

Chroniques génomiques

Le Junk DNA n'est plus ce qu'il était

Bertrand Jordan

Junk or not Junk ?

Lorsqu'il est devenu évident que l'ADN humain ne comportait qu'environ 2 % de séquences codantes (au sens classique¹), les discussions sur la nature des 98 % restants ont oscillé entre deux pôles : certains se refusaient à admettre qu'une partie de notre génome soit inutile et, sans doute influencés par l'idée d'une nature « parfaite », affirmaient que tout cet ADN devait avoir un rôle important quoiqu'encore mystérieux. D'autres, conscients avec François Jacob que la nature est le résultat d'un très long bricolage [1], voyaient dans cet « ADN poubelle » une scorie de l'évolution, un amas de séquences inutiles que la sélection naturelle n'avait pas (encore ?) pris la peine d'éliminer. J'avais moi-même épousé cette thèse dans ma « *Fugu story* » en 1994 [2] (→) à propos de ce poisson proposé comme nouveau modèle par Sydney Brenner, un vertébré dont le génome contient fort peu de *Junk DNA* et que cela ne semble pas gêner le moins du monde. Depuis, de nombreuses études, et surtout le projet ENCODE [3], ont montré que quasiment tout notre génome est transcrit et ont affirmé qu'il s'agissait donc de séquences fonctionnelles, affirmation discutée par plusieurs auteurs [4, 5] (→).

Il a depuis été démontré que certains de ces transcrits ARN jouent un rôle dans la régulation de gènes, et que, dans quelques cas, certains sont même traduits [6]. Un article récent paru dans la revue *Science* [7] réexamine cette question et montre l'existence de nombreux petits ORF (*open reading frame*, ou cadre de lecture) transcrits à partir de cet ADN, qui sont traduits et dont les produits protéiques jouent un rôle dans les fonctions cellulaires. C'est donc un nouveau monde, celui des microORF, qui est révélé par ces travaux d'une grande importance.

¹ Avec un codon d'initiation, un cadre de lecture correspondant à quelques centaines d'acides aminés, éventuellement morcelé en exons séparés par des introns, et un codon de terminaison.

(→) Voir la Chronique génomique de B. Jordan, *m/s* n° 12, décembre 1994, page 1154

(→) Voir le Forum de D. Casane et al., *m/s* n° 6-7, juin-juillet 2015, page 680



UMR 7268 ADÉS, Aix-Marseille, Université /EFS/CNRS ; CoReBio PACA, case 901, Parc scientifique de Luminy, 13288 Marseille Cedex 09, France.
bjordan@orange.fr

Une exploration systématique

Par rapport aux critères classiques définissant un ORF dans une séquence d'ADN (longueur correspondant à au moins 100 acides aminés, présence d'un codon d'initiation ATG, conservation entre espèces) [8], les auteurs ont éliminé tout critère sur la longueur et admis des codons d'initiation non classiques (CTG, GTG, TTG). Par contre, ils ont sélectionné les séquences à analyser en se limitant aux fragments d'ARN encore protégés de la dégradation enzymatique par les ribosomes – donc *a priori* en cours de traduction. Cette technique, appelée *ribosome footprinting* [9, 10] ou *ribosome profiling* [11], repose sur l'isolement de polysomes suivi de leur traitement par une RNase, puis de la récupération et du séquençage des fragments protégés (longs de 20 à 30 nucléotides). Les séquences obtenues sont alors comparées à la séquence du génome pour détecter les ORF, chacun d'entre eux correspondant à plusieurs fragments protégés qui se recouvrent partiellement. Les ORF ainsi découverts correspondent en principe à des séquences transcrites et traduites puisqu'elles sont « couvertes » par des ribosomes dans les polysomes. En réalisant ces expériences sur plusieurs lignées cellulaires (fibroblastes, cardiomyocytes, cellules souche pluripotentes, etc.), les auteurs détectent au total 15411 ORF ; 9490 d'entre eux correspondent à des gènes déjà connus et annotés dans les bases de données, codant des protéines généralement longues de plusieurs centaines d'acides aminés ; 2466 à des variants de ces séquences (épissage alternatif, troncations, extensions), et 3455 à de nouveaux ORF, généralement de petite taille (moins de



