

Rapport d'activité (janvier 2014 – septembre 2016)

Hervé Seitz

Table des matières

1	<i>Curriculum vitæ</i>	2
1.1	Informations générales	2
1.2	Études et expérience post-doctorale	2
1.3	Publications dans des revues à comité de lecture	2
1.4	Collaborations nationales et internationales	2
1.5	Autres activités professionnelles, activités d'expertise	3
1.6	Activités non professionnelles	3
2	Recherche scientifique	4
2.1	Sensibilité des gènes-cibles vis-à-vis des miARN	4
2.2	Les gènes dose-sensibles portent les sites de fixation aux miARN les mieux conservés	6
2.3	Mesure de la titration des miARN par les ARNm	7
2.4	La titration par les ARNm affine le patron d'expression des miARN	10
2.5	Des faux positifs dans les analyses de génomique comparative	12
2.6	Projet terminé : évolution du mode de répression par les miARN chez les Animaux	14
2.7	Collaborations diverses	14
2.8	Références	16
2.9	Place de ma recherche dans l'unité	17
2.10	Distinctions scientifiques	18
2.11	Publications scientifiques sur les 5 derniers semestres	18
3	Enseignement, formation et diffusion de la culture scientifique	19
3.1	Thèses et post-doctorats	19
3.2	Participation à l'enseignement	19
3.3	Diffusion de l'information scientifique	20
3.4	Interventions dans la presse écrite et audiovisuelle	20
3.5	Participation à des travaux d'expertise	20
4	Transfert technologique, relations industrielles et valorisation	22
5	Encadrement, animation et management de la recherche	23

1 *Curriculum vitæ*

1.1 Informations générales

Né le 13 décembre 1977 à Angers (Maine-et-Loire).

Adresse professionnelle : IGH (UPR 1142 du CNRS), 141 rue de la Cardonille, 34396 Montpellier.

email : herve.seitz@igh.cnrs.fr ; site web de l'équipe : <http://www.igh.cnrs.fr/FR/equipe-detail.php?id=72>

1.2 Études et expérience post-doctorale

- 1997 : admis aux trois Écoles normales supérieures, dont l'É.N.S. de la rue d'Ulm avec le rang : 2.
- 2001-2004 : thèse : « Empreinte génomique parentale et petits ARN non-codants » (université Toulouse III Paul Sabatier ; directeur de thèse : Jérôme Cavallé). Manuscrit : <http://www.normalesup.org/~seitz/These.pdf>
- Janvier 2005-janvier 2009 : stage post-doctoral à University of Massachusetts Medical School (Worcester, Massachusetts, États-Unis) ; directeur des travaux : Phillip D. Zamore.
- 2008 : classé 1^{er} au concours de CR2 de la section 21 du C.N.R.S. ; poste pris en janvier 2009.
- 3 novembre 2009 : Habilitation à diriger les recherches. Manuscrit : <http://www.normalesup.org/~seitz/HDR.pdf>
- 2010 : lauréat du Career Development Award (CDA n°00017-2010-C) de Human Frontier Science Program.
- 2010 : classé 1^{er} pour le recrutement d'un jeune chef d'équipe à l'IGH (Montpellier) ; installation en octobre 2011.
- 2010 : Prime d'excellence scientifique du CNRS.
- 2011 : lauréat du programme ATIP-Avenir (contrat initié en mars 2012).
- 2013 : promu CR1 par le CNRS.
- 2014 : Prime d'excellence scientifique du CNRS.

1.3 Publications dans des revues à comité de lecture

Au moins une chaque année depuis 2000 inclus, à l'exception des années 2001, 2005 et 2012.

Liste complète sur http://www.igh.cnrs.fr/equip/Seitz/equipe-publis_par_auteur.html#Publis_Herve.

