



HAL
open science

ANRIL et la régulation génique chez l'Homme : une analyse des multiples facettes d'un long ARN non codant

Sylvain Maenner

► **To cite this version:**

Sylvain Maenner. ANRIL et la régulation génique chez l'Homme : une analyse des multiples facettes d'un long ARN non codant. Sciences du Vivant [q-bio]. Université de Lorraine, 2024. tel-04711136

HAL Id: tel-04711136

<https://hal.science/tel-04711136v1>

Submitted on 26 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**UNIVERSITÉ
DE LORRAINE**

Habilitation à Diriger des Recherches
Science de la Vie et de la Santé

**ANRIL et régulation génique chez l'Homme :
Une analyse des multiples facettes
dun long ARN non -codant**

**Sylvain MAENNER
MCU-UL**

Soutenue le 19 septembre 2024

Jury composé de :

Rapporteurs

Céline MOREY

Eric ADRIAENSSENS

Irwin DAVIDSON

CR INSERM, Paris

Pr Université de Lille

DR CNRS, Strasbourg

Examineurs

Reini LUCO

Clément CARRE

Stéphanie GRANDEMANGE

Bruno CHARPENTIER

DR CNRS, Paris

Pr Sorbonne Université, Paris

Pr Université de Lorraine

Pr Université de Lorraine

PLAN DU PROJET ANRILAct

<i>I. Identification des facteurs protéiques associés à ANRIL par la méthode CARPID</i>	<i>p.4</i>
<i>I.1 Données bibliographiques et données générées par l'équipe qui motivent cette partie du projet</i>	<i>p.4</i>
<i>I.2 Identification exhaustive des facteurs protéiques associées à ANRIL par la méthode CARPID</i>	<i>p.5</i>
<i>I.3. Hiérarchisation des partenaires protéiques les plus pertinents sur la base des données bibliographiques et celles générées par le groupe</i>	<i>p.6</i>
<i>I.3.1. Les modificateurs de la chromatine comme partenaires pertinents d'ANRIL</i>	<i>p.6</i>
<i>I.3.2 Les facteurs de régulation de l'épissage alternatif comme partenaires pertinents d'ANRIL</i>	<i>p.9</i>
<i>I.3.3 Les facteurs de modulation de la topologie nucléaire comme partenaires pertinents d'ANRIL</i>	<i>p.12</i>
<i>I.4. Détection/validation des associations ANRIL/protéines candidates</i>	<i>p.14</i>
<i>I.5. Evaluation de l'effet de l'absence des protéines candidates sur les activités d'ANRIL</i>	<i>p.14</i>
<i>II. Identification des régions ARN d'ANRIL responsables de la fixation des facteurs protéiques</i>	<i>p.15</i>
<i>II.1. Arguments en faveur de cette problématique</i>	<i>p.15</i>
<i>II.2. Evaluation de la présence des protéines candidates dans les RNP formées avec les exons d'ANRIL par chromatographie d'affinité MS2-MBP</i>	<i>p.15</i>
<i>II.3. Génération des lignées déplétées pour les régions ARN identifiées et évaluation de la fixation des facteurs protéiques aux formes tronquées d'ANRIL</i>	<i>p.16</i>
<i>II.4. Caractérisation des activités des RNP protéines/ANRIL</i>	<i>p.16</i>
<i>II.5. Etude des relations Structure/fonction entre les régions ARN et les protéines partenaires d'ANRIL</i>	<i>p.17</i>
<i>III. Extrapolation des données générées avec la lignée cellulaire HEK293 à d'autres lignées cellulaires et recherche de corrélations avec l'état pathologique</i>	<i>p.17</i>
<i>IV. Poursuite du développement de la molécule TDO2 et projets annexes</i>	<i>p.23</i>
<i>V. Identification et étude du rôle des modifications post-transcriptionnelles dans les activités d'ANRIL</i>	<i>p.20</i>
<i>VI. Ressources et bénéfices attendus du projet ANRILAct</i>	<i>p.23</i>
<i>Bibliographie</i>	<i>p.25</i>

ANRIL : un régulateur de l'expression génique à multiples facettes

Le projet ANRILAct vise à identifier les partenaires protéiques, les régions ARN et les modifications post-transcriptionnelles d'ANRIL impliquées dans les activités de ce dernier. Il est basé sur des données générées par le groupe en faveur de l'existence d'activités inédites portées par ANRIL. Cet exposé vise donc à décrire ces évidences en prenant des exemples représentatifs générés par l'équipe et appuyés par les connaissances bibliographiques. Il ne se veut aucunement exhaustif car il se « limite » à mettre en perspectives les éléments de pertinence nécessaires à la bonne compréhension du projet.

I. Identification des facteurs protéiques associés à ANRIL par la méthode CARPID

I.1. Données bibliographiques et celles générées par l'équipe qui motivent ce projet

Les longs ARN non-codants interagissent avec des « protéines qui se lient aux ARN » nommées RBP (RNA binding proteins) par le biais de domaines séquentiels voire structuraux pour former des complexes ribonucléoprotéiques (RNP) fonctionnels. Ces interactions réciproques peuvent influencer la maturation, les modifications post-transcriptionnelles, la stabilité et la fonction des lncRNAs (Jonas *et al.*, 2020). A titre d'exemple, les protéines SRSF1 et AUF1 interagissent avec le lncRNA NEAT pour augmenter et diminuer sa stabilité respectivement (Zhou *et al.*, 2019 ; Yoon *et al.*, 2014). Des expériences d'immunoprécipitation de la chromatine par sélection ARN suivie d'une analyse par spectrométrie de masse (ChIRP-MS) ont mis en évidence que le lncRNA Xist interagit avec 81 protéines impliquées notamment dans les voies de modulation des propriétés de la chromatine ou de la matrice nucléaire (Chu *et al.*, 2015). Cette capacité à associer un nombre conséquent de protéines est vraisemblablement aussi vrai pour ANRIL. Ainsi, les études visant à identifier tous les facteurs protéiques associés à ce dernier ont grand intérêt.

ANRIL associe les complexes PcG et la protéine YY1 (Kotake *et al.*, 2011 ; Yap *et al.*, 2010 ; Zhou *et al.*, 2016). En complément, il interagit aussi avec WRD5 (WD-40 repeat-containing protein 5, dépôt des H3K4me3) et HDAC3 (histone déacétylase 3) pour favoriser leurs recrutements au sein de la chromatine et induire l'expression du gène NOX1 dans les cellules musculaires lisses isolées d'aortes humaines (Zhang *et al.*, 2020). Malgré ces partenaires avérés, le groupe a accumulé des évidences suggérant qu'ANRIL lie très probablement d'autres facteurs encore non identifiés. En effet, au cours de l'analyse de l'activité *-trans* d'ANRIL dans les cellules HEK293, le groupe a constaté que 123 gènes sur les 188 cibles directes sont régulés négativement par ANRIL et que seulement 17% d'entre eux sont susceptibles d'être régulés par un mécanisme dépendant de la présence des H3K27me3 (Alfeghaly *et al.*, 2021). Cette observation suggère donc l'implication de mécanismes indépendants des PcG engagés par ANRIL pour réduire l'expression génique. En faveur de cette hypothèse, des études mettent en évidence que des lncRNAs sont capables de réprimer l'expression génique en recrutant plusieurs catégories de complexes protéiques. C'est notamment le cas de HOTAIR qui peut s'associer aux complexes PRC2 et CoREST responsables du dépôt de H3K27me3 et de la suppression de H3K4me1-2 au locus HOXD, respectivement (Rinn *et al.*, 2007 ; Tsai *et al.*, 2010).

Nous avons également identifié 65 cibles directes régulées positivement par ANRIL. Comme mentionné précédemment, une étude a montré qu'ANRIL interagit avec l'activateur transcriptionnel YY1 dans les cellules HUVEC (Zhou *et al.*, 2015). Or, nous avons trouvé que 63% des cibles primaires trans-activées par ANRIL sont susceptibles d'être liées par YY1 dans les cellules HEK293 (Alfeghaly *et al.*, 2021) ce qui semble suggérer encore un fois que les facteurs protéiques associés à ANRIL ne sont pas tous identifiés.

Afin d'évaluer la capacité d'ANRIL à associer des facteurs protéiques, nous avons réalisé des expériences de chromatographie d'affinité MS2-MBP en utilisant comme ARN tous les exons d'ANRIL individuellement (**Fig.1A**). Le fractionnement des contenus protéiques par SDS-PAGE révèle de multiples bandes témoignant de la présence de nombreuses protéines dans les RNP purifiés. Notons également que de nombreuses différences sont observées lors de la comparaison des profils ce qui suggère que plusieurs protéines sont enrichies dans les RNP formés par des exons distincts (indiquées par les flèches rouges et noires de la **Fig.1B**). A titre d'exemple, les expériences de

western-blot des fractions éluées montrent que les RNP formés avec les exons 1, 4, 19 et 21 sont enrichies pour la protéine hnRNPK.

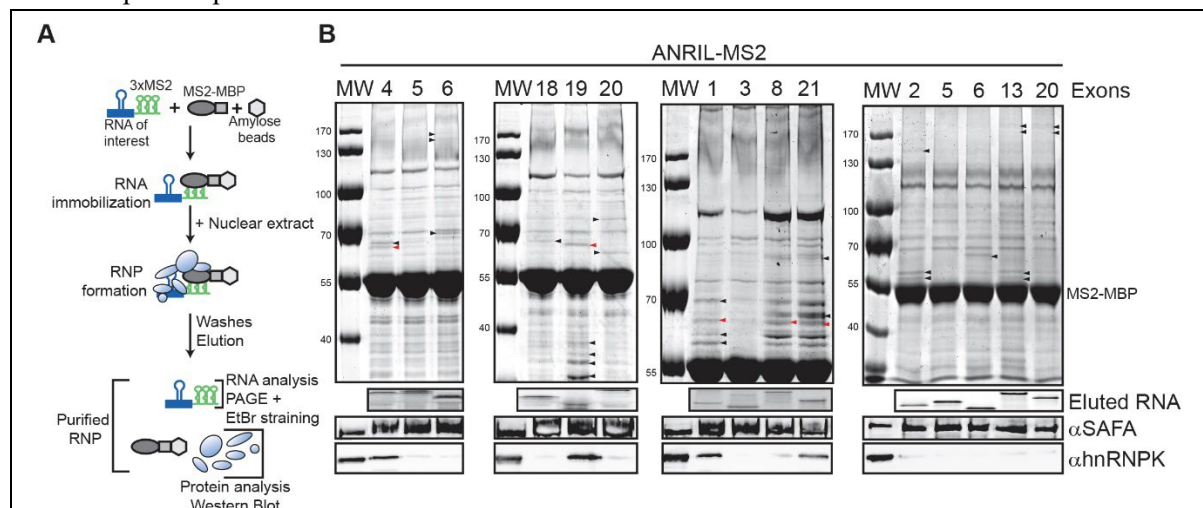


Figure 1 : Fractionnement du contenu protéique des RNP formés avec des exons d'ANRIL dans un extrait nucléaire de cellules HEK293 – Données non publiées. (A) Représentation schématique de la méthode de chromatographie d'affinité MS2-MBP. Un ARN portant l'étiquette MS2 est transcrit *in vitro* et immobilisé sur billes d'amylose par l'intermédiaire de la protéine MS2-MBP. Un extrait nucléaire est ajouté et les RNP sont purifiées et éluées. **(B)** Fractionnement sur gel de SDS-PAGE et analyse par WB des RNP formées avec les exons 1-6, 8, 13, 10-21 d'ANRIL (n=2).

