compare nRNA between the two viewpoints. To do the latter, it is necessary to normalize against the underlie DNA[38].

As our goal is to quantify the amount of nRNA associated with chromatin, we need to ensure that our CRISPR pull-down system has the capability for quantification. Therefore, we set up an *in vitro* experiment to test whether dCas9 can pull-down specific artificial DNA targets from a pool of other non-specific artificial DNA fragments. By testing our system at different specific-to-non-specific DNA target ratios, we could infer if our system can reliably and accurately quantify different amounts of DNA targets.

Co-transcriptional splicing

Co-transcriptional splicing is well-known RNA processing event. It takes place during elongation phase of transcription and is closely associated with it[10]. Furthermore, special focus on whether and how it is coupled with the rest of transcription-associated processes is under intense study[23,45]. In a recent study, Reimer KA, et al presented data from nRNA long read sequencing that, revealing that in more than 2/3 of genes, splicing occurs immediately after the transcription of the 3´-splice site, when RNAPII is located approximately 110-300bp downstream of that site[36]. In contrast, another study suggests that in an inducible gene transcription system, transcription initiation is simultaneous, but splicing occurs at different time points[46].

To determine and verify that our genes of interest undergo co- transcriptional splicing, we utilized qPCR analysis. We used primer pairs that span either partially spliced transcripts or non-spliced ones in an attempt to deduce which one is more abundant. The presence of more partially spliced transcripts indicates higher co-transcriptional splicing and vice versa. We analyzed the qPCR data using $2^{\Delta\Delta C_q}$ method and plotted as $\log_2$ FC. In spite of simplicity and versatility of this analysis method, it is prone to errors since it assumes that the compared samples have primer efficiencies equal to 2 implying that every amplification cycle will output double amount of DNA [37,47]. However, in most cases this assumption does not prove accurate, as reaction efficiencies differ among various primer pairs. Our data showed significant variability and deviation from 2 when we evaluated the efficiencies of primer pairs that target partially spliced transcripts. Therefore, we conclude that our $\Delta\Delta C_q$ analysis results are not valid.

To determine whether co-transcriptional splicing occurs in our genes, we could modify our approach by targeting other sites on the transcripts of interest. Additionally, this time we should choose

shorter amplicons. According to the Hao S, et al[46] it is feasible to target and amplify pre-mRNA transcripts using a single primer pair that spans the 5′-splicing site of the upstream intron. For detecting partially spliced transcript, we could use primer pairs, one of which would span exon-exon junction of the transcript.

Importantly, we note that long introns may undergo recursive splicing. In such cases, previous optimization could potentially result in false negative results since the absence of the transcripts containing the 5′-splice site may be erroneously considered as low partially spliced transcript levels. Despite we didn't identified any such a recursive splicing site in our examination of Kelly, et al dataset, we need to be careful[48], as these data were originated by different cell type.

Grid-seq

Grid-seq is one of the primary methods that study RNA-chromatin interaction in a "all to all" manner. It captures the physical proximity of RNA molecules with their adjacent chromatin regions, using an artificial probe that functions as "bridge" and link these two different molecules in one[30,49]. After Grid-seq, other similar methods have been published and all of them are based on the same principles. Most of them try to modify and optimize the protocol, to overcome the Grid-seq approach limitations[31,50–52]. The most prominent limitations of Grid-seq is 1) its bias for nRNAs and 2) biased chromatin fragmentation due to AluI bias for certain restriction sites. In RADICL-seq, that was published 3 years later, they added a step of RNase H treatment to reduce bias for nRNAs, as RNase H binds on RNA: DNA dimers and digest the RNA strand. In addition, to overcome the bias AluI they used DNase I, a nuclease that digests the genome more homogeneously[31] than the restriction enzymes. I must point out here that, for our aim, bias towards nRNAs is one of the reasons that we chose to use Grid seq.

Our aim is to utilize Grid-seq to determine the spatial organization of nRNA molecules. However, this protocol per se, can only map the 3′-ends of RNAs on the proximal chromatin. Therefore, we can't get spatial information for the rest RNA molecule. To overcome this limitation, it is necessary to add an RNA partial fragmentation step. In such a way, more 3′ free ends from the same nRNA molecule could be mapped against their underlying chromatin. For partial RNA fragmentation we decided to use a chemical treatment with NaOH[53] as we wish avoid the bias that accompanies every enzymatic treatment.

In our experiments we tried the efficiency of double fixation, and it is apparent that this method is functional. However, we met the obstacle of nuclei isolation. Double fixation assay seems to yield much smaller number of nuclei than single FA one. Nuclei from both conditions were isolated based

on the manner that is recommended in the protocol[49]. It is based on SDS treatment at high temperature for sort time (see "Material and Methods"). In addition, many nuclei aggregates were noticed after inspection under microscope (data not shown). This indicates that double fixation could be very harsh, so we either need to change this approach or to modify the fixation protocol. In addition, we consider that the above protocol may not be the suitable to isolate nuclei from our cells, so we would like to experiment with other protocols that use non-ionic detergents.