1.4 Collaborations nationales et internationales

- Depuis 2009 : avec Y. Tomari (University of Tōkyō, Japon) ; publications communes : Kawamata *et al.* (2009) et Tsutsumi *et al.* (2011).
- Depuis 2010 : avec U. Technau (University of Vienna, Autriche), F. Rentzsch (SARS Center, Norvège) et P.D. Zamore (UMass Medical School, États-Unis) ; publication commune : Moran *et al.* (2014).
- Depuis 2012 : avec J. Turner (Crick Institute, Royaume-Uni) ; publication commune : Royo *et al.* (2015).
- Depuis 2012 : avec D. Tagu (INRA Rennes, France) ; financement commun : ANR "MiRNAdapt", ANR-11-BSV6-01701 (Programme Blanc 2011 ; financement : 2012-2015).
- Depuis 2012 : avec F. Apparailly (Inserm Montpellier, France) ; publication commune : en révision à *Genome Research* (soumission initiale le 4 février 2016, révision soumise le 22 août 2016).
- Depuis 2013 : avec M. Méchali (CNRS Montpellier, France) ; publication commune : soumise à *Nature Structural and Molecular Biology* le 26 août 2016.
- Depuis 2013 : avec M. Cock (CNRS Roscoff, France) ; publication commune : Tarver *et al.* (2015).
- Depuis 2013 : avec G. Rappold (University of Heidelberg, Allemagne) ; publication commune : Hoffmann *et al.* (2016).
- Depuis 2014 : avec A. Spradling (Carnegie Institution, États-Unis) ; publication commune en préparation.
- Depuis 2014 : avec E. Levashina (Max Planck Institute for Infection Biology, Allemagne) ; publication commune en préparation.

1.5 Autres activités professionnelles, activités d'expertise

Monitorat à l'Université Toulouse III Paul Sabatier (2001-2004) en biologie moléculaire et en génétique. Cours magistraux sur les petits ARN régulateurs (d'un à trois cours par an depuis 2007 ; aux universités de Toulouse, Paris, Lyon, Rennes, Taipei, Montpellier ; à l'École normale supérieure : cours de L3 en 2010, de L3 en 2014, de L3 en 2015, de L3 puis de M1 en 2016).

Outre les activités de vulgarisation, je tâche d'aider mes collègues biologistes par des formations, formelles ou informelles, dans des domaines qui sont très utiles aux biologistes, mais encore mal connus (la bio-informatique et les statistiques). Je pense que, dans l'intérêt des scientifiques eux-mêmes, la pratique de notre métier devrait être coopérative ; c'est pourquoi je partage systématiquement les programmes que j'écris avec les collaborateurs qui ont sollicité mon aide, et je leur apprends à s'en servir. Je mets en accès libre les formations que je donne à mes collègues, pour qu'elles servent au plus grand nombre.

Formations dispensées aux chercheurs, doctorants et post-doctorants :

- « Initiation à Unix pour les biologistes » en 5 leçons : dispensée à l'University of Massachusetts Medical School (2008), à l'IFR 109 (Toulouse) (2009), à l'IGH (Montpellier) (2013 et 2014).
- « Initiation aux statistiques pour les biologistes, initiation à R » en 5 leçons : dispensée à l'University of Massachusetts Medical School (2010), à l'IFR 109 (Toulouse) (2010), à l'IGH (Montpellier) (2010 et 2014).

Les diaporamas de ces formations sont accessibles sur <http://www.igh.cnrs.fr/equip/Seitz/equipe-tutoriels.html>.

Membre du jury de recrutement d'un maître de conférence à l'université Toulouse III Paul Sabatier (poste 64MCF0300, en 2013). Membre de 17 jurys de thèse, pour des étudiants d'institutions françaises et étrangères (*Cold Spring Harbor Laboratory, Caltech, university of Queensland, ...*). Membre d'un jury d'HDR (Séverine Chambeyron, Montpellier ; 2011).

Rapporteur pour de nombreux journaux dans le domaine de l'ARN, de la bio-informatique et de la génétique moléculaire (*Current Biology, EMBO Reports, Genome Research, Nucleic Acids Research, RNA Biology, ...*).

Rapporteur pour l'ANR (2012 : panel SVSE2 ; 2014 : panel CE10 ; 2016 : panel CE12), l'Idex Sorbonne universités (2014), l'IUF (2015), la FRM Alsace (2011), le GIS IBiSA (2010).

Activité de consulting pour Medesis Pharma S.A. (compagnie spécialisée dans le développement de vecteurs pour l'adressage *in vivo* de composés actifs ; je les ai aidés à choisir des siRNA contre des gènes d'intérêt pour répression *in vivo* chez la Souris) : de mars à mai 2012.

Membre du comité de sélection des projets de la plateforme ABIC de Montpellier (depuis 2016).

Membre de la "Governance Leadership Team" du *Cefic Long-range research initiative* pour les domaines "omics, epigenetics and non-coding RNAs" pour l'amélioration des procédures d'évaluation des risques toxicologiques (depuis 2016).

Interventions dans les médias : *Néon* (2014), *Le Monde* (2015), <http://www.atlantico.fr> (2015, 2016). Articles de vulgarisation : *Pour la Science* (2005 puis 2013).

