

AI for Health

Chroniques AI for Health - Série des « Omiques »

La transcriptomique : l'ARN passé au crible

Temps de lecture : moins de 10 mn

Bienvenue dans ce deuxième épisode de notre série des omiques sur la **transcriptomique** ! Si vous avez manqué le numéro précédent sur la génomique, vous pourrez le retrouver sur notre LinkedIn !

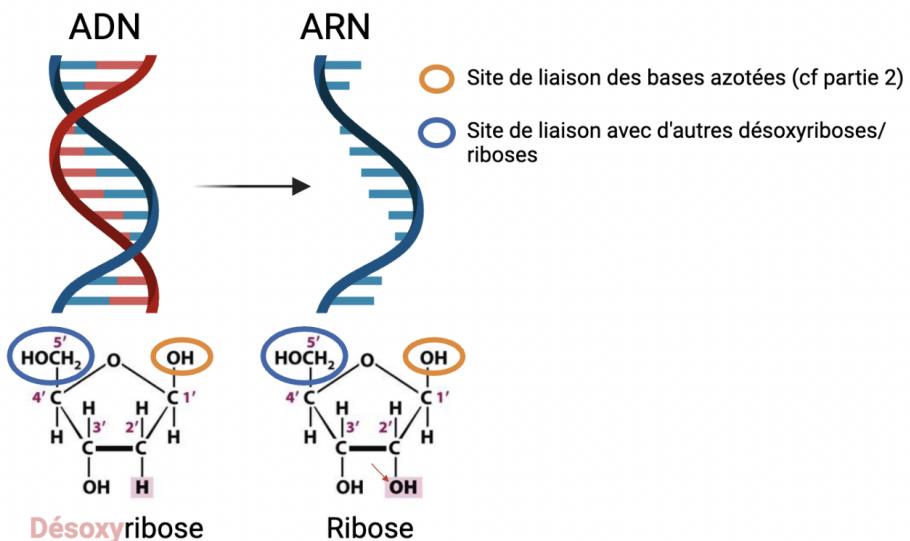
Rappelez-vous, l'étude des omiques peut se découper en 4 grands domaines d'étude : la génomique liée à l'ADN et aux gènes, la transcriptomique pour l'ARN, la protéomique pour les protéines et la métabolomique pour les métabolites.

Tout comme la génomique s'intéresse au génome, la transcriptomique s'intéresse au transcriptome. Mais, à quoi cela correspond-il ?

Partie 1 : l'ARN

Vous vous rappelez du nom complet de l'ADN ? **Acide DésoxyriboNucléique**. L'ARN, ou **Acide RiboNucléique** est une molécule qui “ressemble” à l'ADN à une exception près : elle possède un atome d'oxygène de plus sur sa base (**désoxy**ribo -> ribo). Cette différence va notamment causer le fait que l'ARN est une molécule dite simple brin en opposition à l'ADN qui est double brins.

Si l'ADN et l'ARN se ressemblent tellement, c'est parce que l'une des molécules sert de patron pour créer l'autre par un processus dit de transcription. Si vous avez lu notre premier article sur la génomique, vous savez que c'est l'ADN qui sert de base !



Si l'ARN est utilisé au singulier, en réalité il en existe plusieurs types qui jouent des rôles différents et parfois complémentaires dans les processus biologiques.

AI for Health

Un ARN est tout d'abord dit **codant** ou **non codant**, au même titre qu'une séquence d'ADN dans un gène. Les ARN codant pour des protéines sont appelés ARN messagers (ARNm) et ne représentent que... 5% environ de l'ensemble des ARN.

Les autres types d'ARN assurent différentes fonctions : transfert d'acides aminés* pour former une protéine, interférence avec un autre ARN, formation de ribosomes,... *Quelques notions que nous verrons dans le prochain épisode sur la protéomique !*

Concentrons-nous donc sur l'ARN qui permet à terme la formation d'une protéine : **l'ARN messager**.