Pour toutes ces raisons, il est séduisant de penser qu'un certain nombre des facteurs protéiques d'ANRIL reste à identifier, problématique à laquelle nous proposons de répondre dans cette partie du projet.

1.2. Identification des facteurs protéiques associés à ANRIL par la méthode CARPID

Bien que convaincantes, les expériences de chromatographie MS2-MBP présentées ci-avant ont le désavantage d'informer sur le contenu protéique de RNP formés non pas avec ANRIL pleine taille mais avec des sous-fragments et ceci en présence d'un extrait nucléaire. Ces conditions *in vitro* sont donc susceptibles de ne pas être totalement représentatives du contexte cellulaire. Afin de pallier cette limitation, nous envisageons d'utiliser une méthode nommée CARPID qui repose sur l'emploi du système CRISPR/Cas dont la validité a été démontrée notamment dans le cadre de l'étude du lncRNA Xist (Yi *et al.*, 2020). Rappelons que le système CRISPR/Cas est un des moyens mis en œuvre par les bactéries pour lutter contre les attaques pathogéniques en catalysant le clivage des molécules d'ADN ou ARN viral. Ce système est actuellement largement utilisé en biotechnologie car il offre notamment la possibilité d'éditer finement le génome.

La méthode CARPID est basée sur l'utilisation d'une forme modifiée de la CasRx, une nucléase membre du système CRISPR de type VI-D capable de cibler et cliver les molécules d'ARN. Dans cette méthode, un ARN guide est exprimé *in cellulo* pour cibler l'ARN endogène. Plusieurs ARN guides seront utilisés pour couvrir la séquence de l'isoforme d'ANRIL le plus long. L'expression *in cellulo* de ces ARN guides sera réalisée conjointement avec l'expression de la BASU-dCasRx (vecteur n°153209, Addgene). Suite à une étape d'incubation des cellules à 37°C pendant 15 min en présence de biotine, BASU-dCasRx catalyse la biotinylation des protéines associées à l'ARN d'intérêt. Les protéines ainsi étiquetées sont extraites puis capturées sur billes de streptavidine et analysées par Western-blot ou spectrométrie de masse (**Fig.2**). L'avantage majeur qu'offre cette méthode réside notamment dans la possibilité d'utiliser plusieurs ARN guides, ce qui doit permettre d'augmenter la spécificité de ciblage et diminuer le bruit de fond.

L'interactomique d'ANRIL sera analysée par spectrométrie de masse quantitative en collaboration avec la plateforme UMS2008/US40 IBSLor (Nancy Dir : Pr. I. Motorine).

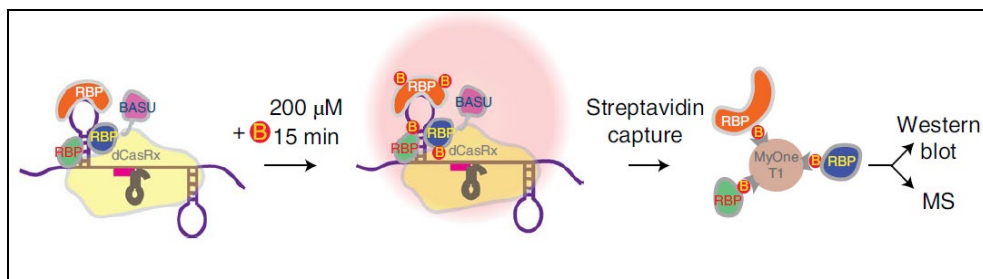


Figure 2 : La méthode de CARPID identifie les protéines associées aux lncRNA *in cellulo*. La biotine est représentée par des cercles rouges marqués par 'B' (pour les détails se référer au texte).

Il n'est cependant pas exclu que nous rencontrions des difficultés expérimentales inattendues, raison pour laquelle nous procéderons en parallèle à l'identification des partenaires protéiques d'ANRIL notamment par la méthode de chromatographie d'affinité MS2-MBP détaillée dans la partie I.1. (Maenner *et al.*, 2013 ; Sanchez et Maenner, 2021).

1.3. Hiérarchisation des partenaires protéiques les plus pertinents sur la base des données bibliographiques et celles générées par le groupe

Après l'identification exhaustive des partenaires protéiques d'ANRIL, nous procéderons à leur hiérarchisation basée sur des critères bibliographiques et sur les données générées par le groupe afin d'établir des liens entre les candidats, leurs caractéristiques fonctionnelles et celles d'ANRIL. Nous nous efforcerons de pleinement annoter la liste des candidats notamment par recherches des activités, d'ontologie, de données de localisation subcellulaires, de domaines fonctionnels ou de fixations aux ARN, d'association à des pathologies en lien avec ANRIL...

Nous porterons une attention toute particulière aux facteurs protéiques importants pour les événements de modulation de l'expression génique impliquant :

- les modificateurs de la chromatine
- les facteurs trans impliqués dans la modulation de l'épissage alternatif
- les modulateurs de la topologie nucléaire

Vous pourrez apprécier dans les lignes suivantes les éléments nous ayant conduit à orienter notre hiérarchisation de cette manière.

1.3.1. Les modificateurs de la chromatine comme partenaires pertinents d'ANRIL

Nous porterons en effet une attention toute particulière aux protéines impliquées dans la modulation du paysage chromatinien. Sachant pertinemment que les notions relatives à la chromatine ont été évoquées en filigrane tout au long de la partie « Rapport d'activité », il m'a semblé tout de même nécessaire d'explicitier brièvement ses propriétés et les facteurs qui conditionnent ces dernières afin d'assurer la compréhension parfaite des motivations de cet axe.

La chromatine est constituée de nucléosomes formés par un fragment ADNdb de 147 pb enroulé autour d'un octamère d'histone composé de deux copies de chacune des 4 protéines d'histone : H2A, H2B, H3 et H4. La protéine histone de liaison H1, s'associe à l'ADN situé aux sites d'entrée/sortie du nucléosome et est impliquée dans les structures chromatinienne d'ordre supérieur (pour une revue, Szerlong et Hansen, 2011). L'espacement des nucléosomes est très dynamique et conditionne l'expression génique. Ces derniers peuvent se « déplacer » le long de la molécule d'ADN et sont soumis à des modifications post-traductionnelles des histones comme, pour ne citer qu'elles, la méthylation (catalysée par les HMT i.e Histone Méthyle Transférases) et l'acétylation (catalysée par les HAT i.e Histone Acétyle Transférases). Ces modifications sont déposées pour la grande majorité au niveau des extrémités N-terminales des queues d'histones, portions non structurées accessibles aux enzymes de modification (pour revue Cutter et Hayes, 2015). Notons que ces dernières sont susceptibles d'être supprimées par l'intervention de HDM (Histone déméthylases) et de HDAC (Histone désacétylases).

Les modifications d'histones influencent les interactions internucléosomiques, le recrutement de facteurs protéiques et concourent à la génération de 2 états chromatiniens : condensé et relaxé appelés respectivement hétérochromatine et euchromatine (pour une revue, Black et Whetstone, 2011). L'hétérochromatine transcriptionnellement inactive est caractérisée par exemple, par la présence de marques répressives telles que les H3K27me3 et H3K9me3. Ces marques sont catalysées par deux entités protéiques : le complexe répresseur Polycomb PRC2 et les HMT SETDB1 / SUV39H1-2 respectivement (**Fig.3**). L'état euchromatinien quant à lui est défini par exemple, par la présence de H3K4me3/H3K9ac qui sont fréquemment présentes au niveau des TSS/promoteurs des gènes actifs alors que le corps de ces derniers sera plutôt enrichi en H3K36me3 (pour revue, Huang et Zhu, 2018). Ainsi, les HATs telles que CPB/p300 sont responsables du dépôt d'acétylation tandis que les HMT comme MLL1-2 contrôlent le statut de méthylation de type H3K4me3 (pour revue, Collins *et al.*, 2019 ; Ernst et Vakoc, 2012). D'autres régions génomiques sont également caractérisées par un paysage chromatinien particulier. C'est le cas des régions dites enhancers, éléments ADN qui régulent positivement la transcription indépendamment de la distance génomique et de l'orientation par rapport au promoteur du gène cible. Ces régions sont capables de favoriser l'augmentation de la concentration locale de la machinerie de transcription d'où leur rôle positif dans la régulation de l'expression génique (pour revue, Pennacchio *et al.*, 2013). Des enrichissements incluant H3K4me1 et H3K27ac, sont observés au niveau des enhancers dits actifs (pour revue Gates *et al.*, 2017 ; Lim *et al.*, 2010) (**Fig.3**).

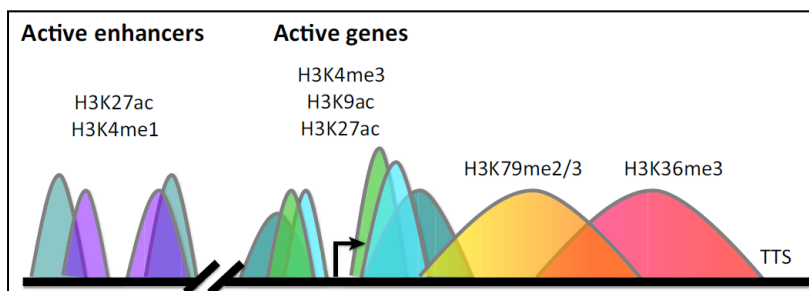


Figure 3 : Localisation génomique des modifications d'histone associées à la transcription active. Illustration simplifiée de l'occupation des modifications d'histone définie par expériences de ChIP-seq. Les régions en couleur représentent les signaux typiques au niveau d'un gène eucaryote transcriptionnellement actif. H3K4me3 et H3K9ac sont associés aux régions promotrices des gènes activement transcrits, tandis que H3K36me3 et H3K79me3 sont localisées dans les corps des gènes transcrits activement. H3K27ac se localise à la fois aux promoteurs actifs des gènes et aux régions enhancer, et H3K4me1 est principalement enrichie au niveau des enhancers. En revanche, les gènes réprimés présentent différentes marques de méthylation et ne présentent pas d'acétylation de l'histone (non montré). La flèche indique le site de début de la transcription (TSS). Abréviations : NDR, région déplétées en nucléosomes ; TTS, site de terminaison de la transcription (d'après Gates *et al.*, 2017).

L'ADN est également le siège de modification comme la méthylation. Il s'agit de l'addition d'un groupement méthyle sur le 5^{ème} atome de carbone du cycle pyrimidine de la cytosine qui est suivi d'un résidu guanine (CpG). Cette marque, lorsqu'elle est présente au niveau des régions promotrices, corrèle avec un état transcriptionnellement réprimé (pour revue, Jang *et al.*, 2017). Les ADN méthyltransférases (DNMTs) telles que DNMT1, DNMT3a et DNMT3b sont responsables de cette modification.