1.6 Activités non professionnelles

Ultra-marathon : médaille de bronze au championnat de France 2012 de course de 100 km, 3ème aux 100 km de Millau 2012, 2ème aux 100 km de Millau 2013 et 100 km de Millau 2014, 3ème aux 100 km de Belvès 2015, vainqueur des 100 km de Millau 2015, vainqueur des 100 km de Belvès 2016.

2 Recherche scientifique

Mon principal sujet de recherche concerne l'impact physiologique de la régulation guidée par les microARN (miARN). Depuis une quinzaine d'années, la communauté scientifique a attribué à ces régulateurs de très nombreuses fonctions, sur la base d'expériences de biologie moléculaires¹ et d'analyses bio-informatiques². Chacune de ces méthodes suggère que les miARN régulent de très nombreuses cibles (plus de 60% des gènes codants : Friedman *et al.*, 2009). Il est donc difficile de trouver un processus biologique que les miARN ne sont pas soupçonnés de contrôler : quel que soit le processus auquel on s'intéresse, au moins un de ses composants sera codé par l'un des 60% de gènes ciblés. Sans surprise donc, ces dernières années ont vu fleurir des milliers d'études impliquant les miARN dans toutes sortes de phénomènes, en contexte sain ou pathologique.

Il nous a semblé étrange que les miARN puissent jouer un rôle aussi important : il est bien connu que l'effet des miARN sur leurs cibles est très modeste (ils répriment les cibles de moins d'un facteur 2 en général), or la plupart des gènes chez les Animaux sont robustes vis-à-vis de petits changements de niveau d'expression (notamment, la plupart des gènes chez les Animaux sont haplo-suffisants). Il est donc possible que ces nombreuses « cibles » identifiées, le plus souvent, sur la foi de gènes rapporteurs par des techniques de biologie moléculaire, soient trop peu réprimées pour être *fonctionnellement* affectées par les miARN. En d'autres termes : ces gènes sont effectivement réprimés à l'échelle moléculaire, mais cette petite répression à l'échelle microscopique ne suffirait pas à provoquer un phénotype macroscopique, elle serait tamponnée par les mécanismes d'homéostasie (Seitz, 2009). Il est important ici de noter que les prédictions informatiques de cibles de miARN sont généralement en très bon accord avec ces tests moléculaires, ce que la communauté a toujours considéré comme une validation solide des prédictions.

Les travaux de mon équipe visent donc principalement à explorer cette nuance entre une « cible » définie au niveau microscopique, et une véritable cible contrôlant un phénotype macroscopique.

2.1 Sensibilité des gènes-cibles vis-à-vis des miARN

Des individus de phénotype sauvage n'expriment pas tous leurs gènes au même niveau : d'un individu à l'autre, des fluctuations de niveau d'expression existent, qui ne se traduisent pas par un phénotype détrimental. Nous avons donc décidé de mesurer la variabilité d'expression des gènes entre individus sauvages, et de la comparer à l'amplitude de la répression par un miARN. Si, comme on l'a cru jusqu'ici, les miARN exercent une régulation fonctionnelle sur toutes leurs cibles, alors la variabilité inter-individus devrait rester inférieure à l'amplitude de la répression par le miARN.

Si au contraire des individus de phénotype sauvage présentent de larges fluctuations d'expression d'un gène donné (fluctuations plus amples que l'effet du miARN sur ce gène), alors il semblera peu probable que la répression de ce gène par le miARN puisse jouer un rôle biologique.

Pour cette expérience, nous avons étudié le miARN miR-223 (qui est spécifique du lignage granulocytaire) dans les neutrophiles de Souris (Chen *et al.*, 2004 ; Johnnidis *et al.*, 2008 ; Baek *et al.*, 2008). L'effet de ce miARN sur chacun des gènes de la Souris est connu grâce aux travaux de Baek *et al.*, 2008. Nous avons donc, de notre côté, mesuré les variations inter-individuelles dans l'expression des gènes chez les neutrophiles de 5 souris sauvages. Dans une expérience contrôle, nous avons mesuré le bruit technique de l'expérience en analysant de manière identique 5 réplicats techniques, issus d'un mélange homogène du sang de 5 autres souris (voir Figure 1A).

Sur les 196 cibles prédites de miR-223, 192 ont pu être analysées par notre expérience. Nos résultats ont montré que, parmi ces 192 cibles, 150 présentent de larges fluctuations entre individus de phénotype sauvage ($p < 0.05$, ce qui signifie que l'effet est réel et ne semble pas dû au bruit technique) (voir Figures 1B et C). Sur l'exemple de ce miARN, il apparaît donc que 80% de ses « cibles » connues sont fonctionnellement insensibles à l'effet du miARN : trop peu réprimées pour contrôler un phénotype macroscopique, elles ne sont probablement pas des cibles au sens fonctionnel. Ces résultats, s'ils se généralisent, pourraient remettre en cause de nombreux travaux de nos collègues.

