

A high-throughput approach to predict A-to-I effects on RNA structure indicates a change of double-stranded content in non-coding RNAs

Riccardo Delli Ponti¹, Laura Broglia², Andrea Vandelli^{3,4}, Alexandros Armaos², Marc Torrent Burgas³, Natalia Sanchez de Groot^{3,*} and Gian Gaetano Tartaglia^{2,5,*}

1. Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), 30 Biopolis Street, Matrix #07-01, 138671 Singapore
2. Center for Human Technologies, Istituto Italiano di Tecnologia, Via Enrico Melen 83, 16152 Genoa, Italy.
3. Department of Biochemistry and Molecular Biology, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain
4. Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain
5. Department of Biology 'Charles Darwin', Sapienza University of Rome, P.le A. Moro 5, Rome 00185, Italy.

ABSTRACT

RNA molecules undergo a number of chemical modifications whose effects can alter their structure and molecular interactions. Previous studies have shown that RNA editing can impact the formation of ribonucleoprotein complexes and influence the assembly of membrane-less organelles such as stress-granules. For instance, N6-methyladenosine (m6A) enhances SG formation and N1-methyladenosine (m1A) prevents their transition to solid-like aggregates. Yet, very little is known about adenosine to inosine (A-to-I) modification that is very abundant in human cells and not only impacts mRNAs but also non-coding RNAs. Here, we built the CROSSalve predictor of A-to-I effects on RNA structure based on high-throughput in-cell experiments. Our method shows an accuracy of 90% in predicting the single and double-stranded content of transcripts and identifies a general enrichment of double-stranded regions caused by A-to-I in long intergenic non-coding RNAs (lincRNAs). For the individual cases of NEAT1, NORAD and XIST, we investigated the relationship between A-to-I editing and interactions with RNA-binding proteins using available CLIP data. We found that A-to-I editing is linked to alteration of interaction sites with proteins involved in phase-separation, which suggests that RNP assembly can be influenced by A-to-I. CROSSalve is available at http://service.tartagliolab.com/new_submission/crossalve.

INTRODUCTION

Although 80% of the human genome is transcribed into RNA, only 2% is translated into proteins (mRNAs), and the rest constitutes the abundant and diverse set of non-coding RNAs (ncRNAs) (1). These RNAs have many important roles such as the regulation of translation, RNA splicing and DNA replication, but also in chromosome structure and immunity (2). This functional diversity is reflected in the wide range of lengths and conformations that ncRNAs can adopt. Accordingly to their size, they can be classified into small ncRNAs (< 200 nt) and long ncRNAs (lncRNA, >200nt). Moreover, this diversity can be furtherly expanded through the addition of post-transcriptional modifications to the nucleotide chain.

As structure confers strength and specificity to molecular interactions, RNA regulation tightly depends on structural properties and their effects on molecular partners recruitment (3). Thus, not only structure determines stability and spatial configuration of RNA molecules but the molecular partners define RNA localization and function. In the case of lncRNAs, their length provides room to incorporate multiple interaction sites providing high valency. As a result, one molecule of RNA can bind several proteins (3, 4). This property is crucial for the formation of ribonucleoprotein (RNP) condensates such as stress granules (SGs) and P-bodies (PBs) where lncRNAs can attract proteins and RNAs (5). Accumulation of proteins and RNAs is a crucial event driving the formation of SGs and PBs via phase separation (4, 6, 7). In biomolecular condensates, RNAs form dynamic interactions with proteins and other RNAs, favoring the exchange of energy and components with the surrounding environment (6). Although all RNAs can be in principle recruited within SGs and PBs (8, 9), some lncRNAs play a special role in this context, such as NORAD, which controls the genomic stability in mammalian cells (10). In order to do so, NORAD can sequester and inhibit Pumilio (PUM) proteins by promoting the formation of phase-separated PUM condensates called NP bodies. Disruption of these condensates leads to genomic instability due to PUM hyperactivity (11, 12). Similarly, the architectural lncRNA NEAT1, has modular domains specifically arranged to bind NONO proteins and form paraspeckle compartments (13). Another interesting example is XIST that is responsible for the X-chromosome inactivation. This is achieved by triggering the formation of large assemblies containing components of Polycomb repressive complexes (PRC1 and PRC2) and many other proteins like CELF, PTBP1 and SPEN (7).

Initially, computational approaches were the only affordable strategies, both technically and timely, to assess RNA structure. However, the lack of experimental data for their training limited their precision (11, 12) until new experimental strategies became available. These approaches can be classified as nuclease cleavage and small molecule-based probing. The former employs nucleases to cut single or double-stranded RNAs (RNases P1, S1 and V1), and the cutting pattern shows the location, at a single nucleotide resolution, of the different strand types (16). The latter employs small chemicals (1M7, DMS, N3-kethoxal and NAI-N3) that bind single-stranded RNA bases and interfere in the reverse transcription step. Then, the pattern of transcription errors (prompt stop or mutations) is used to calculate a structural score for each base (17, 18). Overall, these techniques have allowed achieving

comprehensive RNA secondary structure maps of human coding and non-coding transcriptome (19).

These experimental data have fed a second generation of computational tools to predict RNA structure with high precision. In this context, we developed the CROSS algorithm, which is based on four different genome-wide studies that use PARS (RNases S1 and V1) and SHAPE (NAI-N3), together with 3D structural data from X-ray crystallography and NMR (20). CROSS is able to predict the secondary structure propensity of an RNA sequence at single-nucleotide resolution without sequence length restrictions. As in the cases of XIST (21) and SARS-CoV-2 RNA genome (22), CROSS can be used to identify the structural state of sites where RNA-binding proteins (RBPs) interact.

Post-transcriptional chemical modifications can alter RNA structure and properties, and thus modulate its function and contact network, without changing the template genomic DNA (23). N6-methyladenosine (m6A) is one of the most abundant RNA modifications, present in different RNA species such as mRNA, rRNA, tRNA, snRNA and lncRNA (24). However, it is especially abundant in mRNAs where it can be present in many different positions and can promote their recruitment into SGs (25–27). An *in vivo* screening with NAI-N3 (icSHAPE) showed that m6A promotes the transition from double- to single-stranded regions (18). Based on these data, we developed CROSSalive, a method for the prediction of RNA secondary structure *in vivo* with and without m6A modifications (28).

In lncRNAs, the most abundant editing is adenosine to inosine (A-to-I), which renders adenosine similar to guanosine in terms of chemical properties (**Figure 1A-B**) (29). The A-to-I modification is thought to rewire the RNA contact network, both intramolecularly, influencing the secondary structure, and intermolecularly, affecting binding partners and their function (30). Also occurring in mRNAs, the A-to-I modification changes how a codon is read by the ribosome and translated into protein. In both the RNA classes, the ADAR family of enzymes catalyze the hydrolytic deamination reaction that converts adenosine to inosine, where the amino group at position 6 of the purine ring changes to a carbonyl group (**Figure 1A**). In humans, there are three ADAR enzymes (ADAR1, ADAR2 and ADAR3) that catalyze the A-to-I conversion (31). Despite their abundance and effects, A-to-I conversion in lncRNAs has been scarcely studied at the experimental level. A recent computational screening shows that this modification could appear in about 200,000 positions in humans, and 65% of them are located within sites that can affect secondary structure (32, 33). This structural relationship was previously investigated by Solomon and co-workers who used PARS-seq to study the effect of ADAR1 silencing (34). Upon ADAR1 silencing, the authors observed a general decrease in double-stranded content with respect to the single-stranded one. Additionally, Corbet and co-workers reported that ADAR1 depletion triggers SG formation (35). Changes in RNA secondary structure due to A-to-I editing are relevant during viral infection and help the innate immunity to distinguish between viral and human double-stranded RNA (dsRNA) patterns (36).

Here we used Solomon's PARS-seq data (34) to introduce a new development of CROSSalive that predicts the effects of A-to-I modifications on RNA structure. Specifically,

our method calculates the structural profile of a given RNA sequence (single- or double-stranded state) at single-nucleotide resolution in presence (ADAR+) and absence (ADAR-) of ADAR. We employed this version of CROSSalive, available at http://service.tartagialab.com/new_submission/crossalive, to study the effect of A-to-I modifications on lncRNAs (**Figure 1C**). Using eCLIP data (37) and catRAPID predictions of protein-RNA interactions (38–40) we investigated the occurrence of A-to-I editing in correspondence of RBPs interactions with three long intergenic non-coding RNAs (lncRNAs), NEAT1, NORAD and XIST. As NORAD, NEAT1 and XIST drive the formation of membrane-less organelles, and our predictions indicate that A-to-I editing alters interaction with proteins prone to phase separation, we hypothesize that A-to-I can have an impact on their macromolecular assembly (**Figure 1C**).

EXPERIMENTAL PROCEDURES

PARS data

We processed the read counts from the original PARS experiments on ADAR+ and ADAR- (34) to select nucleotides with a high-propensity to be in single- or double-stranded conformation. In PARS experiments, purified poly-A + RNA samples from control HepG2 cells (ADAR+) and ADAR knockdown (ADAR-) HepG2 cells were extracted and *in vitro* probed with S1 nuclease (S1) and RNase V1 (V1) that cleave RNA at single-stranded and double-stranded regions, respectively. We used the read counts of S1 and V1 to build a score for each single-nucleotide resolution using the formula $\text{Score} = \log(V1/S1)$.

Training set

To build the training set, we followed the approach developed for CROSS (20). Briefly, using the ranking provided by the $\log(V1/S1)$ score, we assessed the structural state of each nucleotide (**Figure 2A**) selecting 5 maxima (double-stranded state) and 5 minima (single-stranded state) per transcript. Upon removal of transcripts with insufficient PARS coverage (zeroes for >40% of nucleotides; 3000 sequences passing the filter) and sequence redundancy, we collected a total of 20000 fragments centered around ‘top’ (double-stranded) and ‘bottom’ (single-stranded) nucleotides. Each fragment has a fixed length of 13nt (6 nucleotides on the left and 6 nucleotides on the right side of the central nucleotide). The sequence redundancy of fragments was removed using CD-HIT with threshold at 90% (41). The one-hot encoding (keras in-built function), was employed to convert the fragments into arrays of binary integers for the training of our Artificial Neural Network (ANN).

Sequences

Sequences of RNAs (20000) reported in the original PARS experiments (34), used in the training, were downloaded from UCSC. cDNA sequences of 22’000 mRNA (**Supplementary Figure 1**) and 12’000 long intergenic non-coding RNAs (lincRNAs; **Figure 3**) were downloaded from Ensembl 105 using Biomart. For mRNAs we used only Ensembl canonical isoforms.

ANN architecture

We used Keras and Tensorflow to train the ANN (42, 43). To keep consistency with our previous algorithm, we used a similar architecture to the one employed for CROSS (20). The architecture is a fully connected Neural Network (*Dense* module in keras), with one hidden layer with 52 internal variables. The layers are based on *relu* and *sigmoid* activation functions, and the training was performed with *adam* optimiser and *binary entropy* loss. The ANN was initially trained for ~3000 epochs, but the final model is trained for 1000 epochs to avoid overfitting (**Figure 2B,C**). 10% of the training set (2000 fragments) was randomly selected and used as an independent test set.