Comme évoqué précédemment, le dépôt des modifications des histones et de l'ADN est un phénomène réversible. La suppression active de ces dernières est donc un événement essentiel dans la gestion de la conversion de la chromatine d'un état transcriptionnellement inactif à actif et réciproquement. Pour ne citer qu'elles, les protéines de type « erasers » comme les protéines JHDM2/KDM3 et celles de la famille TET sont responsables de la déméthylation des histones (H3K9me3) et de l'ADN (CpG) respectivement. Notons également qu'il existe une interconnexion étroite entre H3K9me3 et méthylation de l'ADN. En effet, l'association de la protéine UHRF1 à la chromatine enrichie en H3K9me3 est nécessaire à la stabilité de la DNMT1 et au maintien de la méthylation de l'ADN (Rothbart *et al.*, 2014).

De manière intéressante, plusieurs lncRNAs sont capables de réguler l'expression génique en modulant les propriétés de la chromatine par leurs capacités à interagir avec des modificateurs de la chromatine. Outre ANRIL, citons à titre d'exemple le lncRNA Xist qui favorise le recrutement du complexe répressif Polycomb (PcG) (Sarma *et al.*, 2014), le lncRNA Fendrr qui associe le complexe MLL/TrxG

(Grote *et al.*, 2013), le lncRNA HOTAIR qui associe l'histone déméthylase LSD1 (Tsai *et al.*, 2010), le lncRNA ecCEBPA qui associe l'ADN méthyltransférase DNMT1 (Di Ruscio *et al.*, 2013) et le lncRNA GADD45a la protéine TET1 qui module le degré de méthylation des îlots CpG (Arab *et al.*, 2019). Comme déjà évoqué, ANRIL associe le complexe PcG, les protéines WRD5 et HDAC3 (Kotake *et al.*, 2011 ; Yap *et al.*, 2010 ; Zhang *et al.*, 2020). Malgré ces partenaires avérés, l'équipe a montré notamment que l'extrémité 3' d'ANRIL (Exon terminal n°21) est susceptible de moduler le dépôt des marques H3K27me3 mais aussi celui d'autres marques jamais décrites comme étant dépendante d'ANRIL (données non publiées). En effet, dans le cadre de l'évaluation de l'importance de l'exon 21 dans les fonctions d'ANRIL, nous avons généré une lignée cellulaire (3 clones) portant une délétion homozygote de cet exon par l'approche CRISPR/Cas9 (**Fig.4**).

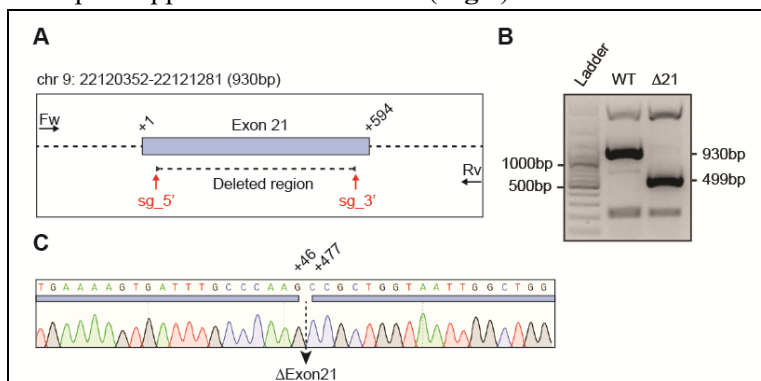


Figure 4 : Génération de la lignée cellulaire HEK293 portant la délétion de l'exon 21 d'ANRIL - données non publiées. (A) Représentation schématique de la région délétée de l'exon 21. La position des guides (ARNg) utilisés pour cibler l'exon 21 est présentée ainsi que le début et la fin de l'exon. (B) Exemple représentatif du criblage par PCR sur l'ADN génomique de clones WT ou Δ Exon21. Les tailles attendues pour les produits PCR correspondant au WT et Δ Exon21 sont respectivement 930 pb et 499 pb. (C) Chromatogramme obtenu après séquençage du produit PCR correspondant au Δ Exon21 confirmant la délétion de la région comprise entre les positions +46 et +477.

L'analyse du transcriptome de cette lignée par microarray a permis d'identifier 557 gènes dont l'expression est affectée par la délétion de l'exon 21 par rapport aux cellules sauvages : 213 présentent une augmentation et 344 une diminution de leur niveau d'expression (**Fig.5A**). Etant donné l'impact de la délétion de cette région de 432 pb sur l'expression génique, la présence d'éléments fonctionnels au sein de l'exon 21 semble évidente. Des expériences de Cut&Run ont donc été réalisées afin d'évaluer les variations éventuelles du paysage chromatinien induites par l'absence de l'exon 21 (**Fig. 5B**, cf. légende figure pour détails). Nous avons testé les différences de dépôt notamment des modifications H3K27me3 entre les lignées sauvage et délétée pour l'exon 21. La **Figure 5C** présente des exemples représentatifs de régions au sein desquelles la mise en place de ces modifications ne dépend aucunement de la présence de l'exon 21 (majorité des régions détectées). Néanmoins, nous avons pu identifier des régions pour lesquelles la présence de ces modifications est pleinement dépendante de ce dernier. En effet, ces analyses nous ont permis d'identifier 296 gènes pour lesquels une diminution de l'enrichissement des modifications H3K27me3 en absence de l'exon 21 est détectée (**Fig. 5D**). Ces résultats suggèrent donc fortement que l'exon 21 d'ANRIL intervient positivement (directement ou indirectement) dans l'activité des complexes PcG au niveau de ces loci. De manière plus surprenante, l'effet inverse pour d'autres modifications testées (nommées délibérément HxKx) est observé. **Le panneau F de la Figure 5** montre une région génomique représentative pour laquelle un enrichissement marqué pour HxKx est observé en absence de l'exon 21. Il est donc séduisant de penser que cet exon pourrait faciliter le recrutement de modificateurs d'histones jamais décrits jusqu'alors comme étant dépendant d'ANRIL, raison supplémentaire pour laquelle nous porterons une attention toute particulière à ces derniers lors de la hiérarchisation des facteurs protéiques identifiés par la méthode CARPID.

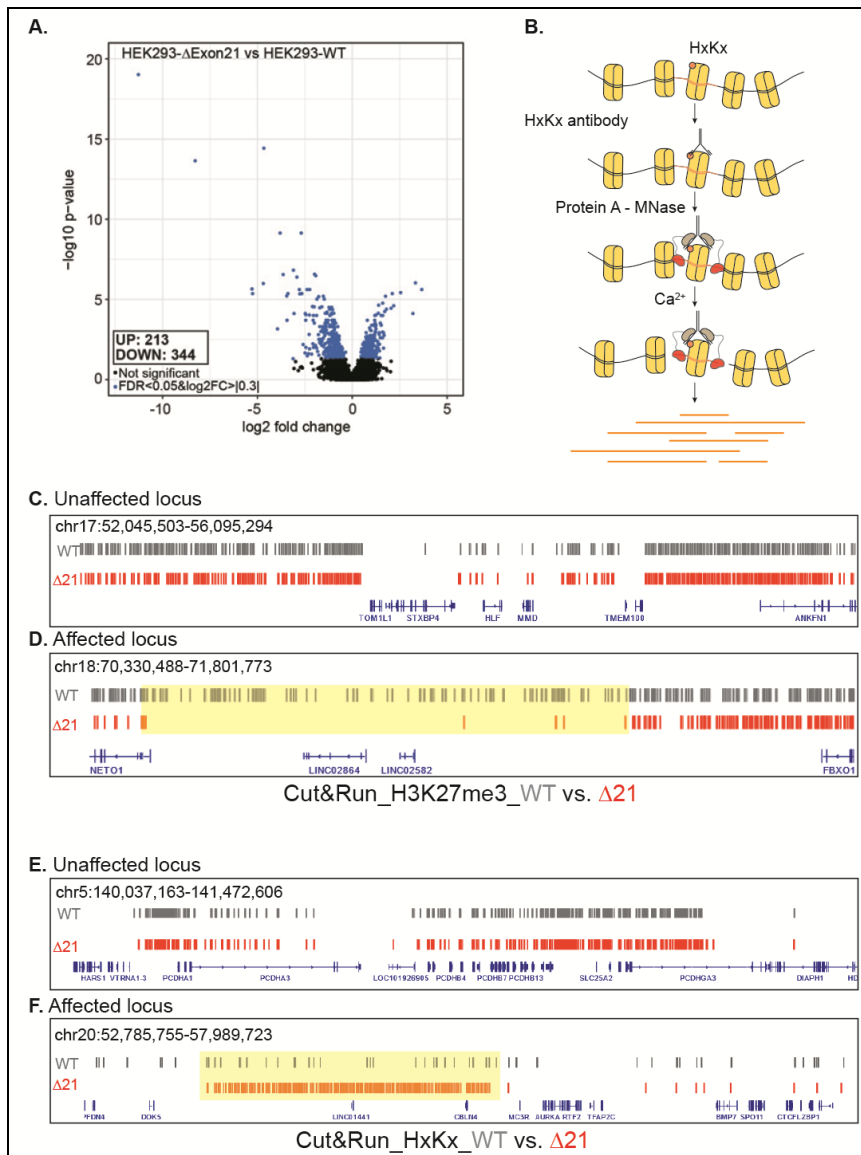


Figure 5 : L'exon 21 affecte l'expression génique et semble moduler le dépôt de modifications d'histone - Données non publiées. (A) Volcano plot représentant la significativité statistique vs "fold change" pour tous les gènes différentiellement exprimés (n=557, FDR < 0.05, log2FC > |0.3|) dans les cellules HEK293-ΔExon21 (n=5). (B) Représentation schématique de la méthode de Cut&Run utilisée pour l'identification des sites d'association génomique des modifications d'histone H3K27me3 ou autre (HxKx). Succinctement, les cellules sont immobilisées sur des billes magnétiques associées à la Concanavaleine A. Les cellules sont ensuite perméabilisées et incubées avec un anticorps dirigé contre H3K27me3 ou autres. La protéine de fusion MNase-Protéine A est ajoutée et s'associe à l'anticorps, la fragmentation ciblée des sites d'association reconnus par les anticorps est alors induite par ajout de Ca²⁺. Les fragments ainsi générés sont élués, convertis en banque et séquencés. (C) et (E) Exemples représentatifs de régions ADN au sein desquelles la mise en place de modifications H3K27me3 et HxKx n'est pas dépendante de l'exon21. (D) et (F) Exemples représentatifs de régions ADN au sein desquelles la mise en place de modifications H3K27me3 et HxKx est dépendante de l'exon21. Un rectangle jaune signale les régions modifiées différemment.

1.3.2 Les facteurs de régulation de l'épissage alternatif comme partenaires pertinents d'ANRIL

Nous porterons également une attention toute particulière aux protéines impliquées dans la modulation de l'épissage alternatif car l'équipe a généré des données suggérant qu'ANRIL est capable de moduler ce processus.

1.3.2.1. ANRIL, un modulateur potentiel de l'épissage alternatif

L'épissage est un événement essentiel à la maturation des ARNm au cours duquel les régions non codantes du transcrit, appelées introns, sont éliminées par excision alors que les exons restants sont reliés entre eux. L'excision des introns et la liaison des exons sont réalisées par un complexe appelé spliceosome, composé de protéines et de petits ARN nucléaires (snRNA) capable de catalyser deux réactions successives de transestérification (Saldanha *et al.*, 1993 ; De Conti *et al.*, 2013). La réaction d'épissage implique trois séquences spécifiques nécessaires au recrutement de la machinerie

d'épissage : le site d'épissage 5' (site donneur), le site d'épissage 3' (site accepteur) et le point de branchement. Ces séquences consensus se trouvent aux extrémités 5' et 3' des introns (donneur et accepteur) et le point de branchement en amont du site d'épissage 3' (**Fig.6**).