1. Par exemple, en co-exprimant un miARN et un gène rapporteur portant la 3' UTR d'une cible présomptive, ou en immunoprécipitant les ARNm recrutés par le complexe effecteur des miARN.

2. La plupart des programmes de prédiction performants utilisent la même stratégie : ils cherchent des sites de complémentarité aux miARN (notamment à leur « graine » : nucléotides 2 à 7 du miARN), puis sélectionnent ceux qui ont été conservés dans l'évolution. En partant du principe que les séquences fonctionnellement importantes sont phylogénétiquement conservées, ces programmes cherchent à éliminer les sites d'interaction qui n'auraient pas d'effet biologique réel.

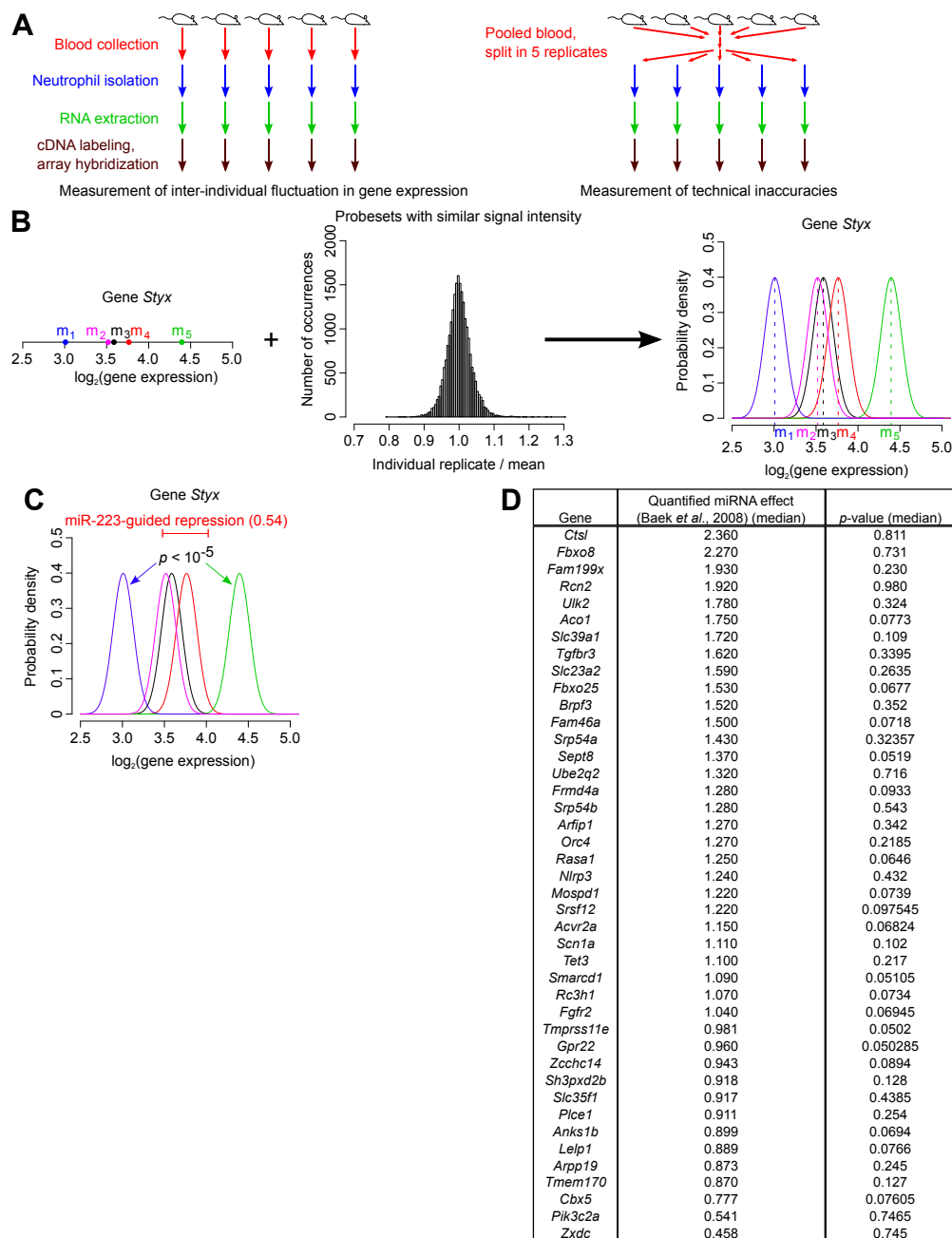


FIGURE 1 – La variabilité inter-individus dans l'expression des cibles de miR-223 dépasse fréquemment l'amplitude de la répression par miR-223. **A**. Principe de l'expérience. **B**. Le signal mesuré par microarray est la somme de la valeur biologique réelle et du bruit technique (ici illustré sur le cas du gène *Styx*). Les signaux mesurés (indiqués m_1 à m_5) sont déconvolués en utilisant nos mesures du bruit technique. **C**. Pour chaque cible prédite de miR-223, l'amplitude de la répression par miR-223 est comparée à l'amplitude de la variabilité d'expression parmi les cinq souris. La *p*-value mesure la probabilité que la variabilité biologique réelle soit plus petite que l'effet du miARN. Nous avons utilisé la répression médiane mesurée par Baek *et al.*, 2008 (représentée par la barre horizontale rouge) pour estimer l'effet répressif de miR-223. **D**. Les gènes dont les fluctuations d'expression entre individus ne sont pas significativement plus grandes que l'effet répressif de miR-223 ($p \geq 0.05$). Colonne du milieu : ampleur de la répression (rapport $\frac{\text{expression dans miR-223}^\Delta}{\text{expression dans wt}}$) d'après Baek *et al.