



de tumeur connus tels que TP53 ou BRCA1 et de plus en plus de mutations dans des régulateurs épigénétiques sont identifiées dans les tumeurs par les séquençages de génomes tumoraux entiers. Le développement de stratégies thérapeutiques visant à inhiber ces oncoprotéines est crucial pour l'éradication des CSC qui sont responsables de l'activité tumorigénique de la tumeur et de sa récurrence. ♦

ZNF703: a novel oncogene involved in breast cancer

CONFLIT D'INTÉRÊTS

Les auteurs déclarent n'avoir aucun conflit d'intérêts concernant les données publiées dans cet article.

RÉFÉRENCES

- Charafe-Jauffret E, Chaffanet M, Bertucci F, et al. Towards an integrated cellular and molecular : definition of breast cancers. *Med Sci (Paris)* 2007 ; 23 : 626-32.
- Charafe-Jauffret E, Monville F, Ginestier C, et al. Cancer stem cells in breast : current opinion and future challenges. *Pathobiology* 2008 ; 75 : 75-84.
- Ginestier C, Korkaya H, Dontu G, Birnbaum D, Wicha MS, Charafe-Jauffret E. The cancer stem cell : the breast cancer driver. *Med Sci (Paris)* 2007 ; 23 : 1133-9.
- Sircoulomb F, Nicolas N, Ferrari A, et al. ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. *EMBO Mol Med* 2011 ; 3 : 153-66.
- Holland D, Burleigh A, Git A, et al. ZNF703, a luminal B breast cancer oncogene, is a transcriptional repressor and differentially regulates luminal and basal progenitors in human mammary epithelium. *EMBO Mol Med* 2011 ; 3 : 167-80.
- Turner N, Pearson A, Sharpe R, et al. FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in breast cancer. *Cancer Res* 2010 ; 70 : 2085-94.
- Zhang J, Liu X, Datta A, et al. RCP is a human breast cancer-promoting gene with Ras-activating function. *J Clin Invest* 2009 ; 119 : 2171-83.
- Bernard-Pierrot I, Gruel N, Stransky N, et al. Characterization of the recurrent 8p11-12 amplicon identifies PPAPDC1B, a phosphatase protein, as a new therapeutic target in breast cancer. *Cancer Res* 2008 ; 68 : 7165-75.
- Kwek SS, Roy R, Zhou H, et al. Co-amplified genes at 8p12 and 11q13 in breast tumors cooperate with two major pathways in oncogenesis. *Oncogene* 2009 ; 28 : 1892-1903.
- Nakamura M, Runko AP, Sagerstrom CG. A novel subfamily of zinc finger genes involved in embryonic development. *J Cell Biochem* 2004 ; 93 : 887-95.

NOUVELLE

De longs ARN non codants activateurs de la transcription des gènes

Thomas Derrien, Roderic Guigó

T. Derrien : Institut de génétique et développement de Rennes 1, CNRS UMR6061, 2, avenue du Pr Léon Bernard, Faculté de médecine, Université de Rennes1, Rennes, France.

R. Guigó : Bioinformatics and genomics group, Center for genomic regulation, Barcelona, Catalonia, Espagne.

toma.derrien@gmail.com

roderic.guigo@crg.cat

► Le transcriptome d'une cellule représente l'ensemble des molécules d'ARN codant et ne codant pas pour des protéines. L'émergence des nouvelles techniques de séquençage à haut débit permet à présent d'appréhender la complexité d'un transcriptome entier. Ces études ont montré qu'une majeure partie du génome, chez l'homme, était transcrite [1] et que la proportion et le rôle des ARN ne codant (ncARN) pas pour des protéines étaient largement sous estimés [2].

La cellule, un monde ARN ?

Il est admis que de multiples classes d'ARN non codants (ncARN) participent aux mécanismes essentiels (voire primitifs) de la machinerie cellulaire. Ces ARN interviennent notamment lors de la traduction des ARN messagers (ARNm) en protéines grâce aux ARN ribosomiaux et aux ARN de transfert. Le contrôle de l'épissage implique, entre autres, les petits ARN nucléaires

(*small nuclear*). Plus récemment, l'inactivation spécifique de l'expression de certains gènes par les microARN (miARN) a été mise en évidence [3, 11] (→).