Performances

We tested the ANNs at each epoch in the training phase by reporting accuracy and loss (**Figure 2B,C**). We used the keras function *train_test_split* to build an independent test set containing 10% (2000 fragments) for ADAR+/- PARS experiments. The networks reached an accuracy of 0.9 on both the independent sets. We tested performances on NEAT1 (>20'000 nt not present in the training set; **Supplementary Information**;) and reported the Area under the ROC curves (AUCs) for both ADAR+/- models using the 10% top (double-stranded) and bottom (single-stranded) nucleotides ranked by log(V1/S1). Both methods achieved an AUC of 0.85 (**Supplementary Figure 2**). Finally, we employed nextPARS (12) experiments on NORAD (**Supplementary Information**) and measured the performances of both ADAR+/- ANNs on the top/bottom 10% nucleotides ranked on nextPARS score, achieving an AUC of 0.77 (**Supplementary Figure 3**). No significant difference exists between PARS and nextPARS protocols, except that nextPARS has higher throughput and sample multiplexing in comparison to PARS (44).

eCLIP analysis

We used human protein–RNA interactions that were identified through eCLIP experiments in either of the two cell lines, K562 and HepG230 (37). The dataset was downloaded on 18 of September 2020. The dataset contains interactions for 151 unique RBPs: 120 proteins in the K562 cell line and 103 in the HepG2 cell line. We used input-normalized eCLIP data (SM input normalization) with stringent cut-offs [$-\log_{10}(\text{p value}) > 3$ and $-\log_2(\text{fold_enrichment}) > 3$]. The data was processed using bioconductor packages for R (<https://www.bioconductor.org/>) to map genomic locations to Ensembl transcript identifiers.

catRAPID analysis

To compute protein-RNA interactions, we used catRAPID omics (40). The *catRAPID* algorithm (38, 39) estimates the binding potential through van der Waals, hydrogen bonding and secondary structure propensities of both protein and RNA sequences allowing identification of binding partners with high confidence. As reported in a recent analysis of about half a million of experimentally validated interactions, the algorithm is able to separate interacting vs non-interacting pairs with an area under the ROC curve of 0.78 (with False Discovery Rate FDR significantly below 0.25 when the Z-score values are > 2) (45) (**Supplementary Information**).

RESULTS

CROSS (Computational Recognition of Secondary Structure) is a neural network-based machine learning approach trained on experimental data (SHAPE, PARS, NMR/X-Ray, and icSHAPE) able to quickly profile large and complex molecules such as lncRNAs and viral genomes without length restrictions (20). In the original manuscript, CROSS was tested on crystallographic structures (Area Under the ROC curve, AUC 0.72 and positive predictive value PPV 0.74) and on DMS data for murine XIST (AUC 0.75) (20). CROSS was used to profile the HIV genome, reaching an Area Under the ROC curve (AUC) of 0.75 with SHAPE data (20). More recently, CROSS was applied to SARS-CoV-2 (22) and Dengue genomes (46), reaching AUCs of 0.73 and 0.85 with experimental data.

The first version of CROSSalive was trained on icSHAPE data measured in presence and absence of METTL3 (m6A writer) (47). It was also tested on murine XIST (28).

Here, we exploited PARS data on purified poly-A⁺ RNA samples from HepG2a experiments by Solomon *et al.* (34) to train a new version of CROSSalive that predicts the effects of A-to-I on the RNA secondary structure profile (**Figure 1**). Briefly, we processed the read counts from the original PARS experiments, both in presence (ADAR+) and absence (ADAR-) of ADAR, to select nucleotides with a high-propensity to be in single- or double-stranded conformation according to the read counts of RNase V1 (V1) and S1 nuclease (S1) (**Figure 2A**). The information contained in the nucleic acid sequences was converted into arrays of binary integers for the training of a connected neural network, following a similar architecture to previous CROSS algorithms (20) (**Experimental Procedures**). Upon cross-validation, both ADAR+ and ADAR- networks showed excellent performances, with accuracies of 0.9 in the cross-validation test set (**Figure 2B,C**).

We applied both models (ADAR+ and ADAR-) on human long intergenic non-coding RNAs (lincRNAs) that are a class of lncRNAs that do not overlap exons of either protein-coding or other non-lincRNA types of genes (12'000 transcripts; **Figure 3A**; **Experimental Procedures**). By analyzing the structural content of each transcript (i.e. the percentage of double-stranded nucleotides, DS%), we noticed that there is an increase in structural levels of lincRNAs when predicting their secondary structure using the ADAR+ model. mRNAs also showed a similar trend (**Supplementary Figure 1**), which is in agreement with previous studies in which it was shown that A-to-I modifications promote the formation of secondary structure (32, 34).

With the aim of validating and exploring putative applications of this algorithm, we studied three lncRNAs, NORAD, NEAT1 and XIST. Their analysis revealed that in presence of ADAR, XIST and NEAT1 exhibit a stronger double-stranded content than NORAD (**Figure 3B**).

NORAD

Systematic investigation of the RNA content of SGs revealed that some lncRNAs such as NORAD are recruited to these condensates (8, 48, 49). NORAD is localized in both the cytoplasm and the nucleus and serves as a platform for PUMILIO (PUM1/2) proteins (50). Although NORAD is not essential for the formation of SGs (49), PUMILIO dependent recruitment of NORAD impacts on SG size and morphology (51). Acting as a PUMILIO sponge, NORAD maintains chromosome stability hampering PUMILIO function (50, 52) and ultimately preventing aberrant mitosis. This is accomplished by the formation of phase-separated condensates (*i.e.* NP bodies) that allow to sequester a high number of PUMILIO proteins. These condensates are generated through the formation of multivalent interactions between NORAD and PUMILIO and also among PUMILIO proteins themselves (6). NORAD contains 12 repeat units (of approximately 300 nt) where the sequence pattern UGURUAUA is repeated (12, 53). Recently it has been shown that NORAD can be methylated at adenosine sites and this would decrease the number of PUMILIO proteins sequestered (54), indicating that RNA modifications can modulate lncRNA activity.

Using CROSSalike we reproduced the secondary structure content, as probed by nextPARS experiments (12), to a remarkable extent, obtaining an AUC of 0.76 for the ADAR+ model and an AUC of 0.77 for the ADAR- model (**Supplementary Image 2; Experimental Procedures**). As the nextPARS experiments were carried out *in vitro* (8), and CROSSalike computes the secondary structure content *in vivo*, the goodness of these performances indicates that NORAD structure does not change significantly in the cell. In presence of ADAR, the most structured regions are the 5' end of NORAD that interacts with RBMX (52) and nucleotides 4100-4650 that comprise repeats 11 and 12 recognized by PUM1/2 (**Figure 4A**). Notably, NORAD 5' end region represents the strongest RBMX binding site in the transcriptome (55). While RBMX was first recognized as essential for NORAD function in maintaining genomic integrity (55), it has been recently shown that RBMX is dispensable and that PUM1/2 is necessary for this function (52). It is possible that RBMX and PUM1/2—by interacting with two different NORAD regions—play distinct roles in NORAD-mediated control of genome integrity. Of note, m6A modification induces a structural rearrangement responsible for rendering NORAD less structured and accessible to RBMX binding (56). Since NORAD interactions with RBMX are predicted to occur in highly structured regions according to our predictions, we speculate that m6A and A-to-I modifications could play a role in balancing the amount of RBMX associated with NORAD.

Nucleotides 1000-1500 of NORAD (between repeat 2 and 3) are predicted to be undergoing structural rearrangements in presence of ADAR (blue box in **Figure 4B**). Specifically, our calculations indicate that there is a strong decrease in secondary structure content. Considering RBP interactions detected by eCLIP in this region as (37), we identified FUBP3, IGF2BP2, PUM1, PUM2, TIA1 and TIAL1 that are SG components, which suggests that ADAR could impact NORAD ability to phase-separate. At nucleotides 500-800 and 4300-

4650 (repeat 12), we predict that ADAR depletion decreases the secondary structure content. In these regions, the SG component KHDRBS1 binds (57) (blue box in **Figure 4B**). In the 5' region corresponding to nucleotides 1-500, interactions with SG proteins DDX3X, DDX6, EIF3G, PCBP1 and RPS3 could be potentially affected by changes in the RNA structure (blue box in **Figure 4B**).

NEAT1

NEAT1 plays a key role in the formation and function of nuclear condensates, namely paraspeckles (58, 59). Paraspeckles—formed through the process of liquid-liquid phase separation—are membraneless subnuclear compartments involved in a variety of cellular processes (60). Paraspeckles not only contribute to gene expression regulation, for instance through sequestration of SFPQ transcriptional factor (61) or by promoting microRNA biogenesis (62), but they also mediate nuclear retention of hyper A-to-I-edited transcripts (63, 64). They are enriched in Drosophila Behavior Human Splicing (DBHS) RBPs such as NONO, SFPQ, and PSPC1, but they also include other RBPs harboring prion like domains *e.g.* FUS, TDP43 and TAF15 (65).

NEAT1 is expressed as two distinct isoforms: a short polyadenylated transcript (NEAT1_1) and longer transcript (NEAT1_2) with a stabilizing triple structure at the RNA 3' end (NEAT1_2) (66). NEAT1_2 is essential for paraspeckle formation, by acting as a scaffold molecule (66). Indeed, assembly of these condensates strictly depends on the transcription of NEAT1_2 and consequently on the proteins that associate to it (58, 67). Genetic analysis revealed that NEAT1_2 contains different functional domains with distinct roles in paraspeckle production (13). Among these domains, the middle region has been demonstrated to recruit the core component NONO, which in turn initiates the paraspeckle assembly (13). NEAT1_2 displays a highly organized spatial arrangement in paraspeckles. Indeed, these condensates form spheroidal structures in which, while the 5' and 3' termini of NEAT1_2 are located at the paraspeckles periphery, the middle region is found in the assembly core (59, 68). Super resolution microscopy revealed that NEAT1 attract paraspeckle core components (*e.g.* FUS and DBHS family proteins) at the center of the spheroidal structure, while 5' and 3' ends of NEAT1_2 and TDBP-43 localize at the paraspeckle shell (59).

NEAT1 has been shown to be chemically modified post-transcriptionally with methylation at cytosines (m5C) and adenosines (m6A) (69, 70). Although the role of RNA modifications on NEAT1 activity needs to be further investigated, recent work has started to shed light on the consequence of m6A modification in NEAT1 and cancer progression (71–73). It has been recently demonstrated that ADAR1 controls NEAT1 stability by editing Alu sequences located in proximity of the AUF1 binding region. The modification decreases the ability of AUF1 to interact with its target (74).

Using CROSSalve we reproduced the secondary structure content of NEAT1 to a remarkable extent (**Experimental Procedures**), with an overall AUC of 0.85 or more compared to experimental data (34) (**Supplementary Image 3A,B**). As NEAT1 was not used in our training set, this result indicates that CROSSalve reliably predicts structural properties of transcripts.