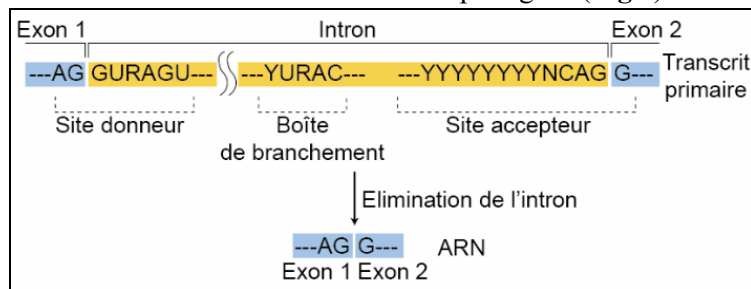


Figure 6 : Des séquences nucléotidiques consensus dans la molécule d'ARN définissent les extrémités introniques chez l'homme. Les trois blocs de séquences nucléotidiques indiqués sont nécessaires pour exclure une séquence intronique de l'ARN mature. Ici, A, G, U et C sont les nucléotides standard de l'ARN ; R représente les purines (A ou G) ; et Y représente les pyrimidines (C ou U). Seuls le GU au début de l'intron et l'AG à son extrémité sont des nucléotides invariants dans les séquences consensus d'épissage (adapté d'après Molecular Biology of the Cell, Sixth edition).

Il existe deux types d'épissage, l'épissage constitutif (EC) et l'épissage alternatif (EA). L'EC concerne les exons qui sont systématiquement insérés dans le transcrit mature final tandis que l'EA est un processus au cours duquel certains exons spécifiques peuvent être soit inclus, soit exclus du transcrit final. Parmi les 20 000 gènes codants présents dans le génome humain, environ 95% d'entre eux sont épissés alternativement ce qui augmente le nombre d'ARNm produits pour un même gène et, par conséquent, augmente le nombre de protéines traduites (Pan *et al.*, 2008 ; Black, 2003).

De manière intéressante, le groupe a découvert que le knock-down d'ANRIL dans les cellules HEK293 affecte le profil global d'EA avec 6936 événements recensés incluant des rétentions d'intron (introns présents dans le transcrit mature), l'utilisation différentielle de sites d'épissage 5' et 3' (exons plus courts ou plus longs selon l'utilisation de sites d'EA présents en 5' ou en 3') et des phénomènes de cassette d'exon (exons soit inclus ou exclus dans l'ARNm mature) (Alfeghaly *et al.*, 2021 et données non publiées). Comme mentionné précédemment, le pouvoir régulateur d'ANRIL peut affecter de façon directe l'expression de certains gènes (cibles primaires), le produit de ces gènes pouvant avoir à leur tour un impact sur l'expression de multiple gènes (cibles secondaires). Il est fort probable qu'il en soit de même pour la modulation de l'EA. Ainsi, le croisement des données d'occupation génomique d'ANRIL (ChIRP-seq) avec celles d'EA obtenues en condition de knock-down d'ANRIL dans les cellules HEK293 a mis en évidence que 24% des gènes contactés par ANRIL présentent une variation d'EA. Ceci représente 939 gènes parmi lesquels 78 ont une intensité d'événements remarquablement forte (Splicing index ≥ 10) (**Fig.7A**). Ces données suggèrent qu'ANRIL pourrait être un régulateur direct de l'EA pour un nombre non négligeable de gènes. La **Figure 7B-C** présente 2 exemples de modulation d'EA dépendantes d'ANRIL puisque sa présence favorise ou défavorise l'inclusion d'exons au sein des ARN MED7 (**panneau B**) et FMR1 respectivement (**panneau C**).

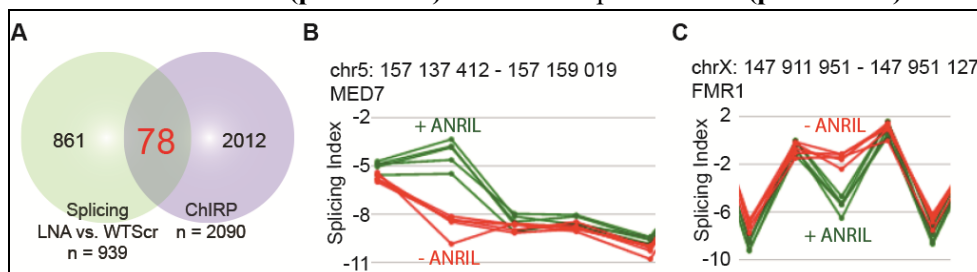


Figure 7 : ANRIL affecte l'épissage alternatif – Données non publiées. ANRIL semble affecter l'épissage alternatif – Résultats obtenus par analyse par microarray Affimetrix (ClariomD) d'ARN totaux extraits à partir de cellules HEK293 sauvages (vert) ou déplétées pour ANRIL (rouge) (n=5). Pour détails, se référer au texte.

1.3.2.2. LncRNA, facteurs trans et modulation de l'épissage alternatif

La combinatoire d'EA repose sur l'utilisation différentielle des sites d'épissage présents au sein de l'ARN pré-mature. Ceci dépend de la force du site lui-même qui corrèle positivement avec le degré de convergence par rapport à la séquence consensus et également du recrutement de facteurs régulateurs d'épissage (éléments régulateurs *trans*) au niveau de sites exoniques ou introniques (éléments

régulateurs *cis*). Les éléments régulateurs *trans* sont des RBP (RNA Binding Proteins). Les protéines SR et hnRNP sont les plus fréquentes de ces RBP. En s'associant aux ESE (Exonic Splicing Enhancer) et ISE (Intronic Splicing Enhancer), les protéines SR peuvent favoriser l'EA en contribuant à la formation et à la stabilisation des complexes d'épissage sur l'ARN pré-mature. En revanche, les protéines hnRNP liées à des ESS (Exonic Splicing Silencer) ou des ISS (Intronic Splicing Silencer) inhibent généralement la formation des complexes d'épissage en diminuant l'accessibilité des sites d'épissage ou en neutralisant les activités des protéines SR (Eperon *et al.*, 2000) (**Fig.8**).

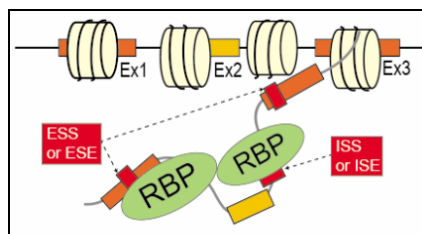


Figure 8 : Régulation de l'EA par les RBP. Adaptée d'après Baralle and Giudice, 2017.

De manière intéressante, le groupe a identifié la protéine hnRNPK comme partenaire privilégié d'ANRIL (données non montrées, travail de thèse d'Aymeric Sanchez, soutenance 12/2022). Il est donc envisageable qu'ANRIL par sa double capacité à lier à la fois la chromatine et des facteurs *trans* puisse aider au recrutement de ces derniers sur un ARN cible en cours de synthèse afin de moduler son EA.

1.3.2.3. LncRNA, paysage chromatinien et modulation de l'épissage alternatif

Il s'avère qu'il existe un lien très étroit entre les caractéristiques de la chromatine et l'EA. En effet, les nucléosomes *via* la combinatoire de leurs modifications d'histones peuvent influencer ce dernier (Segal *et al.*, 2006). A titre d'exemple, notons que les exons sont enrichis en modifications d'histones spécifiques telles que H3K27me2, H3K36me3 et H4K20me3 (Schwartz *et al.*, 2009 ; Andersson *et al.*, 2009 ; Tilgner *et al.*, 2009 ; Saint-André *et al.*, 2011 ; Spies *et al.*, 2009 ; Agirre *et al.*, 2021 ; Hon *et al.*, 2009). La présence de l'une ou l'autre de ces modifications peut affecter l'EA du fait de leurs capacités à lier des facteurs protéiques capables de favoriser le recrutement d'éléments régulateurs *trans* comme les protéines hnRNP ou SR, voire même de recruter directement des composants du spliceosome. Ainsi, il a été démontré qu'un groupe de gènes pour lequel l'EA dépend de la protéine PTB (hnRNP) est enrichi en H3K36me3 dans les régions alternativement épissées et que la modulation de ces marques d'histones est suffisante pour modifier les événements d'inclusion exonique dépendants de PTB. Le mécanisme moléculaire par lequel H3K36me3 agit dans ce cas implique la création d'une plateforme au sein de la chromatine facilitant le recrutement de PTB à l'ARN naissant par l'intermédiaire d'une protéine adaptatrice, MRG15, qui se lie spécifiquement aux H3K36me3. Lorsque les niveaux de H3K36me3 sont élevés au sein des régions alternativement épissées, MRG15 est efficacement recrutée et, en interagissant avec PTB, facilite son recrutement sur l'ARN naissant. Cette cascade d'événements défavorise l'inclusion des exons concernés au sein de l'ARN mature (Sanidas *et al.*, 2014 ; Iwamori *et al.*, 2016) (**Fig.9**). Selon un mécanisme analogue, la présence de H3K9me3 facilite le recrutement de la protéine HP1 qui à son tour facilite le recrutement de SRSF3 et conduit à la réduction de l'inclusion d'exons alternatifs (Yearim *et al.*, 2015) (**Fig.9**).

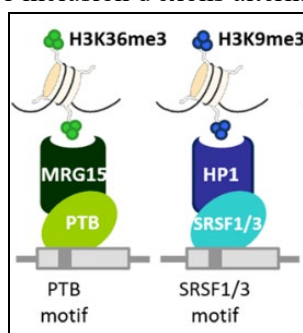


Figure 9 : Les modifications d'histones (exemples : H3K36me3 et H3K9me3) définissent de manière différentielle la chromatine hébergeant des séquences exoniques épissés alternativement ce qui module le recrutement de régulateurs d'épissage (exemples : PTB et SRSF1/3) sur l'ARN pré-messager par l'intermédiaire des complexes adaptateur de la

chromatine (exemples : MRG15 et HP1) (illustration adaptée à partir de celle proposée sur le site web du laboratoire de R. Luco).

Les changements du paysage chromatinien peuvent donc largement affecter l'EA et, en tant que partenaires privilégiés des modificateurs de la chromatine, les lncRNA comme ANRIL sont de très bons candidats pour jouer un rôle dans ces processus. Une étude a d'ailleurs révélé que le lncRNA asFGFR2, produit à partir du locus *FGFR2* humain, favorise l'EA de l'ARNm *FGFR2* en recrutant des modificateurs de la chromatine (Gonzales *et al.*, 2015). En effet, asFGFR2 recrute EZH2 et SUZ12 du complexe PRC2 et l'histone déméthylase KDM2a au locus *FGFR2*. Le domaine chromatinien ainsi façonné par asFGFR2 ($H3K36me2/3 \rightarrow H3K27me3$) limite la liaison du complexe adaptateur de chromatine MRG15-PTB à l'exon IIIb qui favorise finalement son inclusion au sein de l'ARNm *FGFR2*. De manière intéressante, l'isoforme *FGFR2* IIIb joue un rôle antitumoral dans le carcinome hépatocellulaire, ce qui pourrait ouvrir une nouvelle approche pour l'étude et le traitement de cette maladie très agressive (Amann *et al.*, 2010).