Les connaissances sur le nombre, la localisation et le rôle des longs ARN non codants (lncARN) restent partielles, contrairement à celles sur les ncARN de petites tailles. On définit arbitrairement les lncARN comme ayant une taille supérieure à 200 nucléotides et ne codant pas pour des protéines [10]. Cette définition est généralement affinée en fonction de la localisation du lncARN, c'est à dire intergénique ou bien chevauchant (sens ou anti-sens) un (ou des) gène(s) codant(s) pour des protéines. De plus, la quantité de grands transcrits non codants annotés dans le génome humain est certes en constante expansion, mais reste assez variable selon les bases de données (de 4 000 à plus de 10 000 [4]).

(→) Voir l'article de L. Amrouche et al., page 398 de ce numéro

Grâce à la mise à disposition croissante de séquences caractérisant des transcriptomes complets (RNA-Seq)¹, on peut anticiper que ces estimations sont bien loin du nombre réel chez l'homme. Par exemple, une étude exhaustive visant à séquencer plusieurs transcriptomes de souris a démontré que le nombre de lncARN murins se situait autour de 35 000 [5].

À l'instar des 21 000 gènes humains environ codant pour des protéines et dont les fonctions biologiques ne sont pas encore toutes connues, le rôle spécifique du répertoire des lncARN reste à élucider. Cependant, des travaux pionniers sur quelques lncARN permettent de les impliquer dans une régulation de la transcription du génome [10], souvent par leur modification de la structure de la chromatine. Ainsi, le lncARN *Xist*, d'une

¹ RNA-Seq : technique de séquençage haut-débit qui permet d'analyser qualitativement et quantitativement une population entière d'ARN dans un échantillon.

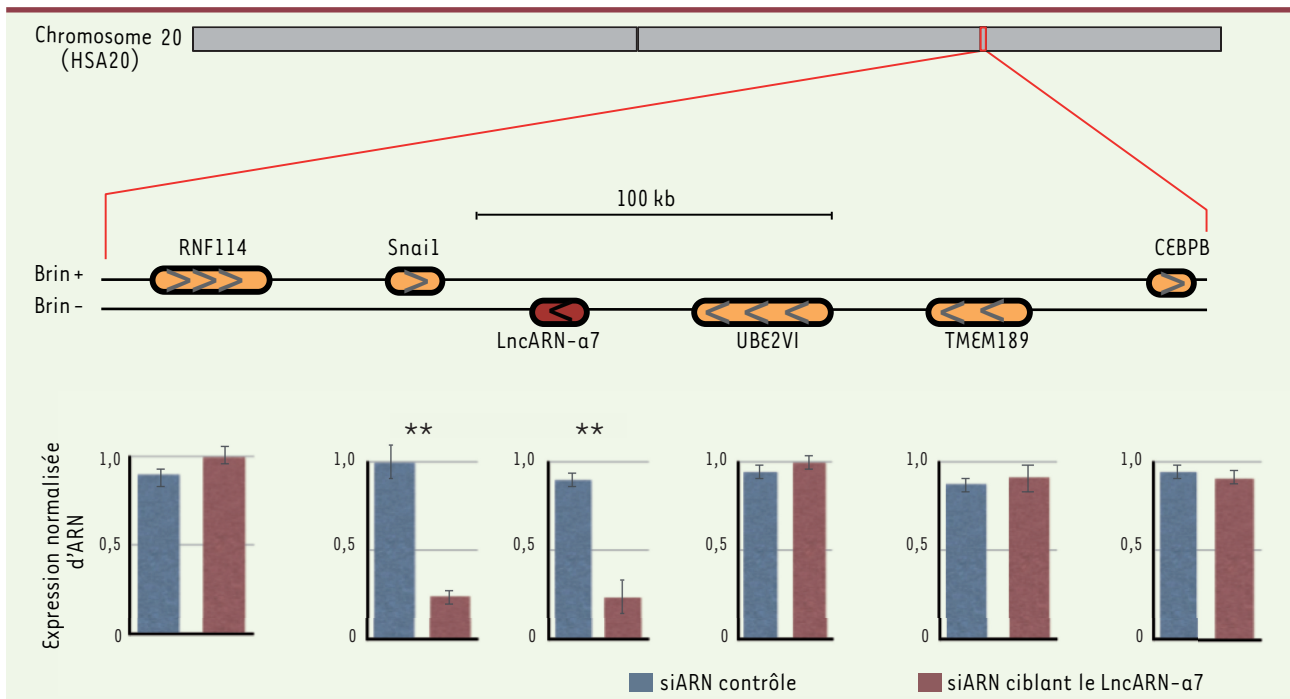


Figure 1. La région chromosomique du chromosome 20 humain (HSA20) aux environs de la position 48Mb est agrandie. (A) Ce locus, schématisé entre les pointillés en rouge, contient le long ARN non codant (LncARN-activateur (a) 7, ici en rouge) ainsi que cinq gènes codant pour des protéines (en orange). Les sens de transcription sont représentés par les flèches et par les brins plus et moins. **(B)** L'inactivation (*knock-out*) du LncARN-a7 par siARN provoque la répression spécifique du gène *Snai1* dont la séquence codante est située en amont sur le brin opposé. L'expression des autres gènes n'est pas altérée. Six expériences indépendantes ont été réalisées ici ; ** $p < 0,01$ (test de Student).

