

Titre du projet de recherche

Analyse des longs ARN non-codants dérivés d'éléments transposables dans le placenta canin

Tuteurs du stage

Anamaria Necsulea (anamaria.necsulea@univ-lyon1.fr)

Thomas Derrien (thomas.derrien@univ-rennes1.fr)

Laboratoire d'accueil

Equipe Bioinformatique, Phylogénie et Génomique Evolutive

Laboratoire de Biométrie et Biologie Evolutive

UMR CNRS 5558, Université Claude Bernard – Lyon 1

<http://lbbe.univ-lyon1.fr/>

Description du projet

Les génomes eucaryotes sont transcrits en une multitude de molécules d'ARN, telles que les ARN messagers (ARNm) qui codent pour des protéines, mais aussi les ARNs non-codants, qui ne sont pas traduits. Parmi ces derniers, la classe des longs ARNs non-codants (lncARN) est particulièrement intéressante, car elle comprend plusieurs molécules avec d'importants rôles fonctionnels. Nous pouvons citer par exemple *Xist*, un lncARN qui est responsable de l'inactivation du chromosome X chez les mammifères (Brown et al. 1991), ou encore *H19*, qui participe au contrôle du développement du placenta (Brannan et al. 1990). Chez les mammifères, des dizaines de milliers de lncARN ont été découverts récemment grâce à l'analyse de données de séquençage de transcriptome à haut débit (Derrien et al. 2012). Même si des approches de transcriptomique comparative inter-espèce ont permis de mieux comprendre l'évolution de ces lncARN (Necsulea et al. 2014; Darbellay and Necsulea 2020), la très grande majorité de ces lncARN n'ont pas encore été caractérisés fonctionnellement et leurs rôles biologiques sont débattus.

Contrairement aux ARNm, les séquences des lncARN des mammifères sont en grande partie constituées de séquences répétées, et plus particulièrement d'éléments transposables (Kapusta et al. 2013). Les éléments transposables fournissent également de nombreux motifs régulateurs de la transcription ou de l'épissage, et participent ainsi de plusieurs manières à la biogénèse des lncARN (Kapusta et al. 2013). L'association entre lncARN et éléments transposables pourrait être adaptative, si les lncARN qui en dérivent étaient fonctionnels. Cette association pourrait également apparaître suite à un processus évolutif non-adaptatif : la transcription des lncARN à partir d'éléments transposables pourrait être une simple conséquence de l'envahissement des génomes par ces séquences « égoïstes ». À ce jour, la question de la fonctionnalité des lncARN dérivés d'éléments transposables reste ouverte.

Dans le cadre de ce projet, nous proposons d'étudier les lncARN dérivés d'éléments transposables dans le placenta. Ce tissu est particulièrement adapté pour cette étude car il fournit de bons exemples de « domestication » des éléments transposables. Par exemple, il est maintenant établi que les éléments transposables de la classe des rétrovirus endogènes sont transcrits et peuvent activer l'expression des gènes dans le placenta (Chuong et al. 2013). D'autres éléments transposables de ce type ont donné naissance aux syncytines, des protéines qui sont importantes pour la fusion entre le tissu maternel et le tissu embryonnaire dans le placenta (Mi et al. 2000).

Nous proposons d'étudier plus particulièrement les lncARN exprimés dans le placenta canin. Le chien est un excellent modèle pour étudier les lncARN et les éléments transposables, car son histoire évolutive comporte un épisode récent de domestication, il y a environ 15,000 ans, suivi par une période intense de sélection artificielle (Frantz et al. 2020). Cette histoire évolutive particulière a mené à la fixation de nombreux éléments transposables dans le génome du chien, à cause de la diminution de l'efficacité de la sélection naturelle. Le génome du chien possède donc la « matière première » pour produire de nombreux lncARN dérivés d'éléments transposables.

Organisation du projet

En pratique, ce projet repose sur l'analyse de données de séquençage de transcriptome (RNA-seq) de placenta canin. Nous proposons d'analyser un jeu de données publiques, comprenant 9 échantillons de RNA-seq (Nowak et al. 2019). En utilisant ce jeu de données, ainsi que les annotations existantes du génome du chien, il sera possible de prédire les transcrits qui sont actifs dans le placenta (Perteza et al. 2015). Nous pourrons ensuite annoter ces transcrits pour distinguer les lncARN des ARNm en utilisant FEELnc (Wucher et al. 2017). Nous proposons ensuite d'identifier les gènes dont la séquence transcrite est dérivée en partie d'éléments transposables, en utilisant les annotations de ces derniers obtenues avec RepeatMasker (Smit et al. 2003). Enfin, nous réaliserons une analyse comparative entre le génome de référence du chien (boxer) et les génomes d'autres races canines ainsi que ceux de canidés, qui nous permettra de définir l'origine évolutive et les pressions de sélection qui agissent sur les lncARN. Ce projet pourra ainsi fournir une première estimation de la fonctionnalité des lncARN dérivés des éléments transposables dans le placenta canin.

Prérequis

Utilisation des programmes en ligne de commande sous Linux ; connaissance des langages de programmation Python et R.

Bibliographie

- Brannan CI, Dees EC, Ingram RS, Tilghman SM. 1990. The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.* 10:28–36.
- Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349:38–44.
- Chuong EB, Rumi MAK, Soares MJ, Baker JC. 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* 45:325–329.
- Darbellay F, Necsulea A. 2020. Comparative transcriptomics analyses across species, organs, and developmental stages reveal functionally constrained lncRNAs. *Mol. Biol. Evol.* 37:240–259.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22:1775–1789.
- Frantz LAF, Bradley DG, Larson G, Orlando L. 2020. Animal domestication in the era of ancient genomics. *Nat. Rev. Genet.* 21:449–460.

- Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* 9:e1003470.
- Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, et al. 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403:785–789.
- Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Grutzner F, Kaessmann H. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505:635–640.
- Nowak M, Rehrauer H, Ay SS, Findik M, Boos A, Kautz E, Kowalewski MP. 2019. Gene expression profiling of the canine placenta during normal and antigestagen-induced luteolysis. *Gen. Comp. Endocrinol.* 282:113194.
- Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33:290–295.
- Smit AFA, Hubley R, Green P. 2003. RepeatMasker Open-4.0. Available from: <http://www.repeatmasker.org>
- Wucher V, Legeai F, Hédan B, Rizk G, Lagoutte L, Leeb T, Jagannathan V, Cadieu E, David A, Lohi H, et al. 2017. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 45:e57.