1.3.2.4. Hypothèse et modèle de travail pour l'implication d'ANRIL dans la modulation de l'EA

Basé sur les données bibliographiques et celles accumulées par le groupe, nous émettons l'hypothèse suivante : ANRIL module de façon directe l'EA en favorisant le recrutement de facteurs *trans* et de modificateurs de la chromatine au niveau d'au moins 78 gènes (Fig.10). Ce projet vise notamment à tester ce modèle.

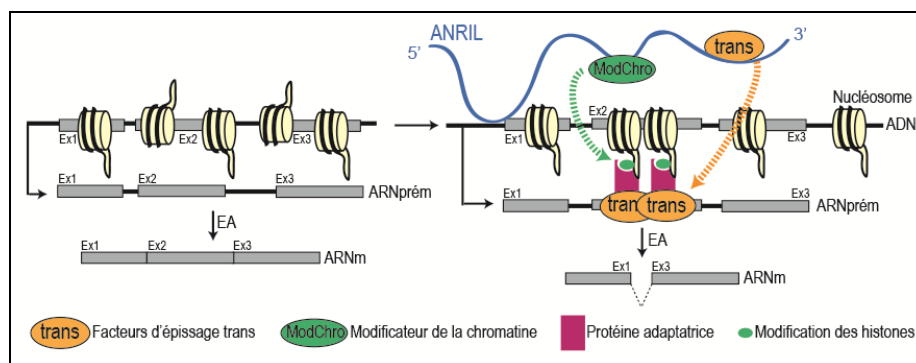


Figure 10 : Modèle de travail. ANRIL associe le génome au niveau de ses gènes cibles et favorise le recrutement des modificateurs de la chromatine. Des protéines adaptatrices lient les nouvelles modifications d'histone qui en résultent et recrutent alors à leur tour des facteurs *trans* au niveau de l'ARN pré-maturé en cours de transcription. Le recrutement de facteurs *trans* pourrait également s'opérer de façon directe via l'intervention d'ANRIL. Cette cascade d'événements module l'EA.

Pour toutes les raisons évoquées précédemment et avec pour objectif de hiérarchiser selon leurs degrés de pertinence les facteurs protéiques identifiés par la méthode CARPID, nous porterons une attention toute particulière aux facteurs protéiques de régulation de l'EA en *trans* (SR et hnRNP) mais également aux modificateurs de la chromatine.

1.3.3 Les facteurs de modulation de la topologie nucléaire comme partenaires pertinents d'ANRIL

Nous porterons également une attention toute particulière aux protéines impliquées dans la modulation de la topologie nucléaire. En effet, le groupe a généré des données suggérant qu'ANRIL serait susceptible de moduler ce processus. Notons que l'organisation 3D du génome au sein du noyau est essentielle à la régulation de l'expression génique et sa perturbation peut induire des processus pathologiques telles que les cancers (Bertero and Rosa-Garrido, 2021).

Les chromosomes se séparent en régions d'interactions préférentielles à longue portée qui forment deux types de chromatine qui s'excluent mutuellement appelés compartiments "A" et "B". Le compartiment A correspond à la chromatine riche en gènes et active, tandis que le compartiment B est principalement enrichi en chromatine répressive. À l'échelle de dizaines ou de centaines de kilobases, les chromosomes se replient en domaines avec des interactions intra-domaines préférentielles alors que leur fréquence d'association avec les régions voisines est drastiquement réduite (pour revue, Szabo *et al.*, 2019). Ces domaines sont communément appelés domaines d'association topologique (TAD) (Fig.11). De manière intéressante, les gènes ont tendance à être coréglés pendant la

différenciation cellulaire lorsqu'ils sont localisés dans le même TAD et les contacts entre les enhancers et les promoteurs de gènes sont principalement limités à l'intérieur des TAD, ce qui fait des TAD des domaines topologiques fonctionnellement organisés (pour revue, Szabo *et al.*, 2019). La formation des TAD et la régulation de l'architecture de la chromatine sont médiées par des complexes protéiques tels que le complexe CCCTC-binding factor (CTCF)/cohésine (pour revue, Song et Kim, 2017). CTCF est une protéine à doigt de zinc qui joue des rôles majeurs dans la répression et l'activation de la transcription par notamment sa capacité à moduler la formation des interactions promoteur-enhancer (pour revue, Dorsett et Ström, 2012 ; de Wit *et al.*, 2015).

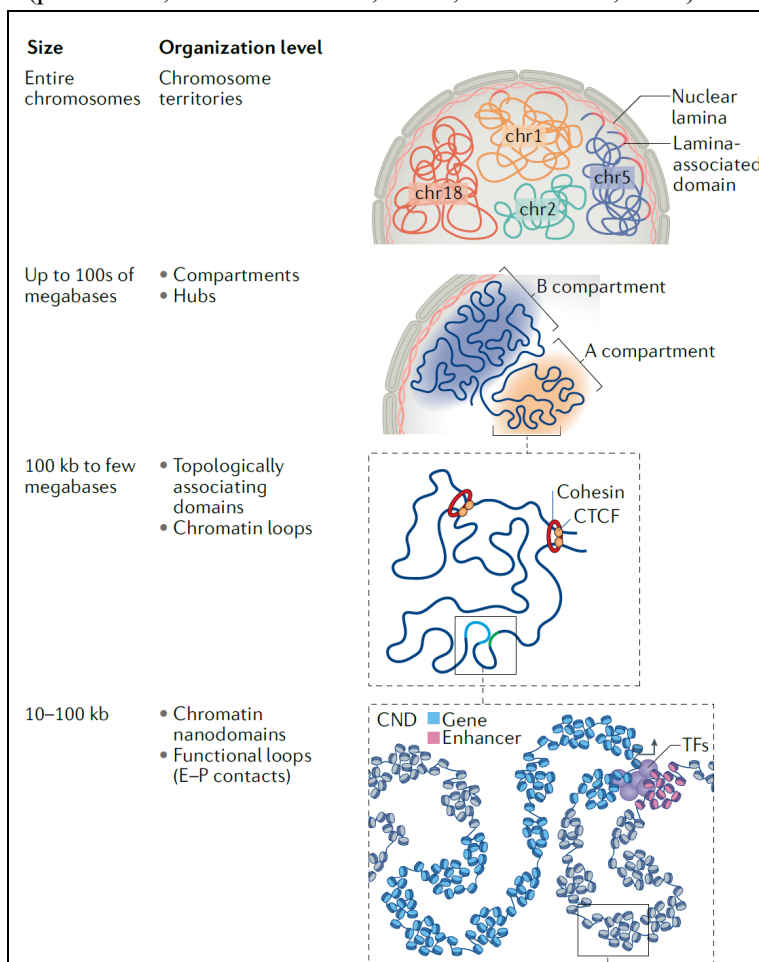


Figure 11 : Les multiples niveaux d'organisation du génome.

L'ADN se replie à plusieurs échelles (indiquées à gauche). L'ADN s'enroule autour des histones et forme les nucléosomes. Les nucléosomes forment des nanodomaines de chromatine (CND) d'une taille d'environ 100 kb, dans lesquels se trouvent la plupart des contacts entre enhancers et promoteurs (E-P). À l'échelle de ~1 Mb, les CND et les boucles de chromatine qui dépendent du facteur de liaison CCCTC (CTCF) - cohésine forment des domaines topologiquement associés (TADs). A plus grande échelle, jusqu'à des centaines de mégabases, la chromatine se sépare en compartiments actifs et inactifs (A et B, respectivement). A l'échelle du noyau, les chromosomes sont organisés en territoires chromosomiques (d'après Jerkovic et Cavalli, 2021).

De manière intéressante, il a été démontré que certains lncRNAs sont capables de moduler la topologie nucléaire en favorisant la mise en place d'interactions inter- et intrachromosomiques. Le lncRNA Firre est l'un des exemples les plus étudiés. Il est transcrit à partir du chromosome X et contient 12 exons répétés (pour une revue, Engreitz *et al.*, 2016). Des approches d'hybridation in situ ont révélé la présence de Firre à des loci autosomiques distaux sur les chromosomes 2, 9, 15 et 17 (Hacisuleyman *et al.*, 2014). Ces loci sont à proximité spatiale du locus Firre, phénomène qui a été attribué à la liaison de la protéine hnRNPU au lncRNA Firre (Hacisuleyman *et al.*, 2014 ; pour une revue, Engreitz *et al.*, 2016). Une étude a également montré l'implication de Firre dans le positionnement du chromosome X inactif près du nucléole et dans le maintien des marques H3K27me3 (Yang *et al.*, 2015).

Notons également que le lncRNA CCAT1-L est capable d'interagir avec CTCF ce qui favorise son recrutement au sein de la chromatine et induit la formation de boucles enhancer/promoteur au niveau du locus MYC dans les cellules HCT116 (Xiang *et al.*, 2014). Avec pour objectif d'évaluer si ANRIL est susceptible de jouer un rôle analogue, des expériences de purification des RNP par

chromatographie d'affinité ont été réalisées en utilisant les exons d'ANRIL et la présence de CTCF a été examinée par WB. Comme le montre la **Figure 12A**, ces expériences ont mis en évidence la présence significative de CTCF dans les RNP formées avec l'exon 21. Des expériences de Cut&Run ont été réalisées en utilisant un anticorps dirigé contre CTCF et les profils obtenus avec les lignées sauvages et déplétées pour l'exon 21 ont été comparés. L'analyse des résultats a pu mettre en évidence que l'absence de l'exon 21 affecte la localisation génomique de CTCF au niveau d'un nombre restreint de loci. La **Figure 12B** présente un exemple représentatif. L'ensemble de ces résultats suggèrent qu'ANRIL pourrait affecter la topologie nucléaire en facilitant le recrutement de CTCF au sein de la chromatine.

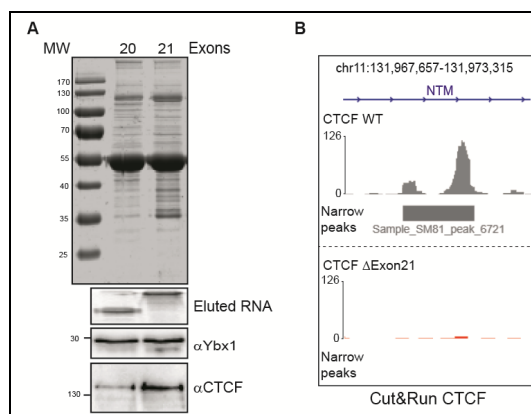


Figure 12 : La localisation génomique de CTCF au niveau d'un nombre restreint de loci semble dépendre de la présence de l'exon 21 d'ANRIL.

(A) Fractionnement sur gel SDS-PAGE et analyse par WB des RNP formées avec les exons 20 et 21 d'ANRIL. La présence des protéines Ybx1 et CTCF ont été détectées. (B) Exemple représentatif d'une région génomique au sein de laquelle l'occupation de CTCF disparaît dans les lignées déplétées pour l'exon 21 d'ANRIL.

Pour toutes les raisons évoquées précédemment, nous porterons une attention toute particulière aux facteurs protéiques susceptibles de moduler la topologie nucléaire parmi ceux identifiés par la méthode CARPID.