References

1. Alonso, D. & Mondragón, A. Mechanisms of catalytic RNA molecules. *Biochem. Soc. Trans.* **49**, 1529–1535 (2021).
2. Cramer, P. Organization and regulation of gene transcription. *Nature* **573**, 45–54 (2019).
3. Gibbons, M. D. *et al.* Enhancer-Mediated Formation of Nuclear Transcription Initiation Domains. *Int. J. Mol. Sci.* **23**, 9290 (2022).
4. Ntini, E. *et al.* Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20**, 923–928 (2013).
5. Rippe, K. & Papantonis, A. RNA polymerase II transcription compartments: from multivalent chromatin binding to liquid droplet formation? *Nat. Rev. Mol. Cell Biol.* **22**, 645–646 (2021).
6. Noe Gonzalez, M., Blears, D. & Svejstrup, J. Q. Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nat. Rev. Mol. Cell Biol.* **22**, 3–21 (2021).
7. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* **33**, 960–982 (2019).
8. Lavigne, M. D., Konstantopoulos, D., Ntakou-Zamplara, K. Z., Liakos, A. & Fousteri, M. Global unleashing of transcription elongation waves in response to genotoxic stress restricts somatic mutation rate. *Nat. Commun.* **8**, 2076 (2017).
9. Kujirai, T. & Kurumizaka, H. Transcription through the nucleosome. *Curr. Opin. Struct. Biol.* **61**, 42–49 (2020).
10. Smolle, M. & Workman, J. L. Transcription-associated histone modifications & cryptic transcription. *Biochim. Biophys. Acta* **1829**, 84–97 (2013).

11. James, K., Gamba, P., Cockell, S. J. & Zenkin, N. Misincorporation by RNA polymerase is a major source of transcription pausing in vivo. *Nucleic Acids Res.* **45**, 1105–1113 (2017).

12. Eaton, J. D. & West, S. Termination of Transcription by RNA Polymerase II: BOOM! *Trends Genet.* **36**, 664–675 (2020).

13. Mendoza-Figueroa, M. S., Tatomer, D. C. & Wilusz, J. E. The Integrator Complex in Transcription and Development. *Trends Biochem. Sci.* **45**, 923–934 (2020).

14. Muniz, L., Nicolas, E. & Trouche, D. RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. *EMBO J.* **40**, e105740 (2021).

15. Kirstein, N., Gomes Dos Santos, H., Blumenthal, E. & Shiekhattar, R. The Integrator complex at the crossroad of coding and noncoding RNA. *Curr. Opin. Cell Biol.* **70**, 37–43 (2021).

16. Tropp, B. E. *Principles of Molecular Biology*. (Jones & Bartlett Publishers, 2012).

17. Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).

18. Bergsma, A. J., van der Wal, E., Broeders, M., van der Ploeg, A. T. & Pim Pijnappel, W. W. M. Chapter Three - Alternative Splicing in Genetic Diseases: Improved Diagnosis and Novel Treatment Options. in *International Review of Cell and Molecular Biology* (ed. Loos, F.) vol. 335 85–141 (Academic Press, 2018).

19. Shenasa, H. & Bentley, D. L. Pre-mRNA splicing and its cotranscriptional connections. *Trends Genet.* **39**, 672–685 (2023).

20. Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in Alternative Pre-mRNA Splicing. *Cell* **144**, 16–26 (2011).

21. Xu, S.-M., Curry-Hyde, A., Sytnyk, V. & Janitz, M. RNA polyadenylation patterns in the human transcriptome. *Gene* **816**, 146133 (2022).

22. Ren, F., Zhang, N., Zhang, L., Miller, E. & Pu, J. J. Alternative Polyadenylation: a new frontier in post transcriptional regulation. *Biomark. Res.* **8**, 67 (2020).

23. Neugebauer, K. M. Nascent RNA and the Coordination of Splicing with Transcription. *Cold Spring Harb. Perspect. Biol.* **11**, a032227 (2019).

24. Skalska, L., Beltran-Nebot, M., Ule, J. & Jenner, R. G. Regulatory feedback from nascent RNA to chromatin and transcription. *Nat. Rev. Mol. Cell Biol.* **18**, 331–337 (2017).

25. G Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B. & Rinn, J. L. Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* **17**, 28 (2016).

26. Saldaña-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* **28**, 723–734 (2014).

27. Kung, J. T. *et al.* Locus-specific targeting to the X-chromosome revealed by the RNA interactome of CTCF. *Mol. Cell* **57**, 361–375 (2015).