taille d'environ 19 kb, contrôle l'inactivation d'un des deux chromosomes X durant la différenciation précoce des cellules souches embryonnaires chez les placentaires femelles [12] (→). Plus récemment, une classe de lncARN intergéniques a été mise en évidence chez la souris par une approche de caractérisation des patrons de méthylation de la chromatine renforçant le lien entre lncARN, épigénétique et niveau d'expression [6]. Un autre exemple de lncARN concerne *HotAIR* qui peut réprimer la transcription du locus *HOXD* en modifiant l'état de méthylation de la chromatine et donc l'expression de ce locus.

(→) Voir m/s juin-juillet 2008, page 584

Identifier et caractériser les longs ARN non codants

Dans le but d'analyser à grande échelle la fonction des lncARN, nous avons utilisé l'annotation Gencode [4] du génome humain, produite dans le cadre du projet ENCODE (*Encyclopedia of DNA elements*) [1], qui vise à identifier tous les éléments fonctionnels du

génom. Pour cela, une combinaison d'outils bioinformatiques, de données expérimentales et de vérifications réalisées par des experts en annotation a permis d'identifier les régions transcrites du génome humain par l'alignement de séquences exprimées totales (ADNc) ou partielles (EST, *expressed sequence tags*) ou de transcriptomes complets. Ensuite, leur définition est affinée en les classant par catégories fonctionnelles : les gènes codant pour des protéines, les gènes non codants, les pseudogènes ou encore les gènes polymorphiques (dont la séquence codante varie en fonction des polymorphismes entre individus).

En utilisant cette annotation, nous avons identifié 3 019 nouveaux transcrits non codants ayant une taille moyenne de 800 nucléotides et exclusivement localisés entre les régions de gènes codant pour des protéines [7]. Ces lncARN ont certaines caractéristiques similaires aux gènes codant pour des protéines puisqu'ils sont épissés et que la moitié d'entre eux possèdent au moins un intron. Dans cette étude,

seulement le tiers du génome couvert par l'annotation Gencode a pu être analysé, ce qui suppose, par projection, que le génome humain posséderait près de 10 000 longs ncARN intergéniques.

Contrairement aux petits ncARN (tels que les miARN), la conservation des séquences des lncARN mesurée grâce à des alignements multiples de 44 génomes de vertébrés [8] apparaît relativement faible ce qui rend leur identification par les approches de similitude de séquences (identification d'homologie) très délicate. Le faible niveau de conservation observé peut refléter une plus grande adaptabilité de ces séquences au cours de l'évolution par acquisition de mutations, à l'opposé des fortes contraintes sélectives qui s'exercent sur les séquences codant pour des protéines. Néanmoins, la conservation de séquences des lncARN, et plus particulièrement de celles de leurs promoteurs, est statistiquement plus forte que celle de régions du génome prises aléatoirement, ce qui renforce la probabilité d'un rôle fonctionnel de ces lncARN.



En utilisant des données de séquençage de 15 transcriptomes issus de 10 tissus et 5 lignées cellulaires [9], il s'avère que plus de 75 % des 3 019 lncARN sont exprimés dans au moins un tissu et cette expression semble spécifique du tissu analysé. Une caractéristique importante des lncARN concerne leur niveau d'expression qui est 10 à 20 fois plus faible que celui des transcrits produits à partir des gènes codant pour des protéines. Ceci souligne à nouveau la nécessité de séquençages en profondeur pour l'identification de transcrits faiblement exprimés.

Mise en évidence du rôle activateur de la transcription des longs ncARN.