The 5' and the 3' ends as well as nucleotides 5000-8500, 14500-15500 and 17500-19000 are highly conserved, as reported by previous studies (75) and enriched in double-stranded content (**Figure 5A; Supplementary Image 3C**). Comparing ADAR+ and ADAR- profiles of secondary structure content, we found that nucleotides 5000-8500 (corresponding to the highly conserved middle region (59), blue box in **Figure 5B**; see also **Supplementary Image 3C**) and especially nucleotides 7000-8500 undergo major structural rearrangements. This could impact RBP recruitment and ultimately RNP assembly. Indeed, as reported by eCLIP (37) (**Experimental Procedures**), this region involves interactions with proteins that undergo phase-separation including CPEB4, EIF4G2, EWSR1, FUBP3, FUS, HNRNPK, HNRNPM, KHSRP, MATR3, NONO, PCBP1, QKI, SFPQ, TARDBP, TIA1 and TIAL1

In the 5' region corresponding to nucleotides 14500-15500 (in the periphery of paraspeckle structure (59); blue box in **Figure 5B**), we predicted a slightly lower decrease in structural content and fewer changes in the binding of phase separating proteins FUBP3, HNRNPK and TIA1 (37). At nucleotides 17500-19000 we predicted an alteration in the structural content upon ADAR depletion but we could not retrieve any protein interactions. In the 5' tail corresponding to nucleotides 21000-22700, structural changes are accompanied by alteration of the binding sites of DDX6, EIF3G, EIF4G2, EWSR1, FUBP3, FUS, HNRNPK, IGF2BP1, KHDRBS1, NONO, SFPQ, TARDBP, TIA1 and TIAL1 (blue box in **Figure 5B**). In the 3' region corresponding to the highly conserved region located at nucleotides 1-1000 (in the periphery of the paraspeckle) structure and (59), we found a milder change of structure and few phase-separating proteins affected, including DDX6, EWSR1, FUBP3, FUS, HNRNPM, PCBP2, and TAF15 (37).

Our predictions suggest that A-to-I editing could impact interactions between NEAT1 and proteins crucial for phase separation. The RNA editing could alter the highly ordered spatial distribution of the assembly components within the paraspeckle core and shell.

XIST

XIST interacts with a plethora of RNA binding proteins (RBPs) involved in gene silencing, RNA modifications and localization. Its interactions with RBPs are mediated by specific RNA regions. XIST harbors six tandem repeats, 'Rep', named from A to F, which act as distinct functional domains essential for RBP binding (76–78). According to experimental and computational analysis, XIST Rep A, B and E are more structured than Rep D. XIST Rep A, B and E are crucial for RBP interactions (20, 39).

XIST Rep A at the 5' region is the most conserved in terms of sequence and copy number and it is probably the most contacted region essential for silencing at the early stages (25, 79). The highly structured Rep A functions as a modular platform for the association of RBPs (*e.g.* SPEN which in turn recruits the SMRT-HDAC3 complex required for gene silencing) and m6A modifications important for XIST functions (80). The cytosine rich Rep B/C modules directly interact with hnRNPK, which subsequently recruit the Polycomb repressive complex 1 (PRC1) (39, 81). The PRC1-mediated ubiquitination of histone H2A is recognized, through the JARID2 cofactor, by Polycomb-repressive complex 2 (PRC2) (82). It has been shown that while the initial recruitment of the Polycomb complexes is mediated by XIST Rep A, the XIST Rep B is crucial in stabilizing the Polycomb complexes and therefore maintaining gene silencing on the inactive X (83). Rep E controls XIST localization, in particular it is important to retain XIST at the inactive X site. Several RBPs have been proved to interact with Rep E, including TDB-43, CELF1, PTBP1 and MATR3 (39, 84, 85). The interaction of these proteins depends on the structural conformation of the Rep E, indicating that it is highly specific (76, 84). It has been hypothesized and proved that Rep E is involved in the formation of phase-separated condensates important for both XIST localization and gene silencing preservation (21, 85). Indeed, it was recently discovered that XIST forms large phase-separated assemblies, similar to stress granules, by progressively attracting proteins (7, 21, 86).

XIST is widely methylated with more than 70 N6-methyladenosine (m6A) residues (79). This modification seems to be crucial for the scaffolding activity of XIST by promoting protein recruitment. For instance, YTHDC1 proteins preferentially recognize m6A residues on XIST (79, 87). It is likely that other RNA modifications also play a crucial role in XIST function and assembly formation. Although ADAR was not identified to bind specifically to XIST in mouse (39), the Lee lab identified two binding partners, Q924C1 and Q9EPU (88). Thus, it is possible that ADAR1 could play an role in XIST activity.

Our calculations indicate that ADAR has the strongest effect on Rep B, C, D and E. Specifically, the double stranded content is significantly reduced upon ADAR depletion (**Figure 6A,B**). The 3' region of XIST, which does not play any functional role in X chromosome inactivation, is mildly affected by ADAR depletion (**Figure 6B**).

Considering RBP interactions in these regions (37), we predict that HNRNPK binding to Rep B may be altered upon A-to-I modifications. A study from the Heard laboratory showed that Rep B is necessary for PRC1 recruitment and the Brockdorff laboratory proved that HNRNPK, which physically interacts with PRC1, is directly involved in RNA binding (89). Importantly, HNRNPK is a SG component (90) and could play a role in XIST phase separation (7, 21). HNRNPK binds to other regions that are altered upon ADAR1 depletion, including Rep D where another SG component binds, HNRNPU (37). Additionally, Rep E interacts with several SG proteins such as MATR3, PTBP1, TARDBP, TIA1 (37).

A hypothesis on A-to-I effects on protein interactions

Our analysis indicates that structural changes in RNA molecules caused by ADAR could occur in regions contacted by RBPs. However, at present very little is known about the effects of A-to-I editing on RBP interactions.

We used *catRAPID* (38–40) to investigate the effects of A-to-I modifications on protein binding (**Figure 7**). As the parameters for “I” are not yet integrated in *catRAPID*, we adopted an approach based on the substitution of “A” with “G” followed by the calculation of RBP interactions. The A-to-I substitution was carried out at random positions within selected regions of NORAD, NEAT1 and XIST corresponding to peaks of ADAR+ profiles.

In the case of NORAD, PCBP1 is predicted to only mildly increase its binding propensity for nucleotides 100-300 depending on A-to-I modification (**Figure 7**). This region is the most structured in NORAD (**Figure 4**), and PCBP1 binding might be relevant for phase separation.

As for NEAT1, one of the most structured regions in presence of ADAR (i.e. nucleotides 7400-7600; **Figure 5**) is predicted to associate with TIA1 in a A-to-I dependent manner and, specifically, to interact tightly when the number of inosines increases (**Figure 7**). As the A-to-I editing is observed in adenosines that are in close proximity to the AU-rich of NEAT and TIA1 is a AU-rich binding protein (74), it is tempting to speculate that TIA1 propensity to phase separation might influence NEAT1 ability to form paraspeckles depending on ADAR editing.

The opposite effect is observed for RBM15 that interacts with Rep A at the 3’ end of XIST. In this case, *catRAPID* predictions indicate that A-to-I modifications decrease the binding ability of RBM15 for nucleotides 300-500 (**Figure 7**). While the structure of this region is only mildly affected by ADAR depletion (**Figure 6**), it is worth mentioning that RBM15 interaction mediates m6A modification of XIST (91), which suggests that A-to-I and m6A can both act in this region.

In summary, our computational analysis indicates that the A-to-I modification can affect RBP interactions in different ways, suggesting that further work in this exciting area is needed.

CONCLUSIONS

Advancements in next-generation sequencing and mass spectrometry allowed the identification of more than 170 different RNA post-transcriptional modifications (92), and the list is rapidly expanding (93). This finding indicates a complex regulatory network for RNAs and highlights the richness of the whole transcriptome. In an extremely versatile manner, chemical modifications adjust the biological functions of all RNA types (tRNA, rRNA, mRNA, snRNA, lncRNA or viral RNA) by altering their secondary structures, interaction networks, localizations and half-lives (94).

The contacts that RNAs establish and their spatial locations are crucial for the formation of membrane-less organelles (95). In the mRNA context, several modifications can influence liquid-liquid phase separation. For example, Ries and co-workers showed that mRNAs with multiple m6A sites can act as a multivalent scaffold for the m6A-binding proteins YTHDF (25). The m6A-mRNAs bring YTHDF low-complexity domains closer and trigger phase separation, thus regulating mRNA localization in compartments such as PBs, SGs or neuronal RNA granules, which ultimately affect stability and translation (25). Also, mRNA recruitment into SGs protects transcripts from irreversible aggregation under stress conditions (96). In agreement with this observation, we recently reported that the m1A modification leads to a more efficient recovery of protein synthesis upon heat shock (97). We also found that m1A RNAs and the enzyme involved in this modification (TRMT6/61A) play a role in SG formation (97). These findings indicate that both m6A and m1A can influence transcriptome turnover and translation by modulating the liquid-liquid phase separation process.

The adenosine deamination to inosine (A-to-I) is a very abundant nucleotide modification in animals. Specifically, millions of editing sites have been reported in humans and most of them are located in non-coding regions of the genome (31, 98). Despite its abundance, we still lack an overall understanding of the purpose and functional effects of this modification (31). The development of high-throughput methods is providing a new set of tools to investigate RNA editing.

In this manuscript, we used PARS data to introduce a new development of CROSSalive that predicts how A-to-I modifications affect the RNA structure (34). In lncRNAs, we detected a general enrichment of double-stranded regions, with just a few transcripts having the opposite behavior. For three functional lncRNAs, NEAT1, NORAD and XIST, we studied the relationship between A-to-I editing and protein interactions using available CLIP data (37) and *catRAPID* (38–40). We found that A-to-I editing is linked to alteration of interactions with several phase separating proteins. This could be a general trend for other lncRNAs (37).

Although protein-RNA interactions are crucial for the formation of membrane-less compartments, emerging evidence has shown that inter- and intra-molecular RNA-RNA interactions are also involved in the RNP condensate formation (51). Relatively to the examples presented here, NORAD tends to lose its structure under stress conditions (*e.g.*

arsenite treatment) and becomes more linear. In these conditions NORAD establishes inter-molecular interactions with G/C-rich mRNAs that accumulate in SGs (99). Similarly, the high local NEAT1 concentration at NEAT1 transcription sites allows the establishment of multivalent interactions among NEAT1 molecules, which could shape paraspeckles construction (51, 100, 101). As for XIST, Rep A has been reported to participate in intra-molecular interactions forming long repeat duplexes whose architecture is crucial for the assembly of SPEN (101). We cannot exclude that XIST establishes inter-molecular interactions also with other RNAs. Indeed, according to the RISE database, NORAD, NEAT1 and XIST engage in inter-molecular interactions with 18, 150 and 184 RNAs, respectively (102). We envision that alterations of RNA-RNA interactions—caused by A-to-I editing—undoubtedly have an impact in condensate formation and future research should aim at dissecting the role of lncRNA chemical modifications in RNA-RNA interactions contributing to cellular assemblies.

CROSSalive could be useful to investigate structural changes in contexts different from phase separation. An intrinsic limitation of our method is that it predicts local properties of structure, yet the ability to reproduce experimental data is significantly high (global accuracy of 0.90), which suggests that tertiary structure should play a minor role in the cellular environment, as also previously reported (103). We envisage that an implementation where long-distance interactions are taken into account would make CROSSalive applicable to a broader range of RNA molecules, especially long non-coding RNAs where tertiary interactions are expected to be relevant (104).

In the long term, we aim to expand our suite of algorithms to predict RNA secondary structure changes of other chemical modifications. CROSSalive can now predict structural changes associated with m6A and A-to-I modifications, and we plan to integrate the individual contributions of other modifications as soon as new experimental data will become available.

ACKNOWLEDGMENTS

The authors would like to thank the ‘RNA initiative’ at IIT and all the members of Tartaglia’s lab at CRG, Sapienza and IIT.