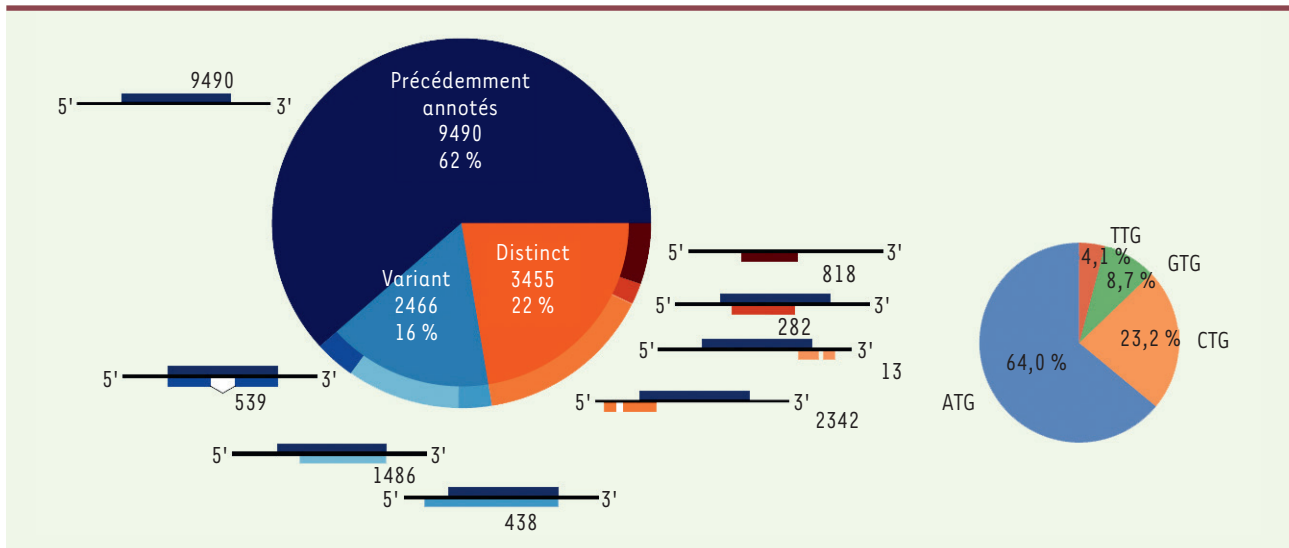


Figure 1. Cadres de lecture ouverts (ORF) repérés par ribosome profiling. À gauche, les trois classes principales : en bleu foncé et bleu clair, les gènes déjà annotés et les variants de ces gènes (type de variant indiqué par les schémas : épissage alternatif, troncation, extension). En bistre : les nouveaux ORF et leurs différentes catégories : isolés, à l'intérieur d'un ORF connu (mais avec un autre cadre de lecture), en 3', en 5'. L'effectif de chaque catégorie est indiqué en-dessous du schéma correspondant. On note que les petits ORF en 3' de gènes sont très minoritaires par rapport à ceux qui sont en 5'. À droite, répartition des codons d'initiation utilisés dans les nouveaux ORF (extrait partiel et modifié de la figure 1 de [7]).

100 acides aminés) et, pour presque la moitié d'entre eux, avec un codon d'initiation de la traduction non-classique (Figure 1).

Les microORF sont réellement exprimés en protéines

Le provenance des séquences qui détectent ces ORF garantit, en principe, qu'il s'agit bien de séquences transcrites et traduites en protéines (fussent-elles de petite taille) puisqu'elles étaient protégées par les ribosomes de la dégradation, donc en cours de traduction. À titre de vérification, les auteurs ont répété les expériences sur des cellules traitées à la harringtonine, un alcaloïde naturel qui immobilise les ribosomes immédiatement après le début de la traduction [12]. Ils ont alors constaté que, comme prévu, les fragments protégés se situaient alors au tout début du cadre de lecture. Ils ont aussi pu montrer que certains des peptides codés par les microORF étaient présentés à la surface des cellules par les molécules du complexe HLA (*human leukocyte antigen*) de classe I. Pour cela, ils ont eu recours à des expériences de « peptidomique HLA » [13] dans lesquelles ces molécules sont isolées à partir d'un lysat cellulaire, puis les peptides associés sont dissociés et analysés par spectrométrie de masse. Ils ont ainsi détecté plus de 200 peptides dérivés de ces microORF et associés aux molécules du complexe HLA de classe I, qui font ainsi partie du répertoire antigénique et peuvent jouer un rôle dans les processus immunitaires.