*, 2008 (notez que certains gènes ont une répression < 1 , apparaissant donc activés, plutôt que réprimés, par le miARN : ces gènes pourraient être indirectement affectés par miR-223). Colonne de droite : *p*-value, mesurée comme indiqué dans le panel C (médiane parmi les sondes pour le gène en question).

D'autre part, il peut être intéressant d'analyser spécifiquement les 42 gènes restants : l'expression de ces gènes semble très peu variable d'un individu à l'autre, ce qui pourrait suggérer que ces gènes contrôlent des phénotypes très sensibles au niveau d'expression – en d'autres termes, il nous semble probable que les vraies cibles de miR-223 (celles qui contrôlent les phénotypes observés chez le mutant miR-223^Δ : Johnnidis *et al.*, 2008) sont enrichies parmi ces 42 gènes. Il est frappant que, parmi ces 42 gènes se trouve *Nlrp3* (voir Figure 1D), dont la dérégulation semblerait justement expliquer le phénotype d'hypermaturation des neutrophiles observé chez les mutants miR-223^Δ (Haneklaus *et al.*, 2012 ; Bauernfeind *et al.*, 2012).

2.2 Les gènes dose-sensibles portent les sites de fixation aux miARN les mieux conservés

Afin de tester la généralité des résultats présentés ci-dessus, nous avons exploré à plus grande échelle la relation entre l'interaction ARNm/miARN et la sensibilité phénotypique vis-à-vis du niveau d'expression d'un gène d'intérêt. En utilisant les données disponibles concernant l'annotation (Dang *et al.*, 2008) ou la prédiction (Huang *et al.*, 2010) de l'haplo-suffisance chez l'Humain, nous nous sommes aperçus que les gènes connus ou prédits pour être haplo-insuffisants, tendent à porter les sites d'interaction aux miARN les mieux conservés (voir Figure 2).