FUNDING

The research leading to these results has been supported by European Research Council [RIBOMYLOME_309545 and ASTRA_855923], the H2020 projects [IASIS_727658 and INFORE_825080], the Spanish Ministry of Science and Innovation (RYC2019-026752-I, PID2020-117454RA-I00/AEI/10.13039/501100011033 and PID2020-114627RB-I00/AEI/10.13039/501100011033) and L'Oréal-UNESCO For Women in Science Programme 2021.

REFERENCES

- [1] Kung, J. T. Y., Colognori, D., and Lee, J. T. (2013) Long noncoding RNAs: past, present, and future. *Genetics* 193, 651–669.
- [2] Marchese, D., de Groot, N. S., Lorenzo Gotor, N., Livi, C. M., and Tartaglia, G. G. (2016) Advances in the characterization of RNA-binding proteins. *Wiley interdisciplinary reviews. RNA* 7, 793–810.
- [3] Sanchez de Groot, N., Armaos, A., Graña-Montes, R., Alriquet, M., Calloni, G., et al. (2019) RNA structure drives interaction with proteins. *Nature Communications* 10, 3246.
- [4] Gotor, N. L., Armaos, A., Calloni, G., Torrent Burgas, M., Vabulas, R. M., et al. (2020) RNA-binding and prion domains: the Yin and Yang of phase separation. *Nucleic Acids Research* 48, 9491–9504.
- [5] Vandelli, A., Cid Samper, F., Torrent Burgas, M., Sanchez de Groot, N., and Tartaglia, G. G. (2022) The Interplay Between Disordered Regions in RNAs and Proteins Modulates Interactions Within Stress Granules and Processing Bodies. *Journal of Molecular Biology* 434, 167159.
- [6] Bolognesi, B., Lorenzo Gotor, N., Dhar, R., Cirillo, D., Baldrighi, M., et al. (2016) A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell Reports* 16, 222–231.
- [7] Cerase, A., Calabrese, J. M., and Tartaglia, G. G. (2022) Phase separation drives X-chromosome inactivation. *Nature Structural & Molecular Biology* 29, 183–185.
- [8] Khong, A., Matheny, T., Jain, S., Mitchell, S. F., Wheeler, J. R., et al. (2017) The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. *Molecular Cell* 68, 808-820.e5.
- [9] Hubstenberger, A., Courel, M., Bénard, M., Souquere, S., Ernoult-Lange, M., et al. (2017) P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Molecular Cell* 68, 144-157.e5.
- [10] Matheny, T., Van Treeck, B., Huynh, T. N., and Parker, R. (2021) RNA partitioning into stress granules is based on the summation of multiple interactions. *RNA* 27, 174–189.
- [11] Elguindy, M. M., and Mendell, J. T. (2021) NORAD-induced Pumilio phase separation is required for genome stability. *Nature* 595, 303–308.
- [12] Chorostecki, U., Saus, E., and Gabaldón, T. (2021) Structural characterization of NORAD reveals a stabilizing role of spacers and two new repeat units. *Computational and Structural Biotechnology Journal* 19, 3245–3254.
- [13] Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y. S., et al. (2018) Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Molecular Cell* 70, 1038-1053.e7.
- [14] Reuter, J. S., and Mathews, D. H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11, 129.
- [15] Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008) The Vienna RNA websuite. *Nucleic Acids Research* 36, W70-74.
- [16] Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., et al. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107.
- [17] Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., et al. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460, 711–716.
- [18] Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., et al. (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* 519, 486–490.
- [19] Wan, Y., Qu, K., Zhang, Q. C., Flynn, R. A., Manor, O., et al. (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709.
- [20] Delli Ponti, R., Marti, S., Armaos, A., and Tartaglia, G. G. (2017) A high-throughput approach to profile RNA structure. *Nucleic Acids Research* 45, e35.
- [21] Cerase, A., Armaos, A., Neumayer, C., Avner, P., Guttman, M., et al. (2019) Phase separation drives X-chromosome inactivation: a hypothesis. *Nature Structural & Molecular Biology* 26, 331–334.
- [22] Vandelli, A., Monti, M., Milanetti, E., Armaos, A., Rupert, J., et al. (2020) Structural analysis of

- SARS-CoV-2 genome and predictions of the human interactome. *Nucleic Acids Research* 48, 11270–11283.
- [23] Helm, M. (2006) Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Research* 34, 721–733.
- [24] Yue, Y., Liu, J., and He, C. (2015) RNA N^6 -methyladenosine methylation in post-transcriptional gene expression regulation. *Genes & Development* 29, 1343–1355.
- [25] Ries, R. J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., et al. (2019) m6A enhances the phase separation potential of mRNA. *Nature* 571, 424–428.
- [26] Anders, M., Chelysheva, I., Goebel, I., Trenkner, T., Zhou, J., et al. (2018) Dynamic m6A methylation facilitates mRNA triaging to stress granules. *Life Science Alliance* 1, e201800113.
- [27] Khong, A., Matheny, T., Huynh, T. N., Babl, V., and Parker, R. (2021) Limited effects of m⁶A modification on mRNA partitioning into stress granules. *Molecular Biology*.
- [28] Ponti, R. D., Armaos, A., Vandelli, A., and Tartaglia, G. G. (2020) CROSSalive: a web server for predicting the in vivo structure of RNA molecules. *Bioinformatics* 36, 940–941.
- [29] Chen, Y. G., Satpathy, A. T., and Chang, H. Y. (2017) Gene regulation in the immune system by long noncoding RNAs. *Nature Immunology* 18, 962–972.
- [30] Yablonovitch, A. L., Deng, P., Jacobson, D., and Li, J. B. (2017) The evolution and adaptation of A-to-I RNA editing. *PLoS genetics* 13, e1007064.
- [31] Nishikura, K. (2016) A-to-I editing of coding and non-coding RNAs by ADARs. *Nature Reviews Molecular Cell Biology* 17, 83–96.
- [32] Silvestris, D. A., Scopa, C., Hanchi, S., Locatelli, F., and Gallo, A. (2020) De Novo A-to-I RNA Editing Discovery in lncRNA. *Cancers* 12, 2959.
- [33] Kazimierczyk, M., and Wrzesinski, J. (2021) Long Non-Coding RNA Epigenetics. *International Journal of Molecular Sciences* 22, 6166.
- [34] Solomon, O., Di Segni, A., Cesarkas, K., Porath, H. T., Marcu-Malina, V., et al. (2017) RNA editing by ADAR1 leads to context-dependent transcriptome-wide changes in RNA secondary structure. *Nature Communications* 8, 1440.
- [35] Corbet, G. A., Burke, J. M., and Parker, R. (2021) ADAR1 limits stress granule formation through both translation-dependent and translation-independent mechanisms. *Journal of Cell Science* 134, jcs258783.
- [36] Samuel, C. E. (2012) ADARs: viruses and innate immunity. *Current Topics in Microbiology and Immunology* 353, 163–195.
- [37] Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., et al. (2020) A large-scale binding and functional map of human RNA-binding proteins. *Nature* 583, 711–719.
- [38] Bellucci, M., Agostini, F., Masin, M., and Tartaglia, G. G. (2011) Predicting protein associations with long noncoding RNAs. *Nature Methods* 8, 444–445.
- [39] Cirillo, D., Blanco, M., Armaos, A., Bunes, A., Avner, P., et al. (2016) Quantitative predictions of protein interactions with long noncoding RNAs. *Nature Methods* 14, 5–6.
- [40] Armaos, A., Colantoni, A., Proietti, G., Rupert, J., and Tartaglia, G. G. (2021) *cat* RAPID *omics* v2.0: going deeper and wider in the prediction of protein–RNA interactions. *Nucleic Acids Research* 49, W72–W79.
- [41] Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics (Oxford, England)* 26, 680–682.
- [42] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, et al. (2015) TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.
- [43] Chollet, F. and others (2015) Keras.
- [44] Saus, E., Willis, J. R., Prysycz, L. P., Hafez, A., Llorens, C., et al. (2018) nextPARS: parallel probing of RNA structures in Illumina. *RNA* 24, 609–619.
- [45] Lang, B., Armaos, A., and Tartaglia, G. G. (2019) RNAAct: Protein-RNA interaction predictions for model organisms with supporting experimental data. *Nucleic Acids Research* 47, D601–D606.
- [46] Delli Ponti, R., and Mutwil, M. (2021) Structural landscape of the complete genomes of dengue virus serotypes and other viral hemorrhagic fevers. *BMC genomics* 22, 352.
- [47] Yang, Y., Hsu, P. J., Chen, Y.-S., and Yang, Y.-G. (2018) Dynamic transcriptomic m6A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Research* 28, 616–624.

- [48] Moon, S. L., Morisaki, T., Stasevich, T. J., and Parker, R. (2020) Coupling of translation quality control and mRNA targeting to stress granules. *The Journal of Cell Biology* 219, e202004120.
- [49] Namkoong, S., Ho, A., Woo, Y. M., Kwak, H., and Lee, J. H. (2018) Systematic Characterization of Stress-Induced RNA Granulation. *Molecular Cell* 70, 175-187.e8.
- [50] Lee, S., Kopp, F., Chang, T.-C., Sataluri, A., Chen, B., et al. (2016) Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* 164, 69–80.
- [51] Aillaud, M., and Schulte, L. N. (2020) Emerging Roles of Long Noncoding RNAs in the Cytoplasmic Milieu. *Non-coding RNA* 6, E44.
- [52] Elguindy, M. M., Kopp, F., Goodarzi, M., Rehfeld, F., Thomas, A., et al. (2019) PUMILIO, but not RBMX, binding is required for regulation of genomic stability by noncoding RNA NORAD. *eLife* 8, e48625.
- [53] Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., et al. (2016) A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nature Communications* 7, 12209.
- [54] Li, G., Ma, L., He, S., Luo, R., Wang, B., et al. (2022) WTAP-mediated m6A modification of lncRNA NORAD promotes intervertebral disc degeneration. *Nature Communications* 13, 1469.
- [55] Munschauer, M., Nguyen, C. T., Sirokman, K., Hartigan, C. R., Hogstrom, L., et al. (2018) The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* 561, 132–136.
- [56] Liu, N., Zhou, K. I., Parisien, M., Dai, Q., Diatchenko, L., et al. (2017) N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Research* 45, 6051–6063.
- [57] Tichon, A., Perry, R. B.-T., Stojic, L., and Ulitsky, I. (2018) SAM68 is required for regulation of Pumilio by the NORAD long noncoding RNA. *Genes & Development* 32, 70–78.
- [58] Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., et al. (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Molecular Cell* 33, 717–726.
- [59] West, J. A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., et al. (2016) Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *The Journal of Cell Biology* 214, 817–830.
- [60] Fox, A. H., Nakagawa, S., Hirose, T., and Bond, C. S. (2018) Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. *Trends in Biochemical Sciences* 43, 124–135.
- [61] Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., Li, R., et al. (2014) NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Molecular Biology of the Cell* 25, 169–183.
- [62] Jiang, L., Shao, C., Wu, Q.-J., Chen, G., Zhou, J., et al. (2017) NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing. *Nature Structural & Molecular Biology* 24, 816–824.
- [63] Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., et al. (2005) Regulating gene expression through RNA nuclear retention. *Cell* 123, 249–263.
- [64] Chen, L.-L., and Carmichael, G. G. (2009) Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Molecular Cell* 35, 467–478.
- [65] Hirose, T., Yamazaki, T., and Nakagawa, S. (2019) Molecular anatomy of the architectural NEAT1 noncoding RNA: The domains, interactors, and biogenesis pathway required to build phase-separated nuclear paraspeckles. *WIREs RNA* 10.
- [66] Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y. F., Goshima, N., et al. (2012) Alternative 3'-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *The EMBO journal* 31, 4020–4034.
- [67] Mao, Y. S., Sunwoo, H., Zhang, B., and Spector, D. L. (2011) Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nature Cell Biology* 13, 95–101.
- [68] Souquere, S., Beauclair, G., Harper, F., Fox, A., and Pierron, G. (2010) Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies. *Molecular Biology of the Cell* 21, 4020–4027.
- [69] Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., et al.