28. Kato, M. & Carninci, P. Genome-Wide Technologies to Study RNA–Chromatin Interactions. *Non-Coding RNA* **6**, 20 (2020).

29. Sridhar, B. *et al.* Systematic mapping of RNA-chromatin interactions in vivo. *Curr. Biol. CB* **27**, 602–609 (2017).

30. Li, X. *et al.* GRID-seq reveals the global RNA–chromatin interactome. *Nat. Biotechnol.* **35**, 940–950 (2017).

31. Bonetti, A. *et al.* RADICL-seq identifies general and cell type–specific principles of genome-wide RNA-chromatin interactions. *Nat. Commun.* **11**, 1018 (2020).

32. Global profiling of RNA–chromatin interactions reveals co-regulatory gene expression networks in Arabidopsis | Nature Plants. https://www.nature.com/articles/s41477-021-01004-x.

33. Schärfen, L. & Neugebauer, K. M. Transcription Regulation Through Nascent RNA Folding. *J. Mol. Biol.* **433**, 166975 (2021).

34. Saldi, T., Riemondy, K., Erickson, B. & Bentley, D. Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. *Mol. Cell* **81**, 1789-1801.e5 (2021).

35. Leidescher, S. *et al.* SPATIAL ORGANIZATION OF TRANSCRIBED EUKARYOTIC GENES.

36. Reimer, K. A., Mimoso, C., Adelman, K. & Neugebauer, K. M. Co-transcriptional splicing regulates 3′ end cleavage during mammalian erythropoiesis. *Mol. Cell* **81**, 998-1012.e7 (2021).

37. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).

38. Fujita, H., Fujita, T. & Fujii, H. Locus-Specific Genomic DNA Purification Using the CRISPR System: Methods and Applications. *CRISPR J.* **4**, 290–300 (2021).

39. Gauchier, M., Van Mierlo, G., Vermeulen, M. & Déjardin, J. Purification and enrichment of specific chromatin loci. *Nat. Methods* **17**, 380–389 (2020).

40. Mekler, V., Minakhin, L., Semenova, E., Kuznedelov, K. & Severinov, K. Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3′-terminal segment of guide RNA. *Nucleic Acids Res.* **44**, 2837–2845 (2016).

41. Slesarev, A. *et al.* CRISPR/Cas9 targeted CAPTURE of mammalian genomic regions for characterization by NGS. *Sci. Rep.* **9**, 3587 (2019).

42. David, S. R., Maheshwaram, S. K., Shet, D., Lakshminarayana, M. B. & Soni, G. V. Temperature dependent in vitro binding and release of target DNA by Cas9 enzyme. *Sci. Rep.* **12**, 15243 (2022).

43. Nagalingam, K. *et al.* Chromatin immunoprecipitation (ChIP) method for non-model fruit flies (Diptera: Tephritidae) and evidence of histone modifications. *PLOS ONE* **13**, e0194420 (2018).

44. Tsui, C. *et al.* dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. *Proc. Natl. Acad. Sci.* **115**, (2018).

45. Custódio, N. & Carmo-Fonseca, M. Co-transcriptional splicing and the CTD code. *Crit. Rev. Biochem. Mol. Biol.* **51**, 395–411 (2016).

46. Hao, S. & Baltimore, D. RNA splicing regulates the temporal order of TNF-induced gene expression. *Proc. Natl. Acad. Sci.* **110**, 11934–11939 (2013).

47. Liu, W. & Saint, D. A. A New Quantitative Method of Real Time Reverse Transcription Polymerase Chain Reaction Assay Based on Simulation of Polymerase Chain Reaction Kinetics. *Anal. Biochem.* **302**, 52–59 (2002).

48. Kelly, S. *et al.* Splicing of many human genes involves sites embedded within introns. *Nucleic Acids Res.* **43**, 4721–4732 (2015).

49. Zhou, B. *et al.* GRID-seq for comprehensive analysis of global RNA-chromatin interactions. *Nat. Protoc.* **14**, 2036–2068 (2019).

50. Bell, J. C. *et al.* Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *eLife* **7**, e27024 (2018).

51. Gavrilov, A. A. *et al.* Studying RNA–DNA interactome by Red-C identifies noncoding RNAs associated with various chromatin types and reveals transcription dynamics. *Nucleic Acids Res.* **48**, 6699–6714 (2020).

52. Wu, W. *et al.* Mapping RNA-chromatin interactions by sequencing with iMARGI. *Nat. Protoc.* **14**, 3243–3272 (2019).

53. Ntini, E., Budach, S., Ørom, U. A. V. & Marsico, A. Predictive modeling of long non-coding RNA chromatin (dis-)association. 2020.12.15.422063 Preprint at https://doi.org/10.1101/2020.12.15.422063 (2020).