les exons ne sont pas forcément codants

publié le 18.01.24 |par pascal combemorel

dans un court article paru dans cell genomics, trois chercheurs rappellent que les exons ne sont pas forcément des séquences traduites en protéines et mettent en garde contre l'emploi de l'expression « séquençage de l'exome » pour désigner une technique qui séquence surtout les exons codants.

alors que certains ouvrages universitaires définissent les exons comme des séquences dont la traduction aboutit à la formation d'une protéine (en particulier « le » alberts [1] et « le » stryer [2]), julie aspden et ses collaborateurs rappellent dans un court article que le terme exon désigne en réalité une séquence d'arn toujours présente après l'épissage, contrairement aux introns [3]. le terme d'exon désigne aussi bien cette séquence sur l'arn que son équivalent dans l'adn.

les exons contenus dans une molécule d'arn messager correspondent donc aussi bien aux séquences traduites en protéine qu'aux séquences non traduites qui la flanquent en 5' (5'utr, pour *untranslated regions*) et en 3' (3'utr). par ailleurs les arn non codants (arnt, arnr, etc.) sont également formés d'exons (figure 1). ainsi, chez l'être humain, seuls 23 % des exons sont codants tandis que 32 % correspondent à des extrémités 5' et 3' non traduites et 37 % à des arn non codants.

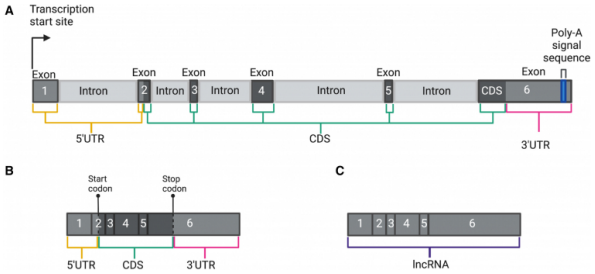


figure 1 - les exons ne correspondent pas forcément à des séquences codantes
a. organisation d'une portion d'adn codant un arn codant lui-même une protéine. les exons 3, 4 et 5 sont constitués exclusivement de séquences codantes, l'exon 1 est entièrement situé dans la région 5'utr, les exons 2 et 6 contiennent à la fois une partie codante et une partie non codante.
b. arnm obtenu après épissage du transcrit produit à partir de la séquence d'adn représentée en a.
c. un long arn non codant, également formé de six exons, tous non codants.
5'utr : extrémité 5' non traduite ; cds : séquence codante ; 3'utr : extrémité 3' non traduite ; lncrna : long arn non codant.

auteur(s)/autrice(s) : aspden et coll., 2023 licence : cc-by-nc-nd
source : cell genomics

par ailleurs, les auteurs insistent sur la dénomination trompeuse donnée à la technique de séquençage de l'exome (ou séquençage de l'exome entier, de l'anglais ***whole-exome sequencing***) qui, contrairement à ce que son nom suggère, séquence surtout les exons codants au détriment des exons non codants. c'est ainsi que 89 % des bases séquencées par séquençage de l'exome correspondent à des séquences codantes, ce qui mène à une importante sous-représentation des extrémités non traduites et des arn non codants dans les bases de données issues des techniques de séquençage de l'exome (figure 2). pour désigner la technique consistant à séquencer essentiellement les transcrits codants, et éviter ainsi toute ambiguïté, les auteurs proposent d'employer l'expression de séquençage de l'exome codant (ces, ***coding exome sequencing***).

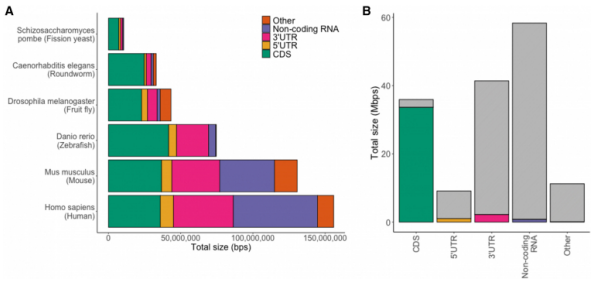


figure 2 - exons et données de séquençage de l'exome
a. comparaison, chez six organismes, de la proportion de bases situées dans des exons et annotées en tant que séquence codante (cds), 5'utr, 3'utr, arn non codant et autre (incluant les éléments transposables et les pseudogènes).
b. histogramme indiquant, pour l'être humain, le nombre total de bases situées dans des exons et annotées de la même manière qu'en figure a. les bases séquencées par les techniques de séquençage de l'exome sont en couleur, celles qui ne sont pas séquencées grâce à ces techniques sont en gris.

auteur(s)/autrice(s) : aspden et coll., 2023 licence : cc-by-nc-nd
source : cell genomics

sans nier l’intérêt de disposer d’informations sur les exons codants, les auteurs soulignent la nécessité de s’intéresser également aux exons non codants. en effet, les extrémités 5’ et 3’utr sont indispensables à la stabilité, à la localisation et à la traduction correcte des arnm. ces extrémités contribuent également à la régulation de la traduction via l’interaction avec des protéines et des microarn. en outre les arn non codants jouent également des rôles importants. par exemple le microarn mir-204 est indispensable au fonctionnement des photorécepteurs de la rétine chez l’être humain, tandis que *xist*, un long arn non codant, est impliqué dans l’inactivation du chromosome x chez les mammifères.

CRÉDITS

AUTEUR(S)/AUTRICE(S)

pascal combemorel

agrégé de svt, il est le responsable éditorial du site planet-vie depuis septembre 2016.

RELECTURE SCIENTIFIQUE

morgane thomas-chollier

maître de conférences en bioinformatique à l'école normale supérieure.

LICENCE DU TEXTE DE L'ARTICLE



creative commons - attribution

bibliographie

1. **alberts, b. ., wilson, j. ., hunt, t. ., & johnson, a. . (éd.). (2008). *molecular biology of the cell* (5. ed). new york, ny: garland science.**
2. **berg, j. m., tymoczko, j. l., gatto, g. j., & stryer, l. . (2015). *biochemistry* (eighth edition). new york: w.h. freeman \& company, a macmillan education imprint.**
3. **aspden, j. l., wallace, e. w., & whiffin, n. . (2023). not all exons are protein coding: addressing a common misconception. *cell genomics*, *3*, 100296. http://doi.org/10.1016/j.xgen.2023.100296 (original work published 2026)**