1.4. Détection/validation des associations ANRIL/protéines candidates

Afin de tester la capacité d'ANRIL à associer les protéines candidates *in cellulo*, nous appliquerons à nouveau l'approche CARPID suivie par l'évaluation de la présence de ces dernières dans les RNP formées avec ANRIL endogène par des expériences de Western-blot. Nous réaliserons également des expériences d'immunoprécipitation des ARN (RIP) avec ou sans pontage aux UV (CLIP) afin de tester la présence des facteurs protéiques sélectionnés dans les RNPs formés par ANRIL mais aussi afin d'appréhender le caractère direct des associations. Notons que cette approche a déjà été déployée par le groupe pour la caractérisation de la RNP [ANRIL/hnRNPK] (cf. rapport d'activité). Cette stratégie devrait donc nous permettre de dresser avec un degré de confiance non négligeable une liste des facteurs protéiques associés à ANRIL *in cellulo*.

1.5. Evaluation de l'effet de l'absence des protéines candidates sur les activités d'ANRIL

La contribution fonctionnelle des partenaires protéiques d'ANRIL sera ensuite évaluée. Pour cela, nous générerons des lignées invalidées pour les gènes codant les candidats par édition CRISPR/Cas9 dans les cellules HEK293, stratégie que nous avons déjà déployée notamment pour la caractérisation de la RNP [ANRIL/hnRNPK]. L'utilisation de ces lignées KO devrait nous permettre de quantifier l'effet de l'absence de ces protéines sur les activités d'ANRIL. Pour cela, des analyses transcriptomiques par RNAseq (nanopore PromethION 2 SOLO) seront réalisées afin d'identifier les gènes dont l'expression ou le profil d'EA sont affectés en absence de cette protéine. L'intersection de ces données avec la liste des cibles primaires d'ANRIL que nous avons identifié dans Alfeghaly *et al.*, 2021, permettra de proposer avec un degré de confiance raisonnable la liste des gènes contactés par ANRIL pour lesquelles l'expression et/ou l'EA dépendent conjointement d'ANRIL et de la protéine candidate.

II. Identification des régions ARN d'ANRIL responsables de la fixation des facteurs protéiques

II.1. Arguments en faveur de cette problématique

Les analyses du transcriptome menées par l'équipe sur des lignées délétées pour les exons 8 et 21 d'ANRIL (CRISPRing) identifient seulement approx. 20% de gènes communs pour lesquels l'expression est affectée par rapport à la lignée sauvage (40 gènes communs parmi 208 et 159 pour $\Delta 8$ et $\Delta 21$ respectivement, données non publiées). Ces données suggèrent fortement que les exons 8 et 21 sont porteurs d'activités différentes. Rappelons que l'isoforme le plus long d'ANRIL mesure 3837-nts et compte 19 exons, autant d'exons capables de porter des activités distinctes. A l'instar des ARN Xist, HOTAIR et roX, ANRIL serait susceptible de posséder plusieurs modules permettant d'associer spécifiquement des facteurs protéiques supports de multiples activités.

La nature de ses déterminants ARN reste à ce jour peu documentée. Notons que l'analyse de la répartition des éléments transposables (ET) au sein des exons d'ANRIL a mis en évidence un enrichissement de l'exon 21 en ET soit 54,2 % de SINE/MIR et 12,6 % de répétitions simples, soit 66,8% de la séquence composant l'exon 21 (**Fig.13**). De manière intéressante, plusieurs études ont révélé que 33 % des séquences codant les lncRNAs sont dérivées d'ET ce qui pourrait suggérer que ces éléments représentent des domaines importants pour les activités des lncRNAs (Kelley and Rinn, 2012 ; Kapusta *et al.*, 2013 ; Carlevaro-Fita *et al.*, 2019). Il a d'ailleurs été démontré qu'ils interviennent dans la localisation ou encore la régulation de l'expression des lncRNAs (Yao *et al.*, 2021 ; Lubelsky *et al.*, 2021 ; pour revue, Fort *et al.*, 2020). Dans cette ligne de pensée, le groupe a notamment mis en évidence l'importance d'ET de type ERVL dans les processus de reconnaissance de la chromatine engagés par ANRIL (Alfeghaly *et al.*, 2021).

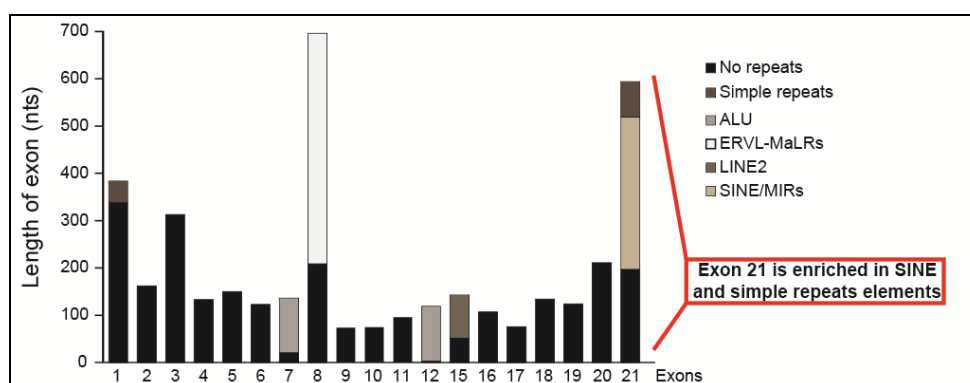


Figure 13 : Analyse de la distribution des éléments transposables au sein des exons d'ANRIL par RepeatMasker.

Il est donc vraisemblable que les ET soient susceptibles d'apporter un gain fonctionnel aux lncRNAs qui les portent. Comprendre leurs contributions dans les fonctions d'ANRIL notamment dans leurs propensions à lier des facteurs protéiques identifiés par CARPID nous paraît donc pertinent (sans pour autant exclure celles impliquant des régions ARN dénuées d'ET). La suite du projet vise notamment à adresser à cette problématique.

II.2. Evaluation de la présence des protéines candidates dans les RNP formés avec les sous-fragments d'ANRIL par chromatographie d'affinité MS2-MBP

En premier lieu, nous réaliserons des expériences de purification des RNP par chromatographie d'affinité MS2-MBP (Sanchez and Maenner, 2021). Nous utiliserons comme ARN d'intérêt des fragments sous-fragments d'ANRIL pour purifier les RNP *in vitro* suivi de l'évaluation de la présence des facteurs protéiques candidats par des expériences de Western-blot. Ces expériences nous permettront d'appréhender les associations ARN/protéine d'intérêt *in vitro* et d'identifier les séquences ARN susceptibles d'être responsables de l'association des protéines identifiées ci-avant.

II.3. Génération des lignées déplétées pour les régions ARN identifiées et évaluation de la fixation des facteurs protéiques aux formes tronquées d'ANRIL

Nous envisageons ensuite de tester la capacité de ces régions ARN à lier les protéines candidates *in cellulo* en générant des lignées cellulaires HEK293 déplétées pour ces dernières par l'approche CRISPR/Cas9. Cette méthode a déjà été utilisée par le groupe pour la génération des lignées déplétées pour les exons 1, 4, 8, 12 et 21 d'ANRIL (Alfeghaly *et al.*, 2021 et données non publiées). Des expériences d'immunoprécipitation des ARN (RIP/CLIP) avec un anticorps dirigé contre la protéine d'intérêt dans les lignées mutées permettront d'évaluer l'impact de la délétion sur la fixation des protéines candidates et donc de tester l'importance *in cellulo* des régions ARN d'ANRIL testées positives pour la fixation de ces protéines *in vitro*. En complément, nous déploierons la stratégie adoptée pour CTCF dans le cas des protéines candidates connues pour associer la fraction chromatinienne (cf **Fig.12**, partie I.3.3). Ainsi, des expériences de Cut&Run seront réalisées en utilisant des anticorps dirigés contre ces protéines et la comparaison des profils des lignées sauvages et tronquées permettra d'évaluer l'effet de l'absence de la région ARN sur la localisation génomique de ces protéines candidates.

II.4. Caractérisation des activités des RNP [protéines/ANRIL]

Grâce aux lignées cellulaires générées comme décrit précédemment, nous proposons d'évaluer l'absence des protéines identifiées et celle des régions ARN responsables de leurs associations dans les événements de modulation génique dépendants d'ANRIL. Par comparaison des lignées sauvage et mutées, nous examinerons les variations du transcriptome (RNAseq nanopore) aussi bien en termes d'expression génique que d'EA. L'intersection de ces données avec la liste des cibles primaires d'ANRIL (Alfeghaly *et al.*, 2021) devrait permettre d'identifier les gènes dont l'expression/l'EA est modulée de manière conjointe par une région spécifique d'ANRIL et sa protéine associée.

Des analyses d'occupation génomique des modifications d'histones par des expériences de Cut&Run (choisies en fonction des propriétés fonctionnelles des protéines identifiées) seront réalisées afin d'évaluer l'effet de l'absence de la région ARN d'ANRIL sur la mise en place de ces dernières. Des corrélations pourront donc être réalisées afin de comprendre l'interconnexion existant entre les changements du paysage chromatinien et la modulation de l'expression génique et de l'EA.

Etant donné que la présence de modifications d'histone comme les H3K9me3 conditionne notamment celle des méthylations de l'ADN, la détection des m5C sera également réalisée par des expériences d'immunoprécipitation de l'ADN méthylé (MeDIP) en utilisant des anticorps dirigés contre les 5-méthylCytosine. Si des variations de méthylation de l'ADN sont observées, nous tenterons d'évaluer leurs impacts sur l'expression génique et l'EA. Ainsi, le croisement des cibles primaires d'ANRIL, des données de transcriptomiques et celles générées par les expériences de Cut&Run/MeDIP devrait nous permettre de corrélérer les modifications de l'épigénome dépendantes conjointement d'ANRIL et des protéines identifiées avec les effets observés sur l'expression génique et l'EA.

Dans le cas où des protéines identifiées par la méthode CARPID ont des rôles décrits dans la modulation de la topologie nucléaire (CTCF, Rad21, cohésine, laminB1...), nous envisageons de réaliser des expériences de capture de conformation chromosomique HiC. Il s'agit d'une méthode qui permet de détecter de manière exhaustive les interactions chromatiniennes au sein du noyau. Brièvement, cette méthode est basée sur la capture de la conformation des chromosomes par pontage de la chromatine à la formaldéhyde suivi d'un fractionnement par digestion enzymatique, d'un couplage des extrémités des fragments ADN avec la biotine puis d'une ligation par proximité de manière à ce que seuls les fragments d'ADN liés de manière covalente forment des produits de ligation. Après isolement de ces fragments d'ADN chimériques par capture sur billes de streptavidine, ces derniers sont séquencés et l'analyse permet de déduire les informations sur leurs origines séquentielles et par conséquent sur leurs localisations physiques dans l'organisation 3D du génome (Belton *et al.*, 2012). La comparaison des fréquences d'association entre les matrices obtenues avec les lignées sauvages et tronquées nous permettra d'apprécier l'effet de l'absence des régions ARN étudiées, leurs protéines partenaires et ceci en lien avec la topologie nucléaire. Les ruptures ou les formations des boucles enhancer/promoteur feront l'objet d'une attention toute particulière étant donné

leurs importances directes dans la modulation de l'état transcriptionnel. Notons que des expériences préliminaires d'évaluation de l'implication d'ANRIL sur la modulation de la topologie nucléaire des cellules HEK293 sont actuellement en cours et ceci en collaboration avec les équipes dirigées par G. Barreto (IMoPA, Nancy) et G. Dobрева, (Université de Mannheim, Allemagne).