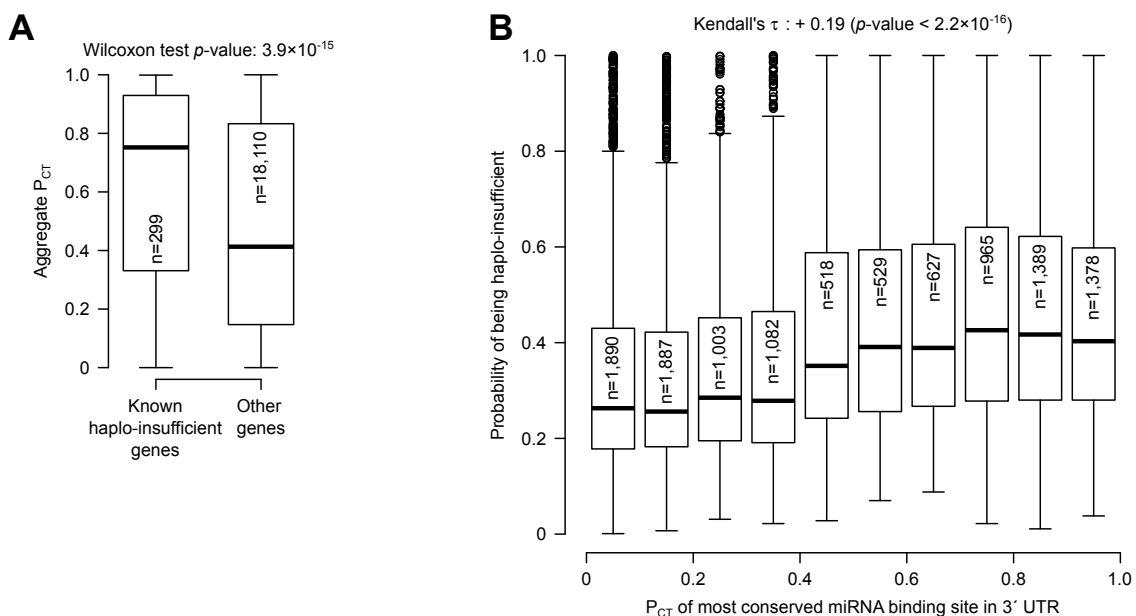


FIGURE 2 – Les gènes haplo-insuffisants humains tendent à porter les sites d'interaction aux miARN les plus conservés. **A.** Les gènes haplo-insuffisants humains connus (Dang *et al.*, 2008) portent des sites d'interaction aux miARN plus conservés que les autres gènes (la conservation des sites est mesurée par la P_{CT} , *probability of conserved targeting*, décrite par Friedman *et al.*, 2009). **B.** La conservation des sites de fixation de miARN corrèle avec la probabilité pour les gènes humains d'être haplo-insuffisants, telle que calculée par Huang *et al.*, 2010 (les gènes ont été groupés en *boxplots* en fonction de leur P_{CT} pour plus de clarté). Dans chaque *boxplot*, le nombre de gènes dans chaque catégorie est indiqué dans les boîtes.

Il semble donc bien que toutes les « cibles » de miARN identifiées ou prédites ne se valent pas : la plupart, peu sensibles au niveau d'expression, sont fonctionnellement insensibles à la petite régulation guidée par les miARN ; seuls les gènes exceptionnellement sensibles au dosage peuvent être efficacement contrôlés par les miARN.

Cette notion a été complètement ignorée jusqu'ici, et la plupart des études se contentent de mesurer l'effet des miARN sur l'accumulation d'ARNm ou de protéines des gènes-cibles : les « cibles » identifiées de cette manière sont probablement très largement contaminées par des faux positifs, dont la régulation par le miARN est sans conséquence biologique.

2.3 Mesure de la titration des miARN par les ARNm

J'avais déjà proposé, en 2009, que la plupart des « cibles » connues de miARN étaient en réalité insensibles à l'effet des miARN (Seitz, 2009). Il fallait cependant expliquer pourquoi ces nombreux gènes portent des sites de fixation aux miARN qui ont été conservés dans l'évolution : si effectivement ces sites sont inutiles, pourquoi ont-ils été conservés ? C'est en effet une caractéristique des cibles prédites informatiquement, puisque les programmes de prédiction de cibles recherchent généralement les sites phylogénétiquement conservés (voir note de bas de page, en p. 4). J'avais donc proposé que ces sites d'interaction sont conservés pour une autre raison : en titrant les miARN, ces ARNm se comporteraient comme des régulateurs de miARN, fonction probablement bénéfique (donc : phylogénétiquement conservée) dans les types cellulaires où l'expression de ce miARN est détrimmentale.

Depuis cette proposition (seulement théorique à l'époque), d'autres laboratoires avaient tenté d'identifier de tels titrateurs de miARN (Poliseno *et al.*, 2010 ; Salmena *et al.*, 2011 ; Cesana *et al.*, 2011). Mais la stratégie expérimentale posait problème, et les titrateurs proposés par ces articles n'en sont probablement pas (Ebert et Sharp, 2010). Le consensus actuel veut plutôt que les ARNm ont peu de chance d'être assez abondants pour jouer un réel rôle de modulateur de miARN (Garcia *et al.*, 2011 ; Wee *et al.*, 2012 ; Denzler *et al.*, 2014) ; il faudrait en tout cas que les concentrations intracellulaires à la fois de l'ARNm, et du miARN, tombent dans des domaines bien précis (Bosson *et al.*, 2014 ; Jens et Rajewsky, 2015).