Les protéines codées par les microORF jouent un rôle dans les cellules

La preuve finale qu'une protéine est fonctionnelle consiste à observer un phénotype provoqué par son inactivation. La technique CRISPR permet en principe d'inactiver chaque microORF en dirigeant le système

par un sgARN (*single-guide RNA*) spécifique – mais il est un peu délicat d'envisager de répéter 3455 fois une telle expérience ! Les auteurs ont donc choisi les 2352 microORF qui, par leur séquence, se prêtaient le mieux à un test d'inactivation, ont fait synthétiser les milliers d'oligonucléotides correspondant aux sgARN (deux par microORF), et ont procédé à la modification CRISPR « en masse » sur des cellules, dans des conditions assurant qu'une cellule subit au plus une modification [14]. Ils ont ensuite cultivé les cellules dans des conditions standard et suivi le devenir des cellules ayant subi l'inactivation d'un microORF donné grâce à la séquence spécifique du sgARN qu'elle contient, dont ils ont mesuré la fréquence par séquençage. Si l'inactivation de ce microORF diminue la viabilité de la cellule, son abondance va décroître au fur et à mesure des passages en culture. Plus de 500 cas où l'inactivation de l'ORF entraîne une perte de viabilité des cellules ont ainsi été identifiés. La Figure 2 montre les résultats pour six de ces ORF qui ont été alors étudiés pour des expériences d'inactivation individuelle. Les auteurs ont également procédé à des séquençages d'ARN sur cellule unique afin de comparer les profils d'expression avec et sans inactivation d'un micro ORF donné, et ont constaté des changements importants dans le profil, suggérant dans certains cas l'implication de cet ORF dans une fonction spécifique. On voit donc qu'au moins une partie des microORF identifiés par les auteurs code des petites protéines qui ont une importance fonctionnelle pour les cellules : ce sont décidément des ORF à prendre au sérieux, et non des artéfacts marginaux.

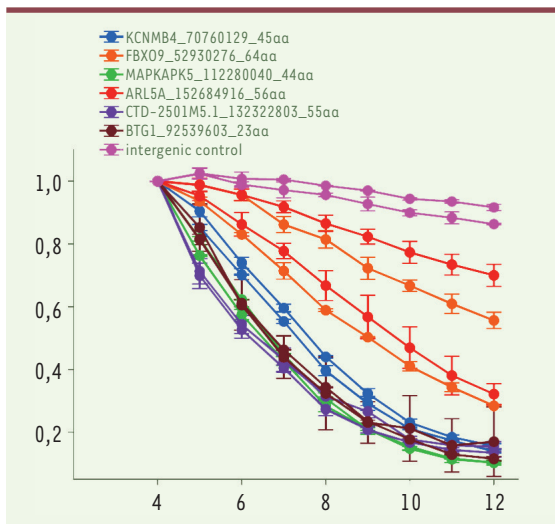


Figure 2. Inactivation de six courts ORF et diminution de la fréquence des cellules correspondantes. En abscisse, jours de culture, en ordonnée, fréquence de chaque clone cellulaire exprimée par rapport à sa fréquence au jour 4. Chaque couleur correspond à un ORF ciblé par deux sgARN différents, la taille du peptide codé est indiquée en haut. Les deux courbes roses correspondent au témoin négatif (ciblage d'une région intergénique) (extrait partiel et remanié de la figure 8 du supplément de [7]).

Un monde nouveau à explorer

La suite de l'article rapporte toute une série d'expériences visant à préciser le rôle des petites protéines codées par les microORF. Les auteurs ont pu préciser leur localisation cellulaire grâce à une technique de fluorescence adaptée, dite *split-fluorescent protein* [15], dans laquelle une petite partie de la protéine fluorescente est fusionnée au peptide, le reste étant exprimé dans la cellule et permettant une complémentation conduisant à un signal fluorescent. Les localisations ainsi révélées vont du réticulum endoplasmique à la membrane plasmique, en passant par les mitochondries ou l'appareil de Golgi, ce qui, avec les profils d'expression, permet parfois d'émettre des hypothèses fonctionnelles. Il a aussi été possible de rechercher des partenaires d'interaction pour ces petites protéines, et de montrer que, dans quelques cas, ce partenaire est la protéine codée par un ORF classique situé en 3' de ce microORF, et de prouver que les deux séquences nucléotidiques sont portés par le même ARN messager : première mise en évidence d'un ARN bicistronique chez des mammifères.

