



## HABILITATION A DIRIGER DES RECHERCHES

Date de la soutenance : **31 janvier 2024**

Nom de famille et prénom de l'auteur : **Madame NECSULEA Anamaria**

Titre des travaux : « *Evolution et fonctionnalité des éléments non-codants dans les génomes des vertébrés* »

### Résumé



Un objectif majeur de la biologie aujourd'hui est de construire des cartes fonctionnelles des génomes, en identifiant les régions du génome qui participent aux fonctions biologiques, et en liant chaque séquence d'ADN à ses rôles moléculaires et aux processus cellulaires auxquels elle participe. Dans ce contexte, l'étude de l'évolution apporte une source importante d'informations pour éclairer la fonctionnalité des séquences d'ADN. Retracer l'histoire évolutive des génomes permet d'identifier les éléments génomiques qui sont soumis à des pressions de sélection purificatrice pour préserver leurs fonctions, ou à des pressions de sélection positive suite à l'acquisition de nouvelles fonctions. Cela permet également de prédire les séquences génomiques qui évoluent de manière neutre, et qui représentent probablement des régions dispensables d'un point de vue fonctionnel. Avec cette motivation, **l'objectif principal de ma recherche jusqu'à présent a été d'étudier la fonctionnalité des génomes avec une approche évolutive.**

Cette approche évolutive pour l'étude de la fonctionnalité des séquences d'ADN est particulièrement utile dans le cas des éléments non-codants (qui ne codent pas pour des protéines). Dans les génomes des organismes complexes, tels que les vertébrés, seule une faible fraction du génome est traduite en protéines. De nombreuses régions génomiques non-codantes sont impliquées dans des fonctions biologiques essentielles, telles que la réplication de l'ADN, la recombinaison, ou encore la régulation de l'activité des gènes. Pour d'autres régions non-codantes, il existe des preuves d'activité biochimique, mais leurs fonctions biologiques ne sont pas avérées. Cette situation est fréquente notamment pour les *loci* qui produisent des longs ARNs non-codants (lncARNs), qui sont des molécules d'ARN que l'on détecte facilement grâce aux techniques de séquençage de transcriptome à haut débit. Des fonctions biologiques ont été démontrées seulement pour une très faible proportion des lncARNs qui ont été identifiés jusqu'à présent.

Dans le premier chapitre de cet ouvrage, je vais présenter mes travaux portant sur **l'évolution et la fonctionnalité des lncARNs**. Au cours de mon séjour postdoctoral, j'ai contribué à l'annotation de lncARNs dans plusieurs espèces de vertébrés et j'ai réalisé une analyse évolutive à grande échelle de ces transcrits, en exploitant des données de séquençage de transcriptome produites pour plusieurs organes dans chaque espèce. J'ai ainsi pu démontrer que les répertoires des lncARNs évoluaient très rapidement chez les vertébrés, avec uniquement une centaine de lncARNs (parmi les dizaines de milliers de lncARNs identifiés pour chaque espèce) partagés entre toutes les espèces étudiées [1]. J'ai également montré que les séquences et les patrons d'expression des lncARNs évoluaient beaucoup plus vite que ceux des gènes codants pour des protéines. Ces résultats suggèrent fortement que les lncARNs sont soumis à moins de pressions de sélection que les gènes codant pour des protéines, et donc qu'une partie non-négligeable des lncARNs pourrait représenter uniquement du « bruit » transcriptionnel. J'ai continué ces travaux en rajoutant une dimension temporelle à l'analyse comparative, en examinant les répertoires des

IncARNs actifs à plusieurs stades de développement dans des organes majeurs du rat et de la souris [2]. Cette étude a permis de démontrer que les IncARNs actifs lors des stades précoces du développement embryonnaire étaient plus conservés au cours de l'évolution que les IncARNs actifs dans les organes adultes. Nous avons également observé que des contraintes fonctionnelles plus fortes agissaient sur certaines parties des *loci* des IncARNs, notamment les promoteurs et les éléments impliqués dans la régulation de l'épissage. Ces résultats étaient en accord avec des hypothèses proposées récemment, selon lesquelles les *loci* produisant des IncARNs pourraient porter d'autres fonctions biologiques indépendantes de la production de la molécule d'ARN, par exemple à travers la présence d'éléments régulateurs de l'expression des gènes au niveau des promoteurs. En parallèle de l'approche évolutive que j'ai employée pour explorer la fonctionnalité des IncARNs à grande échelle, j'ai participé à des études qui visaient à tester au laboratoire les fonctions prédites pour certains de ces *loci* [3, 4]. En particulier, j'ai participé à une étude dont l'objectif était de tester la fonction proposée pour le IncARN *Hotair* dans le développement du squelette axial et des membres chez la souris. En comparant les phénotypes et les patrons d'expression des souris sauvages et des souris comportant une délétion de *Hotair*, nous avons pu démontrer que ce locus était dispensable pour le développement de la souris. Nous avons également mis en évidence que la délétion du locus produisant ce IncARN avait des conséquences indépendantes de l'absence de la molécule d'ARN de *Hotair*, notamment la production d'un transcrit antisens et la présence d'une région non-traduite étendue pour le gène codant voisin. Ces résultats renforcent l'hypothèse de la présence de fonctions indépendantes de la molécule d'ARN pour les *loci* produisant des IncARNs.