- (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206.
- [70] Squires, J. E., Patel, H. R., Nusch, M., Sibbritt, T., Humphreys, D. T., et al. (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Research* 40, 5023–5033.
- [71] Zhang, J., Guo, S., Piao, H.-Y., Wang, Y., Wu, Y., et al. (2019) ALKBH5 promotes invasion and metastasis of gastric cancer by decreasing methylation of the lncRNA NEAT1. *Journal of Physiology and Biochemistry* 75, 379–389.
- [72] Yao, F.-Y., Zhao, C., Zhong, F.-M., Qin, T.-Y., Wen, F., et al. (2021) m(6)A Modification of lncRNA NEAT1 Regulates Chronic Myelocytic Leukemia Progression via miR-766-5p/CDKN1A Axis. *Frontiers in Oncology* 11, 679634.
- [73] Wen, S., Wei, Y., Zen, C., Xiong, W., Niu, Y., et al. (2020) Long non-coding RNA NEAT1 promotes bone metastasis of prostate cancer through N6-methyladenosine. *Molecular Cancer* 19, 171.
- [74] Vlachogiannis, N. I., Sachse, M., Georgiopoulos, G., Zormpas, E., Bampatsias, D., et al. (2021) Adenosine-to-inosine Alu RNA editing controls the stability of the pro-inflammatory long noncoding RNA NEAT1 in atherosclerotic cardiovascular disease. *Journal of Molecular and Cellular Cardiology* 160, 111–120.
- [75] Adriaens, C., Rambow, F., Bervoets, G., Silla, T., Mito, M., et al. (2019) The long noncoding RNA *NEAT1_1* is seemingly dispensable for normal tissue homeostasis and cancer cell growth. *RNA* 25, 1681–1695.
- [76] Pintacuda, G., Young, A. N., and Cerase, A. (2017) Function by Structure: Spotlights on Xist Long Non-coding RNA. *Frontiers in Molecular Biosciences* 4, 90.
- [77] Raposo, A. C., Casanova, M., Gendrel, A.-V., and da Rocha, S. T. (2021) The tandem repeat modules of Xist lncRNA: a swiss army knife for the control of X-chromosome inactivation. *Biochemical Society Transactions* 49, 2549–2560.
- [78] Yen, Z. C., Meyer, I. M., Karalic, S., and Brown, C. J. (2007) A cross-species comparison of X-chromosome inactivation in Eutheria. *Genomics* 90, 453–463.
- [79] Patil, D. P., Chen, C.-K., Pickering, B. F., Chow, A., Jackson, C., et al. (2016) m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* 537, 369–373.
- [80] Lu, Z., Guo, J. K., Wei, Y., Dou, D. R., Zarnegar, B., et al. (2020) Structural modularity of the XIST ribonucleoprotein complex. *Nature Communications* 11, 6163.
- [81] Cerase, A., and Tartaglia, G. G. (2020) Long non-coding RNA-polycomb intimate rendezvous. *Open Biology* 10, 200126.
- [82] Kasinath, V., Beck, C., Sauer, P., Poepsel, S., Kosmatka, J., et al. (2021) JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. *Science (New York, N.Y.)* 371, eabc3393.
- [83] Colognori, D., Sunwoo, H., Wang, D., Wang, C.-Y., and Lee, J. T. (2020) Xist Repeats A and B Account for Two Distinct Phases of X Inactivation Establishment. *Developmental Cell* 54, 21–32.e5.
- [84] Smola, M. J., Christy, T. W., Inoue, K., Nicholson, C. O., Friedersdorf, M., et al. (2016) SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proceedings of the National Academy of Sciences of the United States of America* 113, 10322–10327.
- [85] Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancina Leon, W. R., et al. (2020) A protein assembly mediates Xist localization and gene silencing. *Nature* 587, 145–151.
- [86] Jachowicz, J. W., Strehle, M., Banerjee, A. K., Blanco, M. R., Thai, J., et al. (2022) Xist spatially amplifies SHARP/SPEN recruitment to balance chromosome-wide silencing and specificity to the X chromosome. *Nature Structural & Molecular Biology* 29, 239–249.
- [87] Huisman, B., Manske, G., Carney, S., and Kalantry, S. (2017) Functional Dissection of the m6A RNA Modification. *Trends in Biochemical Sciences* 42, 85–86.
- [88] Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., et al. (2015) Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science (New York, N.Y.)* 349.
- [89] Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B. A., Solcan, N., et al. (2017) hnRNP Recruits

- PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Molecular Cell* 68, 955-969.e10.
- [90] Markmiller, S., Soltanieh, S., Server, K. L., Mak, R., Jin, W., et al. (2018) Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. *Cell* 172, 590-604.e13.
- [91] Coker, H., Wei, G., Moindrot, B., Mohammed, S., Nesterova, T., et al. (2020) The role of the Xist 5' m6A region and RBM15 in X chromosome inactivation. *Wellcome Open Research* 5, 31.
- [92] Boccaletto, P., Machnicka, M. A., Purta, E., Piątkowski, P., Bagiński, B., et al. (2018) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Research* 46, D303–D307.
- [93] Schaefer, M., Kapoor, U., and Jantsch, M. F. (2017) Understanding RNA modifications: the promises and technological bottlenecks of the 'epitranscriptome.' *Open Biology* 7, 170077.
- [94] Armaos, A., Zacco, E., Sanchez de Groot, N., and Tartaglia, G. G. (2021) RNA-protein interactions: Central players in coordination of regulatory networks. *BioEssays* 43, 2000118.
- [95] Gomes, E., and Shorter, J. (2019) The molecular language of membraneless organelles. *The Journal of Biological Chemistry* 294, 7115–7127.
- [96] Marcelo, A., Koppenol, R., de Almeida, L. P., Matos, C. A., and Nóbrega, C. (2021) Stress granules, RNA-binding proteins and polyglutamine diseases: too much aggregation? *Cell Death & Disease* 12, 592.
- [97] Alriquet, M., Calloni, G., Martínez-Limón, A., Delli Ponti, R., Hanspach, G., et al. (2021) The protective role of m1A during stress-induced granulation. *Journal of Molecular Cell Biology* 12, 870–880.
- [98] Porath, H. T., Knisbacher, B. A., Eisenberg, E., and Levanon, E. Y. (2017) Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundance. *Genome Biology* 18, 185.
- [99] Ziv, O., Farberov, S., Lau, J. Y., Miska, E., Kudla, G., et al. (2021) Structural features within the NORAD long noncoding RNA underlie efficient repression of Pumilio activity. *Molecular Biology*.
- [100] Lin, Y., Schmidt, B. F., Bruchez, M. P., and McManus, C. J. (2018) Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. *Nucleic Acids Research* 46, 3742–3752.
- [101] Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., et al. (2016) RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* 165, 1267–1279.
- [102] Gong, J., Shao, D., Xu, K., Lu, Z., Lu, Z. J., et al. (2018) RISE: a database of RNA interactome from sequencing experiments. *Nucleic Acids Research* 46, D194–D201.
- [103] Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J. S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505, 701–705.
- [104] Zampetaki, A., Albrecht, A., and Steinhofel, K. (2018) Long Non-coding RNA Structure and Function: Is There a Link? *Frontiers in Physiology* 9.

FIGURES LEGENDS

Figure 1. A-to-I editing and its possible implication on condensate formation. **A.** A-to-I editing is catalyzed by adenosine deaminases acting on RNA (ADAR) proteins that convert the adenosine (green) into inosine (purple) by hydrolytic deamination at position C6. **B.** Inosine mimics guanosine and therefore base-pairs favorably with cytosine. Editing causes a structural rearrangement of target RNA. **C.** Modified lincRNAs, due to the altered structure, attract different RNA binding proteins (RBPs) compared to unedited RNAs. We predict that A-to-I editing mediates the recruitment of distinct RBPs and could have an impact on the formation of condensates such as SGs and PBs in the cell.

Figure 2. The CROSSalve algorithm. **A.** Workflow for the processing of PARS data, including the selection of high-propensity nucleotides and of the RNA fragments, used for training CROSSalve (log(S1) and log(V1) are the read counts measured upon S1 nuclease (S1) and RNase V1 (V1) cleaving; **Experimental Procedures**). **B.** Training statistics of ADAR- model including loss (blue line) and training accuracies (orange) at each epoch. The accuracy on an independent test set excluded from the training (i.e. 10% of the training set) is reported. **C.** Training statistics of ADAR+ model including loss (blue line) and training accuracies (orange) for each epoch. The accuracy on an independent test set excluded from the training (i.e. 10% of the training set) is reported.

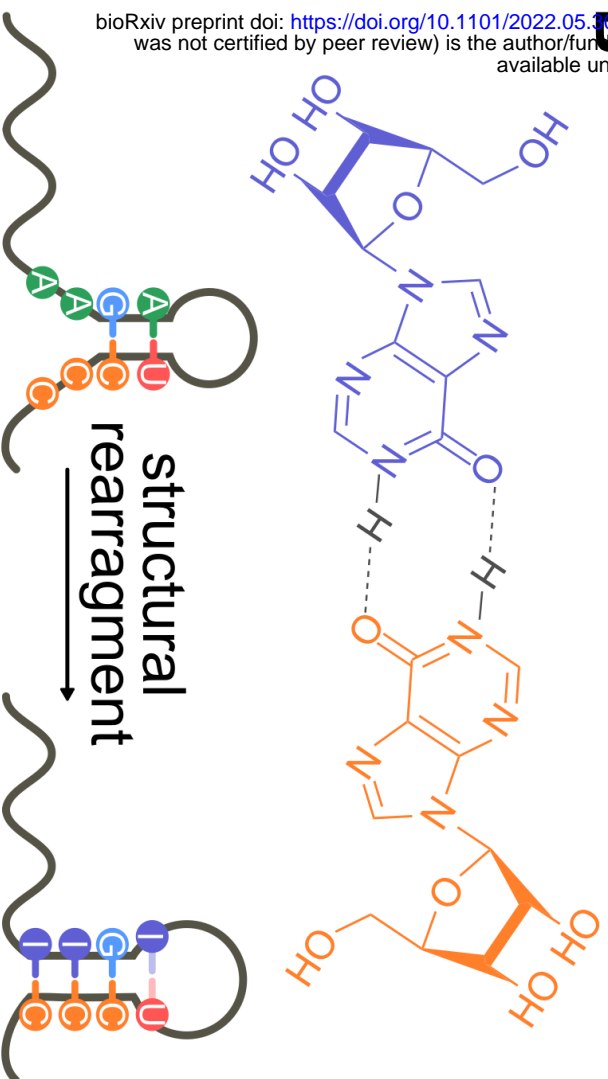
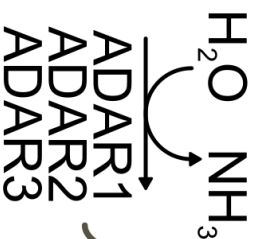
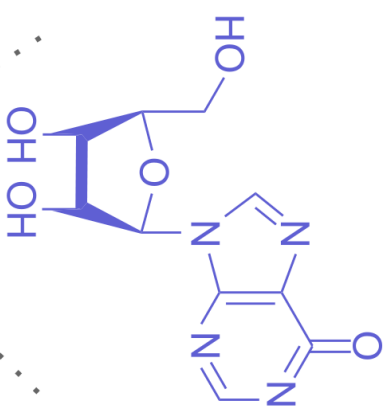
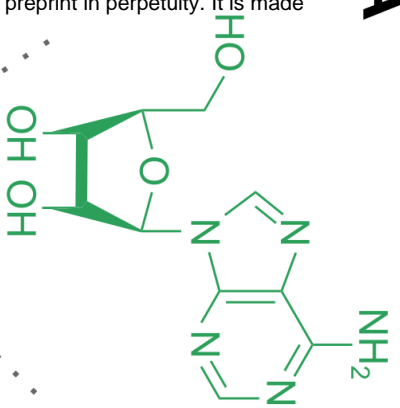
Figure 3. Effects of A-to-I on the structure of long non-coding RNAs. **A.** Boxplots showing the distribution of structural content (i.e. DS% or double-stranded nucleotides) predicted in absence (“-” in red) and presence of ADAR (“+” in gray) for long intergenic non-coding RNAs (lincRNAs). Significance is reported (Kolmogorov-Smirnov). **B.** For lincRNAs of interest (NORAD, XIST and NEAT1), we report differences in the structural content considering absence (“-”) and presence of ADAR (“+”).