Cependant pour l'heure, personne n'a encore mesuré les concentrations intracellulaires des miARN et des ARNm-cibles de manière précise et répliquée. Nous avons donc entrepris de le faire, dans un système cellulaire où les miARN sont connus pour jouer un rôle physiologique : la différenciation des cellules C2C12 (Chen *et al.*, 2006 ; Kim *et al.*, 2006 ; Goljanek-Whysall *et al.*, 2012). Cette lignée cellulaire murine a des caractéristiques de myoblaste, et il est possible d'induire, en culture, sa différenciation en myotubes.

En utilisant des *Northern blots* calibrés par des oligos synthétiques, et du RNA-Seq calibré par des transcrits *in vitro* introduits en quantités contrôlées, nous avons pu mesurer la concentration intracellulaire de miR-1a/miR-206 et miR-133 (des miARN spécifiques de la lignée myocytaire ; voir Figure 3A et B), ainsi que de l'ensemble des ARNm. Pour déterminer, parmi les ARNm, lesquels interagissent physiquement avec ces miARN, nous avons réalisé une immuno-précipitation des ARNm après inhibition spécifique d'un miARN ou de l'autre (voir Figure 3C). Nous avons donc identifié expérimentalement les ARNm qui fixent ces miARN dans les cellules C2C12 (voir Figure 3D), et, à partir de leurs concentrations intracellulaires, déterminé l'effet de chacun sur la titration des miARN (voir Figure 3E et F). Il apparaît que plusieurs ARNm sont tellement abondants qu'ils peuvent titrer, individuellement, plusieurs dizaines de pourcents de miARN. Notamment, l'ARNm du gène *Tmsb4x* semble titrer la moitié du miARN miR-1a/miR-206 (s'il perdait son site pour ce miARN, la concentration de miR-1a/miR-206 libre serait multipliée par deux).

Pour vérifier expérimentalement ce résultat, nous avons utilisé la technologie CRISPR/Cas9 pour muter le site de fixation de miR-1a/miR-206 dans le gène *Tmsb4x*. À l'aide d'une nucléase Cas9 haute-fidélité (Kleinstiver *et al.*, 2016) et en utilisant la recombinaison homologue, nous avons généré plusieurs lignées polyclonales de C2C12 mutées ou non sur ce site, et portant un rapporteur luciférase pour mesurer l'activité de miR-1a/miR-206 dans les cellules (voir Figure 4A).

Nous avons donc mesuré l'activité du rapporteur luciférase dans chacune des 5 lignées polyclonales où le site miR-1a/miR-206 est sauvage, et dans les 4 lignées où il est muté. Les résultats (voir Figure 4B) montrent que le rapporteur est davantage réprimé dans les lignées mutantes, confirmant que la mutation du site de miR-1a/miR-206 sur le gène *Tmsb4x* libère une quantité notable de miARN.

Il s'agit de la première mesure rigoureuse de la titration des miARN par un ARNm, et cette mesure a été réalisée dans un type cellulaire où les miARN étudiés sont connus pour jouer un rôle biologique. Ces résultats montrent que le gène *Tmsb4x* module effectivement la disponibilité de miR-1a/miR-206 dans les cellules C2C12 ; nous avons également identifié une demi-douzaine d'autres gènes dont les concentrations intracellulaires semble suffisantes pour qu'ils titrent efficacement les miARN myocytaires.