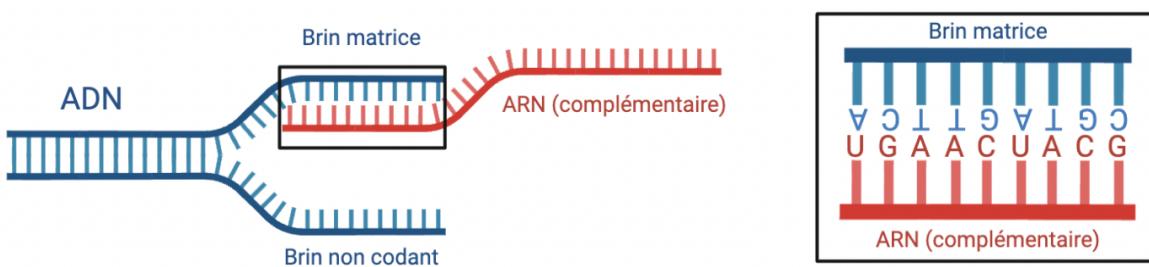
Comment est-il formé et quels sont ses intérêts ?

Partie 2 : la transcription

Des régions spécifiques de l'ADN forment, rappelons-le, des gènes. Ces gènes sont à l'origine de la production de protéines par l'intermédiaire d'un type d'ARN dit "messager" ou **ARNm**. Le "passage" de l'ADN vers l'ARNm est appelé transcription.

Si nous reprenons la métaphore du langage utilisée dans notre épisode précédent, le mécanisme de transcription porte bien son nom car il consiste à copier très exactement un brin d'ADN (matrice) en un brin d'ARN **complémentaire** de sa séquence.

Une image vaut mille mots :

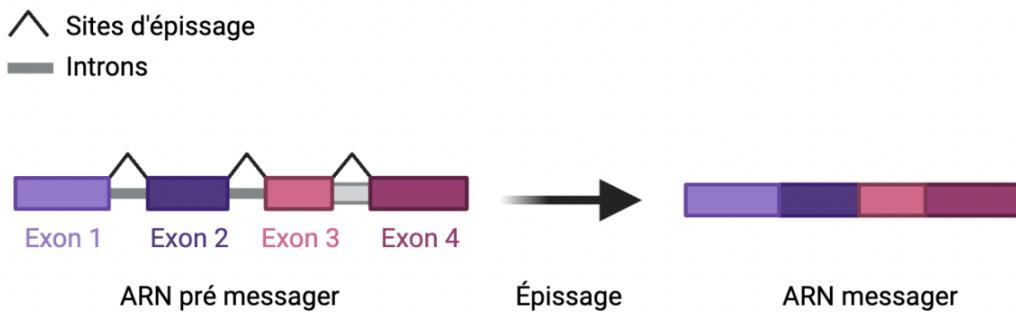


Avez-vous noté l'intrus dans les différentes bases azotées (lettres) qui composent l'ARN ? **A** (adénine) est complémentaire de **T** (thymine) et **C** (cytosine) de **G** (guanine)... Mais dans l'ARN, T est remplacé par sa molécule cousine : **U** pour uracile !

L'ARN nouvellement formé est dit pré-messager, et se situe donc toujours à l'intérieur du noyau, là où se trouve l'ADN.

Avant d'être qualifié d'ARNm, l'ARN pré-messager subit une étape clé appelée **épissage**. Grâce à une machinerie cellulaire spécifique, certains blocs de séquences appelés **introns** sont retirés de l'ARN afin de ne laisser que des **exons**.

AI for Health



Cette étape accomplie, l'ARN est déplacé à l'extérieur du noyau de la cellule via les pores nucléaires et arrive dans le cytoplasme. C'est à cet endroit qu'aura lieu le prochain processus donnant lieu à la production de protéines : la traduction.

Chasse aux fake news : Du fait même de sa structure moléculaire, un ARNm est dit "séquestré" dans le cytoplasme et ne peut donc pas rentrer dans le noyau. Ainsi, **il n'est pas probable qu'un ARNm modifie le matériel génétique** qui, lui, réside à l'intérieur du noyau. Un vaccin à ARN, par exemple, ne peut donc pas entraîner de modification au niveau du génome.