Figure 4. Structural profile of NORAD. **A.** Structural content in presence of ADAR (ADAR+). The 5’ end as well as nucleotides 4100-4650 are enriched in double-stranded regions. **B.** Difference between the structural content in presence (ADAR+) and absence of ADAR (ADAR-). Red bars indicate 12 repeat units (of approximately 300 nt) enriched in the sequence pattern UGURUAUA. Blue boxes indicate regions with the most significant structural changes (i.e. structural peaks and valleys).

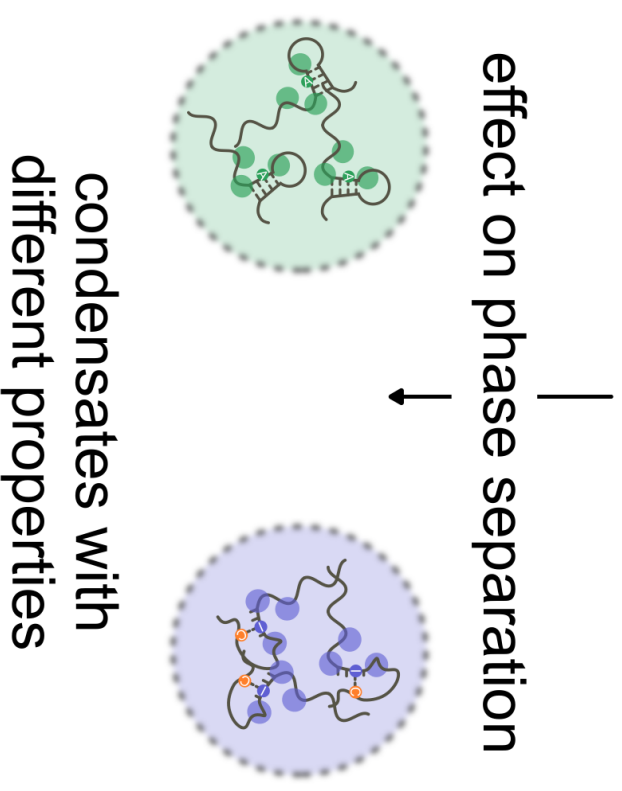
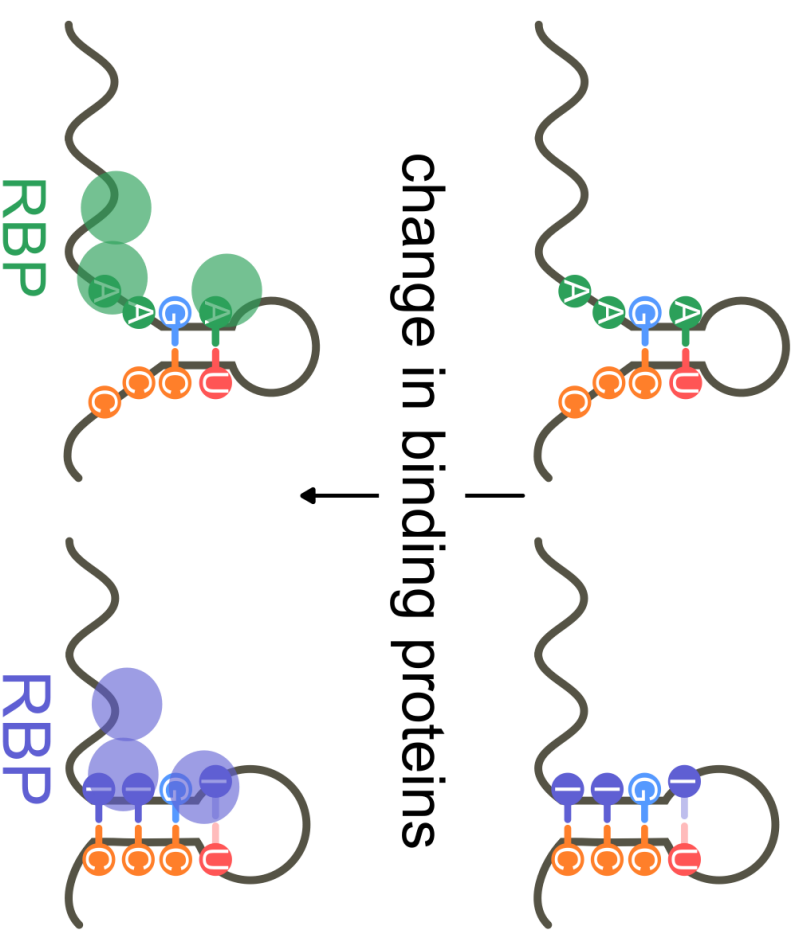
Figure 5. Structural profile of NEAT1. **A.** Structural content in presence of ADAR (ADAR+). The 5’ and the 3’ ends as well as nucleotides 5000-8500, 14500-15500 and 17500-19000 are enriched in double-stranded regions. **B.** Difference between the structural content in presence (ADAR+) and absence of ADAR (ADAR-). Red bars indicate evolutionary conserved regions that are important for paraspeckle structure (59); (**Supplementary Figure 3**). Blue boxes indicate regions with the most significant structural changes (i.e. structural peaks and valleys).

Figure 6. Structural profile of XIST. **A.** Structural content in presence of ADAR (ADAR+). Rep A, B, C, D and E are enriched in double-stranded regions. **B.** Difference between the structural content in presence (ADAR+) and absence of ADAR (ADAR-). Red bars indicate repeat units (78). Blue boxes indicate regions with the most significant structural changes.

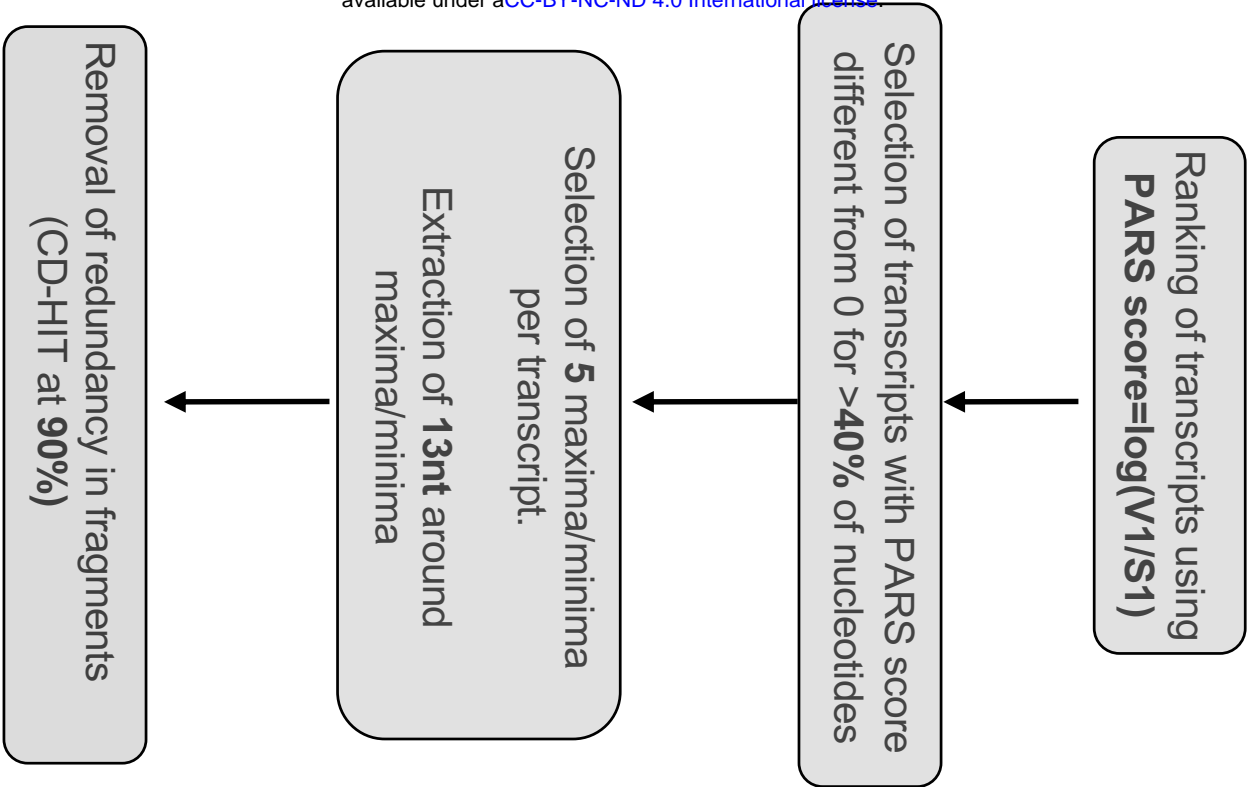
Figure 7. Effects of A-to-I editing on protein binding. *catRAPID* predictions indicate that PCBP1 only mildly increases its binding propensity for nucleotides 100-300 of NORAD depending on A-to-I modification. By contrast, nucleotides 7400-7600 of NEAT1 associate with TIA1 more tightly when the number of inosines increases. The opposite effect is observed for RBM15 that interacts with Rep A in the 3' end of XIST. We note that all the interaction propensities are >0, indicating that protein binding is always predicted to occur.