II.5. Etude des relations Structure/fonction entre les régions ARN et les protéines partenaires d'ANRIL

La fonction d'un ARN étant connue pour dépendre notamment de sa structure secondaire (2D), nous prévoyons de réaliser l'analyse expérimentale (inédiée) de la structure 2D des éléments fonctionnels d'ANRIL identifiés ci-dessus. A cette fin, l'utilisation d'une version adaptée du ChIRP permettant de capturer *in cellulo* des motifs structuraux spécifiques de l'ARN d'intérêt en complément d'expériences de SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) (idéalement par nanopore, nanoSHAPE, Stephenson et al., 2022) nous permettra d'accumuler des données expérimentales que nous utiliserons comme contraintes pour construire un modèle de structure 2D pour ANRIL (Maenner *et al.*, 2013 ; Low et Weeks, 2010 ; Smola *et al.*, 2015). Il est à noter que la mise en place de la méthode SHAPE vise à être intégrée de manière durable au sein du groupe. Cette démarche nécessitera l'expertise d'un laboratoire tiers, avec lequel des contacts ont déjà été établis pour définir les bases d'une collaboration future. Les données recueillies devraient nous permettre de caractériser en termes de structures 2D les éléments ARN critiques pour les activités d'ANRIL. La fonction des structures 2D identifiées sera élucidée en générant des lignées tronquées/mutées pour les structures 2D identifiées ci-avant par édition du génome CRISPRCas9 dans les HEK293. Les effets des formes mutées d'ANRIL sur sa capacité à lier ses partenaires protéiques (RIP, CLIP, Cut&Run) et à induire ses activités régulatrices (analyse transcriptomique, Cut&Run, ChIRP) seront évaluées par la mise en place de la stratégie explicitée dans la partie 1.5.

III. Extrapolation des données générées avec la lignée HEK293 à d'autres lignées cellulaires et recherche de corrélations avec l'état pathologique

Notons que les parties I. et II. du projet sont basées sur l'utilisation des cellules HEK293. Ces cellules embryonnaires humaines de rein en plus d'être relativement simples à manipuler, sont largement utilisées dans le domaine de recherche relatif à ANRIL. Toutefois, l'utilisation de modèles cellulaires plus en lien avec les pathologies associées à ANRIL est une étape incontournable notamment dans le but de mieux comprendre les effets d'ANRIL sur l'apparition/renforcement pathologique à l'échelle moléculaire. Dans cet axe, nous envisageons de mettre à l'épreuve l'existence des mécanismes d'action identifiés pour ANRIL dans les cellules HEK293 dans d'autres lignées cellulaires.

Afin de définir les lignées cellulaires les plus pertinentes, nous avons entrepris dans un premier temps l'évaluation du taux d'expression d'ANRIL (tout isoformes confondus) par RTqPCR dans 66 lignées isolées à partir de tissus tumoraux comme les cancers du sein, du pancréas, du poumon et du colon (**Fig.14**). Ces expériences montrent qu'ANRIL est exprimé dans la majorité des lignées analysées et qu'une hétérogénéité marquée est observée aussi bien parmi les différents cancers que parmi les différentes lignées appartenant au même tissu tumoral et ceci conformément à la littérature (Sanchez et al., 2023). A titre d'exemple pour le cancer d'origine pulmonaire, les cellules NCI-H149 présentent un niveau d'expression d'ANRIL de x4700 relatif à la lignée A549 (n'exprimant que très faiblement ANRIL). Notons que les cellules NCI-H149 sont décrites comme étant plus agressives que les cellules A549, SCLC (small cell lung cancer) vs. NSLSC (non-small cell lung cancer) respectivement, observation en faveur de la corrélation positive existant entre abondance d'ANRIL et sévérité tumorale.

Bibliographie

1. Agirre, E., Oldfield, A.J., Bellora, N., Segelle, A., and Luco, R.F. (2021). Splicing-associated chromatin signatures: a combinatorial and position-dependent role for histone marks in splicing definition. *Nat Commun* *12*, 682. 10.1038/s41467-021-20979-x.
2. Alfeghaly, C., Aigueperse, C., Maenner, S., and Behm-Ansmant, I. (2021). Non-Coding RNA Silencing in Mammalian Cells by Antisense LNA GapmeRs Transfection. *Methods Mol Biol* *2300*, 31–37. 10.1007/978-1-0716-1386-3_4.
3. Alfeghaly, C., Sanchez, A., Rouget, R., Thuillier, G., Igel-Bourguignon, V., Marchand, V., Branlant, C., Motorin, Y., Behm-Ansmant, I., and Maenner, S. (2021). Implication of repeat insertion domains in the trans-activity of the long non-coding RNA ANRIL. *Nucleic Acids Res* *49*, 4954–4970. 10.1093/nar/gkab245.
4. Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., and Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res* *19*, 1732–1741. 10.1101/gr.092353.109.
5. Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., and Niehrs, C. (2019). GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat Genet* *51*, 217–223. 10.1038/s41588-018-0306-6.
6. Belton, J.-M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* *58*, 268–276. 10.1016/j.ymeth.2012.05.001.
7. Bertero, A., and Rosa-Garrido, M. (2021). Three-dimensional chromatin organization in cardiac development and disease. *J Mol Cell Cardiol* *151*, 89–105. 10.1016/j.yjmcc.2020.11.008.
8. Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* *72*, 291–336. 10.1146/annurev.biochem.72.121801.161720.
9. Black, J.C., and Whetstone, J.R. (2011). Chromatin landscape: methylation beyond transcription. *Epigenetics* *6*, 9–15. 10.4161/epi.6.1.13331.
10. Burd, C.E., Jeck, W.R., Liu, Y., Sanoff, H.K., Wang, Z., and Sharpless, N.E. (2010). Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet* *6*, e1001233. 10.1371/journal.pgen.1001233.
11. Carlevaro-Fita, J., Polidori, T., Das, M., Navarro, C., Zoller, T.I., and Johnson, R. (2019). Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. *Genome Res* *29*, 208–222. 10.1101/gr.229922.117.
12. Cerasuolo, A., Buonaguro, L., Buonaguro, F.M., and Tornesello, M.L. (2020). The Role of RNA Splicing Factors in Cancer: Regulation of Viral and Human Gene Expression in Human Papillomavirus-Related Cervical Cancer. *Front Cell Dev Biol* *8*, 474. 10.3389/fcell.2020.00474.
13. Cho, H., Li, Y., Archacki, S., Wang, F., Yu, G., Chakrabarti, S., Guo, Y., Chen, Q., and Wang, Q.K. (2020). Splice variants of lncRNA RNA ANRIL exert opposing effects on endothelial cell activities associated with coronary artery disease. *RNA Biol* *17*, 1391–1401. 10.1080/15476286.2020.1771519.
14. Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. *Cell* *161*, 404–416. 10.1016/j.cell.2015.03.025.
15. Collins, B.E., Greer, C.B., Coleman, B.C., and Sweatt, J.D. (2019). Histone H3 lysine K4 methylation and its role in learning and memory. *Epigenetics Chromatin* *12*, 7. 10.1186/s13072-018-0251-8.
16. Congrains, A., Kamide, K., Ohishi, M., and Rakugi, H. (2013). ANRIL: molecular mechanisms and implications in human health. *Int J Mol Sci* *14*, 1278–1292. 10.3390/ijms14011278.
17. Cutter, A.R., and Hayes, J.J. (2015). A brief review of nucleosome structure. *FEBS Lett* *589*, 2914–2922. 10.1016/j.febslet.2015.05.016.
18. De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip Rev RNA* *4*, 49–60. 10.1002/wrna.1140.
19. de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* *60*, 676–684. 10.1016/j.molcel.2015.09.023.
20. Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P., et al. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* *503*, 371–376. 10.1038/nature12598.
21. Dorsett, D., and Ström, L. (2012). The ancient and evolving roles of cohesin in gene expression and DNA repair. *Curr Biol* *22*, R240–250. 10.1016/j.cub.2012.02.046.
22. Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol* *17*, 756–770. 10.1038/nrm.2016.126.
23. Eperon, I.C., Makarova, O.V., Mayeda, A., Munroe, S.H., Cáceres, J.F., Hayward, D.G., and Krainer, A.R. (2000). Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol* *20*, 8303–8318. 10.1128/MCB.20.22.8303-8318.2000.
24. Ernst, P., and Vakoc, C.R. (2012). WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Brief Funct Genomics* *11*, 217–226. 10.1093/bfgp/els017.
25. Folkersen, L., Kyriakou, T., Goel, A., Peden, J., Mälarstig, A., Paulsson-Berne, G., Hamsten, A., Hugh Watkins, null, Franco-Cereceda, A., Gabrielsen, A., et al. (2009). Relationship between CAD risk genotype in the chromosome 9p21 locus and gene expression. Identification of eight new ANRIL splice variants. *PLoS One* *4*, e7677. 10.1371/journal.pone.0007677.
26. Fort, V., Khelifi, G., and Hussein, S.M.I. (2021). Long non-coding RNAs and transposable elements: A functional relationship. *Biochim Biophys Acta Mol Cell Res* *1868*, 118837. 10.1016/j.bbamcr.2020.118837.
27. Gates, L.A., Foulds, C.E., and O'Malley, B.W. (2017). Histone Marks in the “Driver’s Seat”: Functional Roles in Steering the Transcription Cycle. *Trends Biochem Sci* *42*, 977–989. 10.1016/j.tibs.2017.10.004.
28. Ghigna, C., Valacca, C., and Biamonti, G. (2008). Alternative splicing and tumor progression. *Curr Genomics* *9*, 556–570. 10.2174/138920208786847971.
29. Gonzalez, I., Munita, R., Agirre, E., Dittmer, T.A., Gysling, K., Misteli, T., and Luco, R.F. (2015). A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nat Struct Mol Biol* *22*, 370–376. 10.1038/nsmb.3005.
30. Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., et al. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell* *24*, 206–214. 10.1016/j.devcel.2012.12.012.
31. Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D.G., Sauvageau, M., Kelley, D.R., et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol* *21*, 198–206. 10.1038/nsmb.2764.
32. Holdt, L.M., Stahring, A., Sass, K., Pichler, G., Kulak, N.A., Wilfert, W., Kohlmaier, A., Herbst, A., Northoff, B.H., Nicolaou, A., et al. (2016). Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun* *7*, 12429. 10.1038/ncomms12429.
33. Hon, G., Wang, W., and Ren, B. (2009). Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput Biol* *5*, e1000566. 10.1371/journal.pcbi.1000566.
34. Huang, C., and Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. *Biophys Rep* *4*, 170–177. 10.1007/s41048-018-0063-1.
35. Iwamori, N., Tominaga, K., Sato, T., Riehle, K., Iwamori, T., Ohkawa, Y., Coarfa, C., Ono, E., and Matzuk, M.M. (2016). MRG15 is required for pre-mRNA splicing and spermatogenesis. *Proc Natl Acad Sci U S A* *113*, E5408–5415. 10.1073/pnas.1611995113.
36. Jang, H.S., Shin, W.J., Lee, J.E., and Do, J.T. (2017). CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes (Basel)* *8*, E148. 10.3390/genes8060148.