Dans le deuxième chapitre de ce mémoire, je vais présenter des projets plus récents, portant sur **l'évolution des éléments non-codants impliqués dans la régulation de l'expression des gènes**, ainsi que sur l'évolution des structures tridimensionnelles du génome dans lesquelles ils interviennent. J'ai commencé à m'intéresser aux éléments non-codants régulateurs de l'expression des gènes à la suite des travaux que j'ai effectués sur les IncARNs, pour lesquels on propose également souvent des rôles dans le contrôle de l'activité des gènes. Je me suis particulièrement intéressée aux éléments *cis*-régulateurs (c'est-à-dire, aux séquences régulatrices non-codantes qui sont situées sur le même chromosome que les gènes qu'elles contrôlent) situées à longue distance par rapport aux gènes cibles. Ces relations régulatrices à longue distance sont fréquentes dans les génomes des vertébrés, et elles peuvent être détectées grâce aux techniques de capture de conformation de la chromatine (Hi-C et techniques apparentées). Dans le cadre de la thèse d'Alexandre Laverre, que j'ai co-encadré entre 2018 et 2022, nous nous sommes intéressés à l'évolution de ces relations régulatrices, en lien avec l'évolution de l'expression des gènes. Après avoir confirmé que les contacts de chromatine centrés sur les promoteurs des gènes permettaient d'améliorer la prédiction des éléments régulateurs de l'expression des gènes, nous avons montré que ces contacts de chromatine étaient plus conservés qu'attendu par hasard entre l'humain et la souris [5]. Nous avons également démontré que la vitesse d'évolution des contacts de chromatine entre promoteurs des gènes et éléments régulateurs était positivement corrélée à la vitesse d'évolution des patrons d'expression des gènes. Enfin, nous avons trouvé que les réarrangements génomiques à grande échelle pouvaient influencer de manière significative la régulation de l'expression des gènes chez les vertébrés, en perturbant les relations régulatrices à longue distance.

Dans le troisième chapitre du mémoire, je vais présenter mes projets de recherche à moyen terme, qui combinent des idées qui ont émergé lors des études des IncARNs et des contacts de chromatine à longue distance entre promoteurs des gènes et éléments régulateurs non-codants. Notamment, suite aux travaux que nous avons effectués sur les influences croisées entre réarrangements génomiques et structure des contacts de chromatine à rôle régulateur, j'aimerais explorer le lien entre réarrangements génomiques et la perte de fonction des gènes. L'objectif de ce projet serait d'évaluer la fréquence des pertes de fonction des gènes suite à des perturbations drastiques de leurs paysages régulateurs. Des résultats préliminaires ont été obtenus dans le cadre des stages de Master de Thomas Lahaie et Victor Lefebvre, qui ont exploré cette hypothèse en analysant la distribution des pertes des gènes ayant eu lieu chez l'ancêtre des mammifères placentaires. Je vais également présenter un projet de recherche collaboratif visant à explorer de manière conjointe l'évolution de l'expression des gènes, des contacts de chromatine à rôle régulateur et des phénotypes, chez les oiseaux.

#### Références :

- [1] [A. Necsulea](#), M. Soumillon, M. Warnefors, A. Liechti, T. Daish, U. Zeller, J.C. Baker, F. Grützner, H. Kaessmann (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, 505:635-640.
- [2] F. Darbellay, [A. Necsulea](#) (2020). Comparative transcriptomics analyses across species, organs and developmental stages reveal functionally constrained lncRNAs. *Molecular Biology and Evolution*, 37:240-259.

- [3] A.R. Amândio, [A. Necșulea](#), E. Joye, B. Mascrez, D. Duboule (2017). *Hotair* is dispensable for mouse development. *PLoS Genetics*, 12:e1006232.
- [4] J. Zakany, F. Darbellay, B. Mascrez, [A. Necșulea](#), D. Duboule (2017). Control of growth and gut maturation by *HoxD* genes and the associated lncRNA *Hagl1*. *Proceedings of the National Academy of Science USA*, 114:E9290-E9299.
- [5] A. Laverré, E. Tannier, [A. Necșulea](#) (2022). Long-range promoter-enhancer contacts are conserved during evolution and contribute to gene expression robustness. *Genome Research*, 32:280-296.