C different lincRNA structures

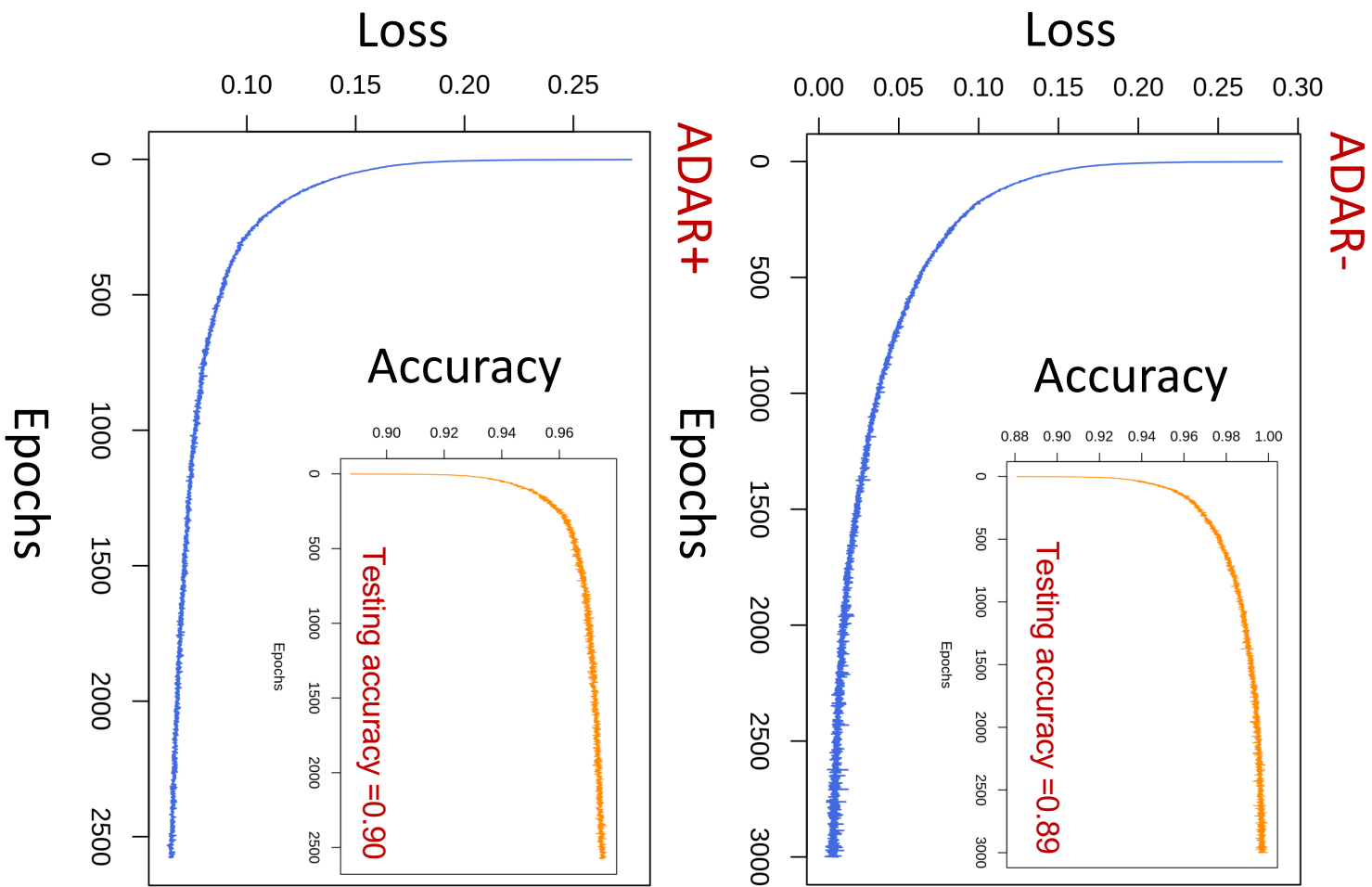


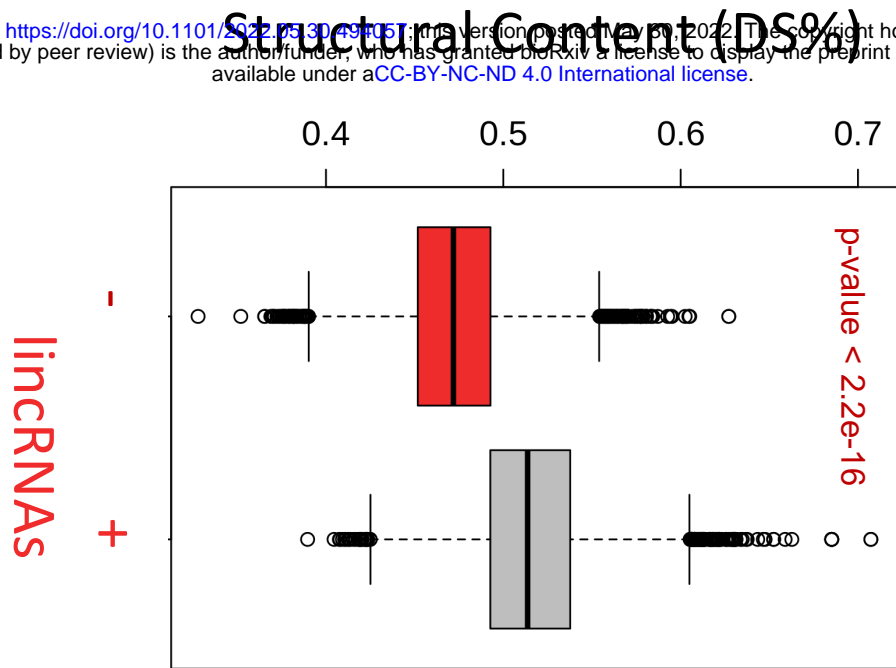
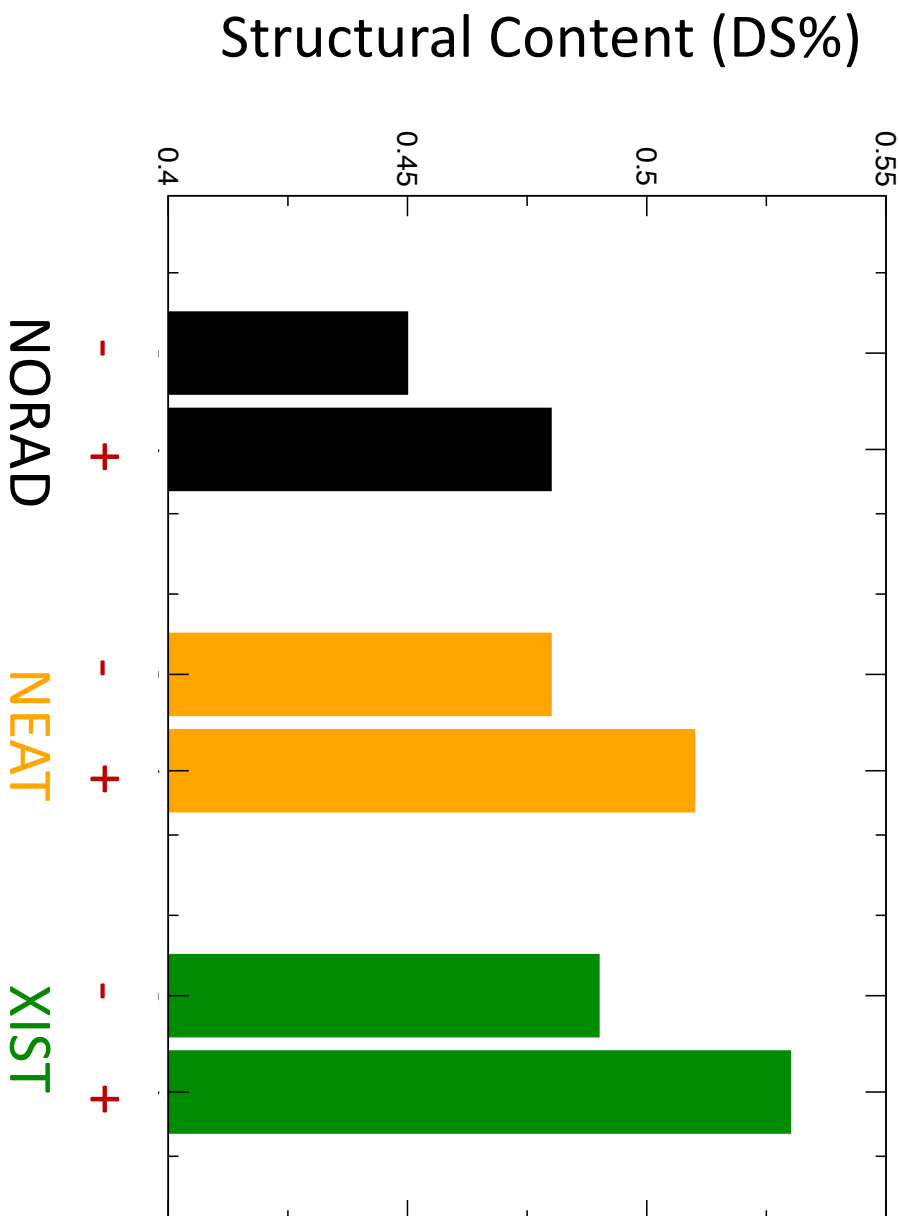
A