La plupart des fonctions des lncARN décrites dans l'inactivation du chromosome X ou dans l'empreinte parentale (*imprinting*) ont mis en évidence leur rôle dans la répression de l'expression des gènes voisins codant pour des protéines. Afin de tester cette hypothèse, nous avons analysé la corrélation d'expression entre les lncARN et tous les gènes codants dans les 15 échantillons (provenant de tissus et cellules) mentionnés ci-dessus. De façon surprenante, nous avons trouvé que l'expression des lncARN est essentiellement corrélée positivement avec l'expression des gènes codant pour des protéines et que cette corrélation est plus forte avec des gènes voisins plutôt qu'avec des gènes distants sur le génome. Pour tester expérimentalement ces résultats, des petits ARN interférents (siARN) ciblant exclusivement les lncARN ont été utilisés. Un premier lncARN candidat (ncARN- $\alpha 7$) a été spécifiquement réprimé par l'utilisation de siARN. Nous avons ensuite évalué les conséquences de cette inactivation en

mesurant le niveau d'expression des gènes situés dans une fenêtre de 1 Mb autour du lncARN (Figure 1). Comme prédit au niveau bioinformatique, l'inactivation du lncARN entraîne une répression concomitante de l'expression d'un gène voisin *Snail*, ce qui implique un rôle activateur du ncARN- $\alpha 7$ sur la transcription du gène *Snail*, dont l'expression a un rôle dans l'adhésion et la migration cellulaires. Nous avons répété ces expériences d'interférence pour plusieurs lncARN candidats et suggérons que ce mécanisme d'activation de la transcription serait généralisable. Les gènes activés par les lncARN ont des fonctions variées. Parmi ces gènes, *ECM1* est impliqué dans la constitution de la matrice extra-cellulaire, *KLHL12* (codant une E3 ubiquitine ligase) est un régulateur négatif de la voie Wnt bêta-caténine et *SCL/TALI* (*stem cell leukemia*, ou *T cell acute leukemia*) est un facteur important de la régulation de l'hématopoïèse.

Le génome humain possède donc une forte proportion de longs ARN non codants pour des protéines. Une nouvelle classe de lncARN, certes modestement conservés et faiblement exprimés, régule l'activation de la transcription de gènes voisins au même titre que les régions activatrices de la transcription (*enhancer*). Cependant, à l'inverse des séquences *enhancer* qui activent la transcription par fixation de facteurs de transcription à l'ADN, c'est bien l'ARN qui est ici l'élément central de l'activation de la transcription même si le mécanisme d'activation reste à préciser. Une hypothèse intéressante serait que le lncARN servirait de plate-forme par similitude de séquence avec son gène cible pour favoriser le recrutement de protéines activatrices de

la transcription. L'achèvement de l'annotation Gencode pour la totalité du génome humain permettra d'identifier l'ensemble du répertoire des lncARN et ainsi de préciser leurs rôles fonctionnels. ♦

Long non-coding RNAs with enhancer-like function in human cells

CONFLIT D'INTÉRÊTS

Les auteurs déclarent n'avoir aucun conflit d'intérêts concernant les données publiées dans cet article.

RÉFÉRENCES

1. ENCODE Projet Consortium, Birney E, Stamatoyannopoulos JA, et al. Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. *Nature* 2007 ; 447 : 799-816.
2. Amaral PP, Dinger ME, Mercer TR, Mattick JS. The eukaryotic genome as an RNA machine. *Science* 2008 ; 319 : 1787-9.
3. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 1993 ; 75 : 843-54.
4. Harrow J, Denoeud F, Frankish A, et al. GENCODE : producing a reference annotation for ENCODE. *Genome Biol* 2006 ; 7 Suppl 1 : S4.1-9.
5. Carninci P, Kasukawa T, Katayama S, et al. The transcriptional landscape of the mammalian genome. *Science* 2005 ; 309 : 1559-63.
6. Guttman M, Amit I, Garber M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 2009 ; 458 : 223-7.
7. Ørom UA, Derrien T, Beringer M, et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell* 2010 ; 143 : 46-58.
8. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005 ; 15 : 1034-50.
9. Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008 ; 456 : 470-6.
10. Pasmant E, Laurendeau I, Sabbagh A, et al. ANRIL ou l'étrange histoire d'un grand ARN non codant. *Med Sci (Paris)* 2010 ; 26 : 564-6.
11. Amrouche L, Bonifay R, Anglicheau D. MicroARN et maladies rénales : un intérêt grandissant. *Med Sci (Paris)* 2011 ; 27 : 398-404.
12. Augui S, Heard E. Inactivation du chromosome X : comment une cellule sait compter jusqu'à 2X. *Med Sci (Paris)* 2008 ; 24 : 584-5.



Tarifs d'abonnement m/s - 2011

Abonnez-vous
à médecine/sciences

> Grâce à m/s, vivez en direct les progrès
des sciences biologiques et médicales

Bulletin d'abonnement
page 368 dans ce numéro de m/s