De nombreuses pistes de recherche sont esquissées dans cette partie particulièrement touffue de l'article²,

et vont sans doute faire l'objet de développements très intéressants. Comme l'évoque une analyse parue dans le même numéro de *Science* [16], cette découverte pourrait même avoir des conséquences pour la compréhension de maladies génétiques liées à l'expansion de triplets de nucléotides (maladie de Huntington, syndrome de l'X fragile, etc.) en indiquant la possibilité et les modalités de transcription et traduction de ces zones répétées. Il s'agit en tout cas d'une avancée importante, certes déjà entrevue [8, 9], mais ici systématisée et approfondie. Quoi qu'il en soit, la vision un peu simpliste du *Junk DNA* qui avait cours dans les années 1990, est définitivement écartée. ♦

SUMMARY

Junk DNA is out of fashion

A systematic search for non-conventional open reading frames in human DNA reveals a large number of small ORFs encoding peptides generally smaller than 100 amino-acids. These ORFs are transcribed and translated into small proteins, which are demonstrated to have functional significance by bulk CRISPR inactivation. Evidence is also found for bicistronic mRNAs including such a small ORF upstream of a canonical coding sequence. These findings add a new facet to our understanding of biological processes. ♦

LIENS D'INTÉRÊT

L'auteur déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

- Jacob F. Evolution and tinkering. *Science* 1977 ; 196 : 1161-6.
- Jordan B. Fugu story. *Med Sci (Paris)* 1994 ; 10 : 1154-6.
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012 ; 489 : 57-74
- Casane D, Fumey J, Laurenti P. L'apophénie d'ENCODE ou Pangloss examine le génome humain. *Med Sci (Paris)* 2015 ; 31 : 680-6
- Palazzo AF, Lee ES. Non-coding RNA: what is functional and what is junk ? *Front Genet* 2015 ; 6 : 2.
- Li LJ, Leng RX, Fan YG, Pan HF, Ye DQ. Translation of noncoding RNAs: focus on lncRNAs, pri-miRNAs, and circRNAs. *Exp Cell Res* 2017 ; 361 : 1-8.
- Chen J, Brunner AD, Cogan JZ, et al. Pervasive functional translation of noncanonical human open reading frames. *Science* 2020 ; 367 : 1140-6.
- Basrai MAI, Hieter P, Boeke JD. Small open reading frames: beautiful needles in the haystack. *Genome Res* 1997 ; 7 : 768-71
- Bazzini AA, Johnstone TG, Christiano R, et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J* 2014 ; 33 : 981-93.
- Eastman G, Smirich P, Sotelo-Silveira JR. Following ribosome footprints to understand translation at a genome wide level. *Comput Struct Biotechnol J* 2018 ; 16 : 167-76.
- McGlinchey NJ, Ingolia NT. Transcriptome-wide measurement of translation by ribosome profiling. *Methods* 2017 ; 126 : 112-29.
- Fresno M, Jiménez A, Vázquez D. Inhibition of translation in eukaryotic systems by harringtonine. *Eur J Biochem* 1977 ; 72 : 323-30.
- Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol Cell Proteomics* 2015 ; 14 : 658-73.
- Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet* 2015 ; 16 : 299-311.
- Pedelaq JD, Cabantous S. Development and applications of superfolder and split fluorescent protein detection systems in biology. *Int J Mol Sci* 2019 ; 20 : 3479.
- Wei LH, Guo JU. Coding functions of noncoding RNAs. *Science* 2020 ; 367 : 1074-5.

² Ce qui reflète sans doute les demandes des *referees*, l'article ayant été soumis en mai 2019 puis re-soumis (après corrections, donc) fin novembre de la même année.

TIRÉS À PART

B. Jordan