Partie 3 : la transcriptomique

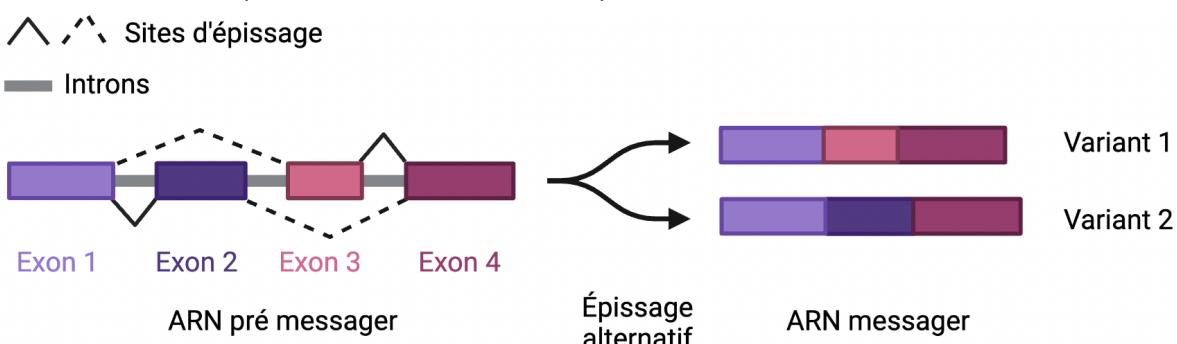
Revenons sur une particularité biologique merveilleuse : l'épissage.

Le génome humain possède entre 20 000 et 25 000 gènes en fonction des estimations. Cependant, des estimations portent le nombre de protéines produites à plus de 2 millions chez l'Homme.

Si on suivait un schéma intuitif, un gène devrait donner un ARNm puis une protéine... Mais le vivant est plein de ressources lorsqu'il s'agit d'optimisation ! Le génome est en réalité capable de générer à terme de très nombreuses protéines à partir de "peu" de gènes grâce à l'**épissage alternatif**.

À la différence de l'épissage "classique" dit constitutif, l'épissage alternatif permet une "sélection" des exons formant l'ARNm. Ainsi, **un seul ARN pré messager peut donner plusieurs ARNm différents**, appelés variants d'épissage !

C'est un peu comme si vous ouvriez un dossier compressé et que vous sélectionniez uniquement le fichier correspondant à votre utilité.



AI for Health

La **transcriptomique** est une discipline étudiant l'ensemble des ARN, d'un point de vue quantitatif et qualitatif. Dans le cas des ARN codant, la notion de transcrits (ou variants) est essentielle.

Dans ce domaine, la récolte des données, leur structuration et leur analyse est également fondamentale. En effet, l'on ne peut qu'imaginer la quantité de données différentes issues des analyses d'un ensemble de molécules aussi complexes et avec des fonctions différentes.

L'intelligence artificielle joue un rôle majeur dans le développement de nouvelles méthodes de screening, de simulations *in silico* ou encore de modélisation.

Quelques **exemples d'application** de la transcriptomique épaulée par la data science et l'IA :

- Identification de cibles thérapeutiques ;
- Quantification de la présence de transcrits ARNm ;
- Modélisation d'interaction d'ARN interférents (ou autre) à visée thérapeutique ;
- Identification de nouveaux variants d'épissage pour la caractérisation de protéines impliquées dans une maladie ;
- Reproduction ou création d'ARNm pour leur usage dans des vaccins ou autres thérapies ;
- ...

NB : Dans cet épisode, nous nous sommes intéressés plus fortement aux ARNm pour la création d'un fil rouge entre les différents articles, mais la transcriptomique permet bien une analyse de l'ensemble des types d'ARN et ne se limite donc pas aux seuls ARN codant pour des protéines.

AI for Health Summit 2022

Chez **AI for Health**, nous cherchons à accueillir le plus grand nombre sur les enjeux de datas, IA et santé dans un monde en constante évolution. Des problématiques et enjeux divers, incluant les évolutions en transcriptomique, seront adressées lors de notre **Summit annuel le 16 novembre 2022** à Station F, Paris !

Enfin, si vous voulez continuer de suivre les étapes menant à la production de protéines et de leurs métabolites, ne manquez pas les prochains épisodes de la série des Omiques !

Prochain épisode : "La protéomique : le royaume de la fonction des gènes"

Vous souhaitez nous contacter pour nous rejoindre, nous proposer une mission ou simplement vous renseigner ? **Ecrivez-nous à ai@startupinside.com.** Nous sommes également preneurs de vos propositions de sujets d'articles !