37. Jerkovic, I., and Cavalli, G. (2021). Understanding 3D genome organization by multidisciplinary methods. *Nat Rev Mol Cell Biol* 22, 511–528. 10.1038/s41580-021-00362-w.
38. Jonas, K., Calin, G.A., and Pichler, M. (2020). RNA-Binding Proteins as Important Regulators of Long Non-Coding RNAs in Cancer. *Int J Mol Sci* 21, E2969. 10.3390/ijms21082969.
39. Jorjani, H., Kehr, S., Jedlinski, D.J., Gumienny, R., Hertel, J., Stadler, P.F., Zavolan, M., and Gruber, A.R. (2016). An updated human snoRNAome. *Nucleic Acids Res* 44, 5068–5082. 10.1093/nar/gkw386.
40. Jorjani, H., Kehr, S., Jedlinski, D.J., Gumienny, R., Hertel, J., Stadler, P.F., Zavolan, M., and Gruber, A.R. (2016). An updated human snoRNAome. *Nucleic Acids Res* 44, 5068–5082. 10.1093/nar/gkw386.
41. Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* 9, e1003470. 10.1371/journal.pgen.1003470.
42. Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* 13, R107. 10.1186/gb-2012-13-11-r107.
43. Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., and Xiong, Y. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 30, 1956–1962. 10.1038/ncr.2010.568.
44. Kunkler, C.N., Schiefelbein, G.E., O’Leary, N.J., McCown, P.J., and Brown, J.A. (2022). A single natural RNA modification can destabilize a U•A-T-rich RNA•DNA-DNA triple helix. *RNA* 28, 1172–1184. 10.1261/rna.079244.122.
45. Lim, P.S., Shannon, M.F., and Hardy, K. (2010). Epigenetic control of inducible gene expression in the immune system. *Epigenomics* 2, 775–795. 10.2217/epi.10.55.
46. Lou, N., Liu, G., and Pan, Y. (2020). Long noncoding RNA ANRIL as a novel biomarker in human cancer. *Future Oncol* 16, 2981–2995. 10.2217/fo-2020-0470.
47. Low, J.T., and Weeks, K.M. (2010). SHAPE-directed RNA secondary structure prediction. *Methods* 52, 150–158. 10.1016/j.ymeth.2010.06.007.
48. Lubelsky, Y., Zuckerman, B., and Ulitsky, I. (2021). High-resolution mapping of function and protein binding in an RNA nuclear enrichment sequence. *EMBO J* 40, e106357. 10.15252/embj.20210106357.
49. Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell* 144, 16–26. 10.1016/j.cell.2010.11.056.
50. Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science* 327, 996–1000. 10.1126/science.1184208.
51. Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianféran, S., Van Dorsselaer, A., Clerc, P., Avner, P., et al. (2010). 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol* 8, e1000276. 10.1371/journal.pbio.1000276.
52. Maenner, S., Müller, M., Fröhlich, J., Langer, D., and Becker, P.B. (2013). ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol Cell* 51, 174–184. 10.1016/j.molcel.2013.06.011.
53. Ou, M., Li, X., Zhao, S., Cui, S., and Tu, J. (2020). Long non-coding RNA CDKN2B-AS1 contributes to atherosclerotic plaque formation by forming RNA-DNA triplex in the CDKN2B promoter. *EBioMedicine* 55, 102694. 10.1016/j.ebiom.2020.102694.
54. Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40, 1413–1415. 10.1038/ng.259.
55. Pasmant, E., Laurendeau, I., Héron, D., Vidaud, M., Vidaud, D., and Bièche, I. (2007). Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res* 67, 3963–3969. 10.1158/0008-5472.CAN-06-2004.
56. Patil, D.P., Pickering, B.F., and Jaffrey, S.R. (2018). Reading m6A in the Transcriptome: m6A-Binding Proteins. *Trends Cell Biol* 28, 113–127. 10.1016/j.tcb.2017.10.001.
57. Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. *Nat Rev Genet* 14, 288–295. 10.1038/nrg3458.
58. Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Bruggmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323. 10.1016/j.cell.2007.05.022.
59. Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M., et al. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat Struct Mol Biol* 19, 1155–1160. 10.1038/nsmb.2391.
60. Saint-André, V., Batsché, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1 γ favor inclusion of alternative exons. *Nat Struct Mol Biol* 18, 337–344. 10.1038/nsmb.1995.
61. Saldanha, R., Mohr, G., Belfort, M., and Lambowitz, A.M. (1993). Group I and group II introns. *FASEB J* 7, 15–24. 10.1096/fasebj.7.1.8422962.
62. Sanchez, A., and Maenner, S. (2021). Purification of RiboNucleoProtein Particles by MS2-MBP Affinity Chromatography. *Methods Mol Biol* 2300, 99–106. 10.1007/978-1-0716-1386-3_10.
63. Sanidas, I., Polytaichou, C., Hatzia Apostolou, M., Ezell, S.A., Kottakis, F., Hu, L., Guo, A., Xie, J., Comb, M.J., Iliopoulos, D., et al. (2014). Phosphoproteomic screen reveals akt isoform-specific signals linking RNA processing to lung cancer. *Mol Cell* 53, 577–590. 10.1016/j.molcel.2013.12.018.
64. Sarkar, D., Oghabian, A., Bodiya badu, P.K., Joseph, W.R., Leung, E.Y., Finlay, G.J., Baguley, B.C., and Askarian-Amiri, M.E. (2017). Multiple Isoforms of ANRIL in Melanoma Cells: Structural Complexity Suggests Variations in Processing. *Int J Mol Sci* 18, 1378. 10.3390/ijms18071378.
65. Sarma, K., Cifuentes-Rojas, C., Ergun, A., Del Rosario, A., Jeon, Y., White, F., Sadreyev, R., and Lee, J.T. (2014). ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell* 159, 869–883. 10.1016/j.cell.2014.10.019.
66. Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* 16, 990–995. 10.1038/nsmb.1659.
67. Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I.K., Wang, J.-P.Z., and Widom, J. (2006). A genomic code for nucleosome positioning. *Nature* 442, 772–778. 10.1038/nature04979.
68. Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A., and Weeks, K.M. (2015). Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc* 10, 1643–1669. 10.1038/nprot.2015.103.
69. Song, S.-H., and Kim, T.-Y. (2017). CTCF, Cohesin, and Chromatin in Human Cancer. *Genomics Inform* 15, 114–122. 10.5808/GI.2017.15.4.114.
70. Spies, N., Nielsen, C.B., Padgett, R.A., and Burge, C.B. (2009). Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* 36, 245–254. 10.1016/j.molcel.2009.10.008.
71. Stephenson, W., Razaghi, R., Busan, S., Weeks, K.M., Timp, W., and Smibert, P. (2022). Direct detection of RNA modifications and structure using single-molecule nanopore sequencing. *Cell Genomics* 2, 100097. <https://doi.org/10.1016/j.xgen.2022.100097>.
72. Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Sci Adv* 5, eaaw1668. 10.1126/sciadv.aaw1668.
73. Szerlong, H.J., and Hansen, J.C. (2011). Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem Cell Biol* 89, 24–34. 10.1139/O10-139.

74. Tajbakhsh, A., Khorrami, M.S., Hassanian, S.M., Aghasizade, M., Pasdar, A., Maftouh, M., Tabatabai, E., Parizadeh, S.M.R., Fazeli, M., Ferns, G.A., et al. (2016). The 9p21 Locus and its Potential Role in Atherosclerosis Susceptibility; Molecular Mechanisms and Clinical Implications. *Curr Pharm Des* 22, 5730–5737. 10.2174/1381612822666160628082453.
75. Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., and Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol* 16, 996–1001. 10.1038/nsmb.1658.
76. Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693. 10.1126/science.1192002.
77. Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693. 10.1126/science.1192002.
78. Wang, Z.-W., Pan, J.-J., Hu, J.-F., Zhang, J.-Q., Huang, L., Huang, Y., Liao, C.-Y., Yang, C., Chen, Z.-W., Wang, Y.-D., et al. (2022). SRSF3-mediated regulation of N6-methyladenosine modification-related lncRNA ANRIL splicing promotes resistance of pancreatic cancer to gemcitabine. *Cell Rep* 39, 110813. 10.1016/j.celrep.2022.110813.
79. Xiang, J.-F., Yin, Q.-F., Chen, T., Zhang, Y., Zhang, X.-O., Wu, Z., Zhang, S., Wang, H.-B., Ge, J., Lu, X., et al. (2014). Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 24, 513–531. 10.1038/cr.2014.35.
80. Yang, F., Deng, X., Ma, W., Berletch, J.B., Rabaia, N., Wei, G., Moore, J.M., Filippova, G.N., Xu, J., Liu, Y., et al. (2015). The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol* 16, 52. 10.1186/s13059-015-0618-0.
81. Yao, J., Tu, Y., Shen, C., Zhou, Q., Xiao, H., Jia, D., and Sun, Q. (2021). Nuclear import receptors and hnRNPK mediates nuclear import and stress granule localization of SIRLOIN. *Cell Mol Life Sci* 78, 7617–7633. 10.1007/s00018-021-03992-7.
82. Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.-M. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* 38, 662–674. 10.1016/j.molcel.2010.03.021.
83. Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J.-P., Nissim-Rafinia, M., Cohen, A.-H.S., Rippe, K., Meshorer, E., et al. (2015). HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Rep* 10, 1122–1134. 10.1016/j.celrep.2015.01.038.
84. Yi, W., Li, J., Zhu, X., Wang, X., Fan, L., Sun, W., Liao, L., Zhang, J., Li, X., Ye, J., et al. (2020). CRISPR-assisted detection of RNA-protein interactions in living cells. *Nat Methods* 17, 685–688. 10.1038/s41592-020-0866-0.
85. Yoon, J.-H., De, S., Srikantan, S., Abdelmohsen, K., Grammatikakis, I., Kim, J., Kim, K.M., Noh, J.H., White, E.J.F., Martindale, J.L., et al. (2014). PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nat Commun* 5, 5248. 10.1038/ncomms6248.
86. Zhang, C., Ge, S., Gong, W., Xu, J., Guo, Z., Liu, Z., Gao, X., Wei, X., and Ge, S. (2020). LncRNA ANRIL acts as a modular scaffold of WDR5 and HDAC3 complexes and promotes alteration of the vascular smooth muscle cell phenotype. *Cell Death Dis* 11, 435. 10.1038/s41419-020-2645-3.
87. Zhou, X., Han, X., Wittfeldt, A., Sun, J., Liu, C., Wang, X., Gan, L.-M., Cao, H., and Liang, Z. (2016). Long non-coding RNA ANRIL regulates inflammatory responses as a novel component of NF-κB pathway. *RNA Biol* 13, 98–108. 10.1080/15476286.2015.1122164.
88. Zhou, X., Li, X., Yu, L., Wang, R., Hua, D., Shi, C., Sun, C., Luo, W., Rao, C., Jiang, Z., et al. (2019). The RNA-binding protein SRSF1 is a key cell cycle regulator via stabilizing NEAT1 in glioma. *Int J Biochem Cell Biol* 113, 75–86. 10.1016/j.biocel.2019.06.003.