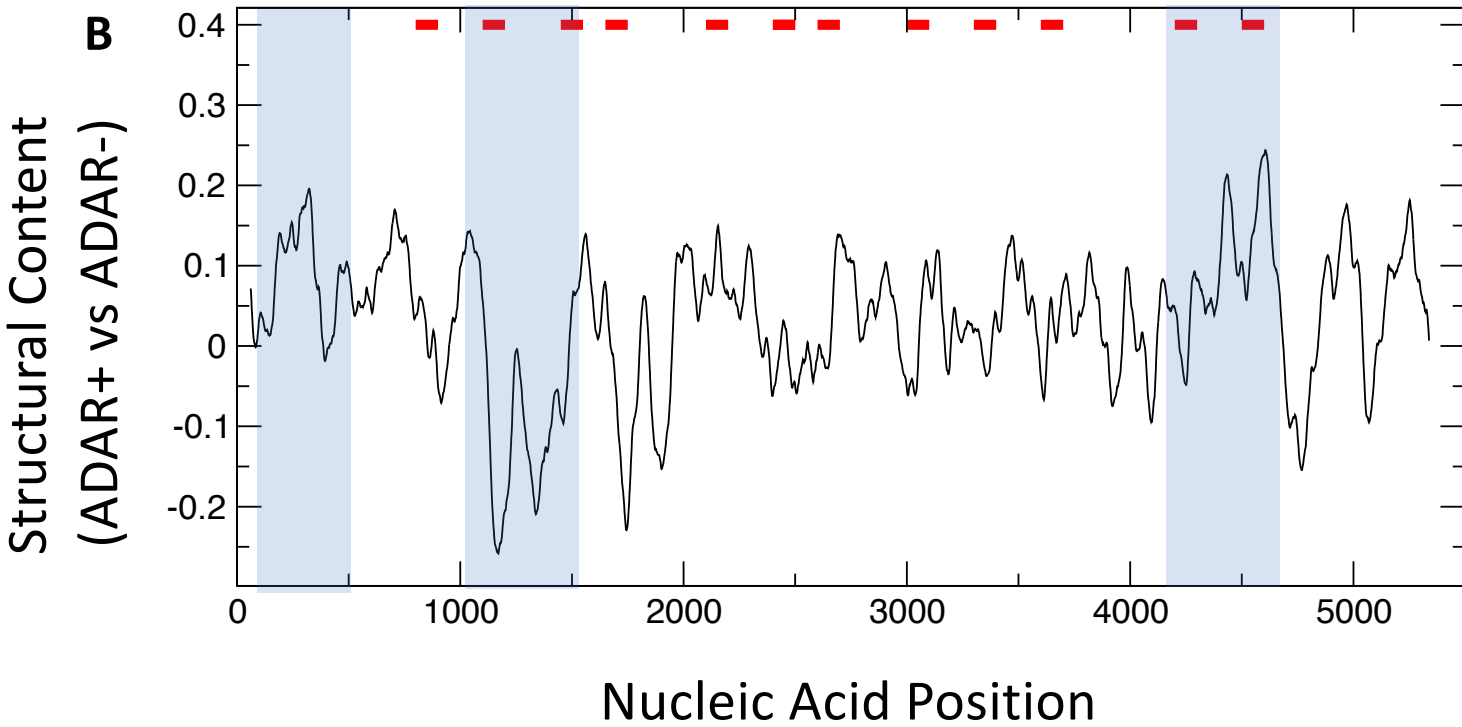
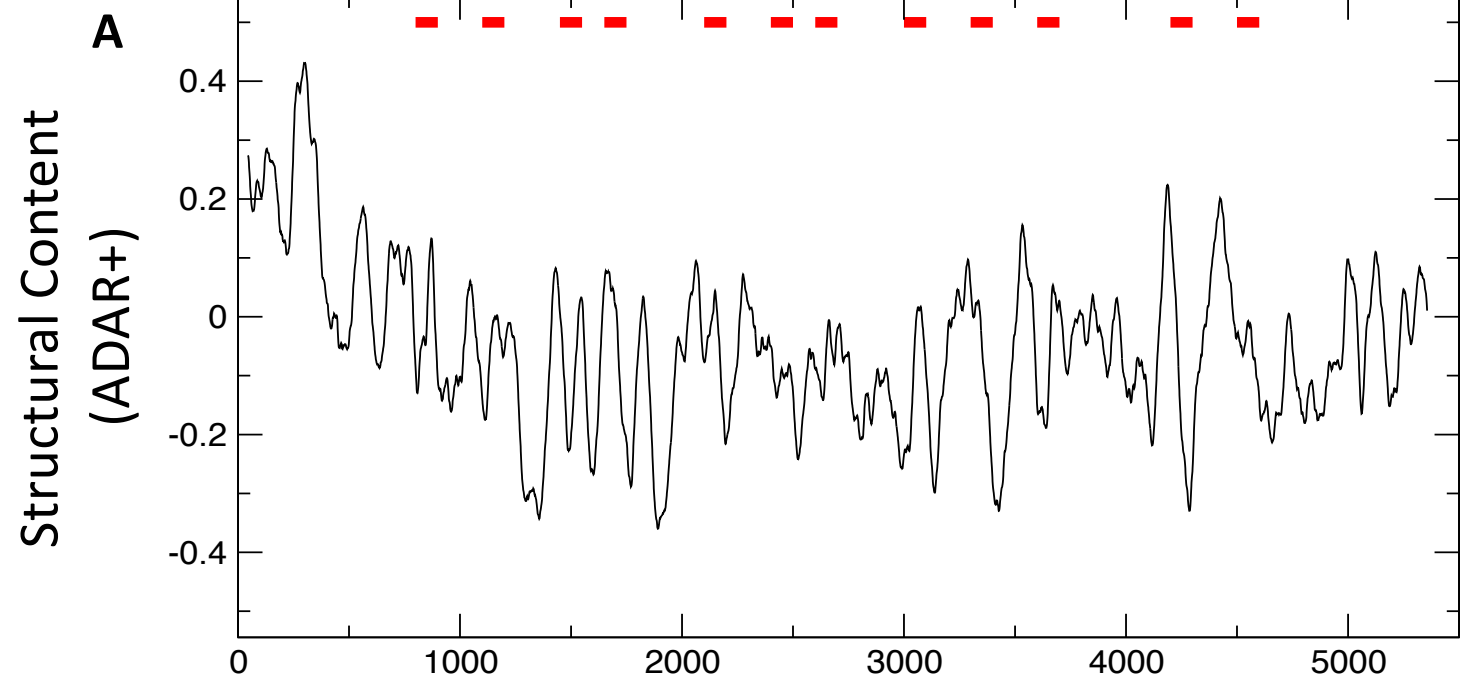


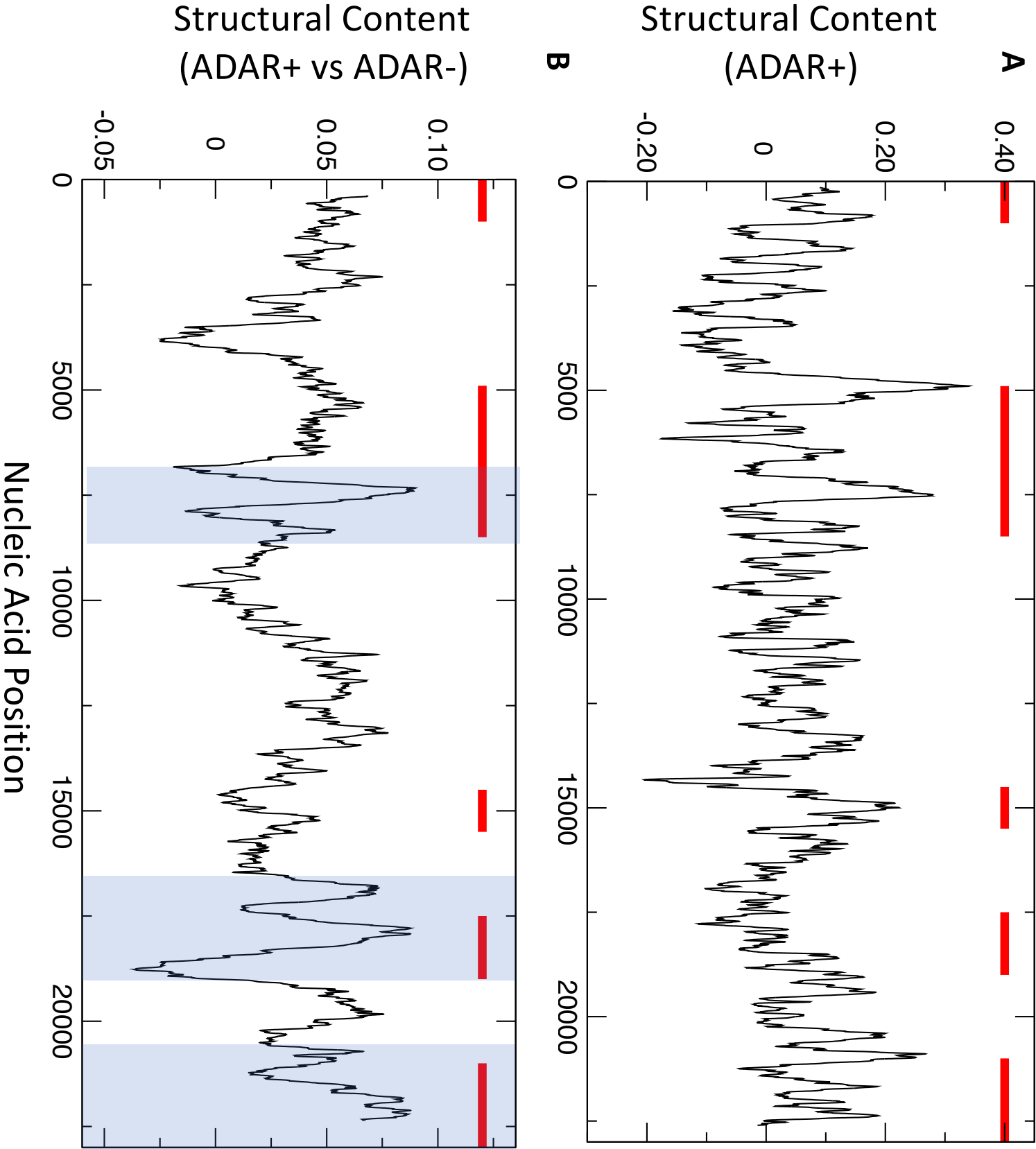
B

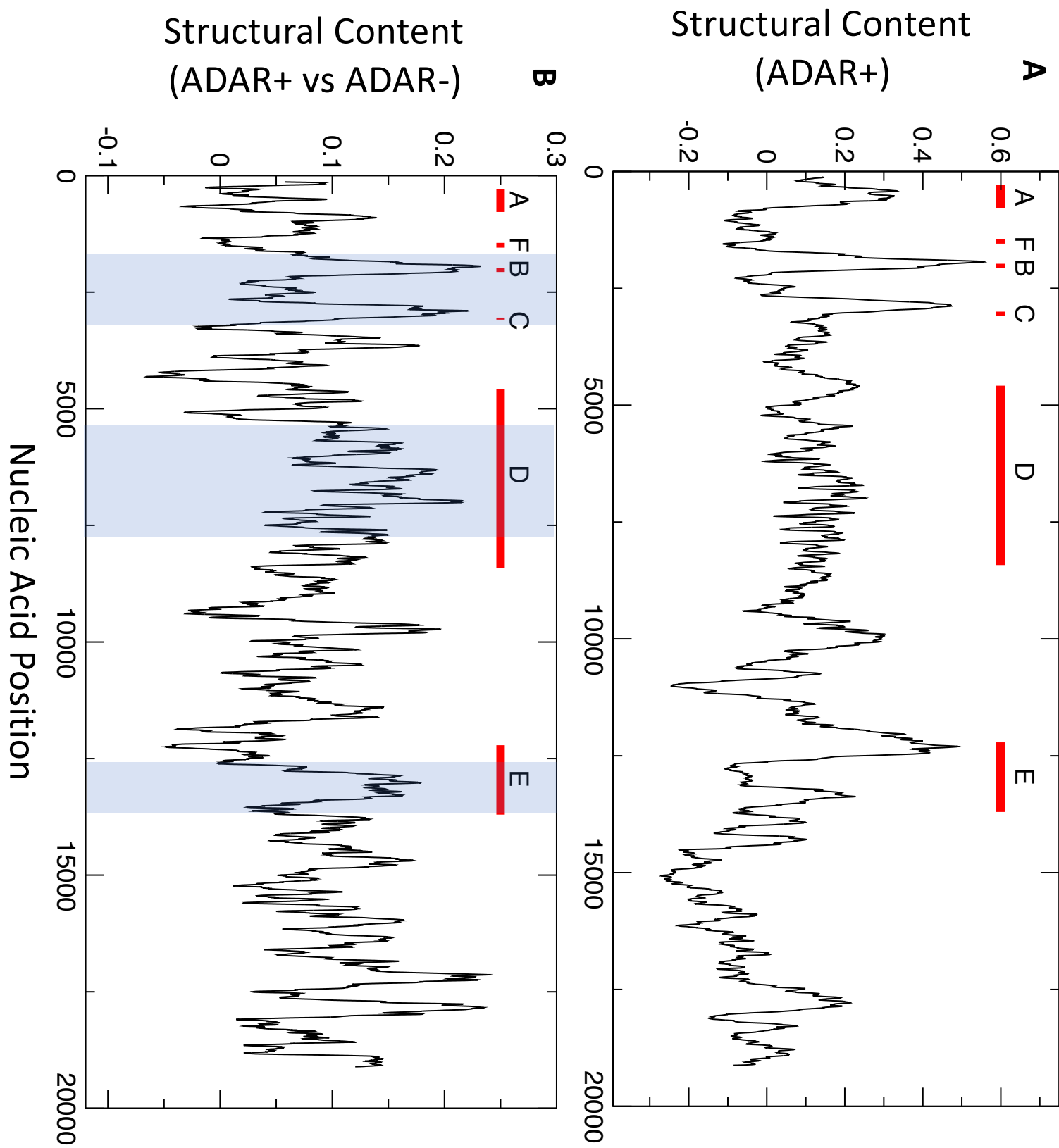
C



A**B**







Interaction Propensity