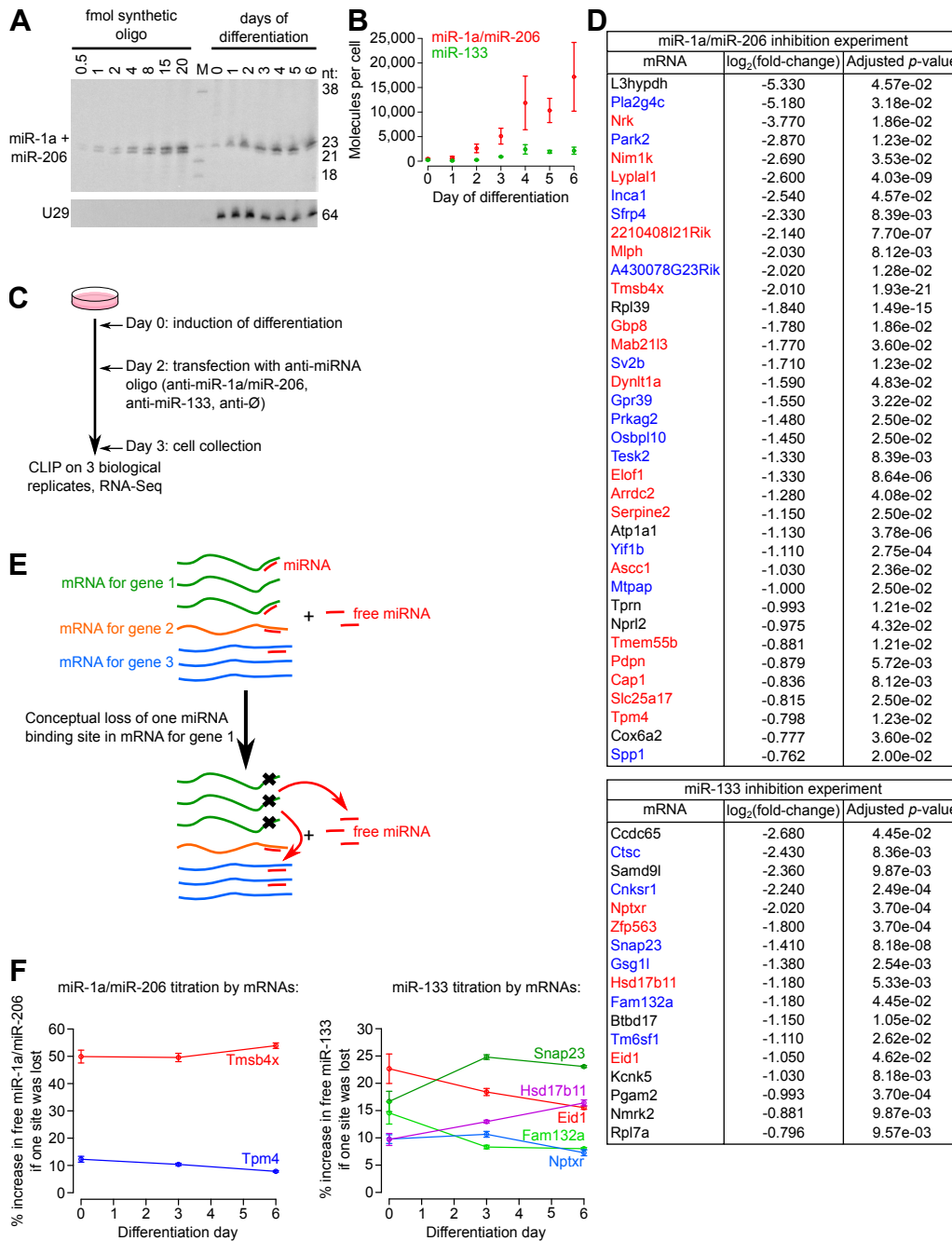


FIGURE 3 – Des ARNm individuels suffisamment abondants pour titrer les miARN spécifiques des myotubes dans les cellules C2C12 en différenciation. **A.** Pistes de gauche : miR-1a et miR-206 synthétiques (pour calibration). M : marqueur de taille. Pistes de droite : 20 μ g d'ARN total de C2C12 en différenciation. **B.** Quantification de trois réplicats biologiques de l'expérience montrée en panel A pour chaque famille de miARN (moyenne \pm erreur-type). **C.** Identification expérimentale des cibles de miR-1a/miR-206 et de miR-133 dans les cellules C2C12. Les cellules ont été transfectées avec des oligos 2'-O-Me dirigés contre miR-1a et miR-206, contre miR-133, ou contre aucun miARN murin ("anti- \emptyset "). Les ARNm immunoprécipités avec le complexe effecteur des miARN ont été quantifiés par RNA-Seq indépendant de la queue poly(A). **D.** Les cibles identifiées pour miR-1a/miR-206 (haut) et miR-133 (bas). Rouge : ARNm avec une complémentarité parfaite à la graine du miARN dans leur 3' UTR. Bleu : ARNm dont la meilleure complémentarité à la graine dans la 3' UTR est l'une des trois complémentarités imparfaites les plus fréquentes (CNATTCC, CATNCC ou CNTTCC pour miR-1a/miR-206 ; GACCANA, GNACCAA ou GACNCAA pour miR-133). **E.** Les concentrations de miARN libre et fixé ont été calculées d'après nos mesures de concentrations intracellulaires, et d'après la valeur de la constante de dissociation miARN/ARNm mesurée par Wee *et al.*, 2012. Le même calcul a été refait après perte conceptuelle du site de fixation de miARN d'intérêt. **F.** Les sites de fixation de miARN qui exercent la plus haute activité titratrice (augmentation de $> 10\%$ de la concentration de miARN libre si le site était perdu).

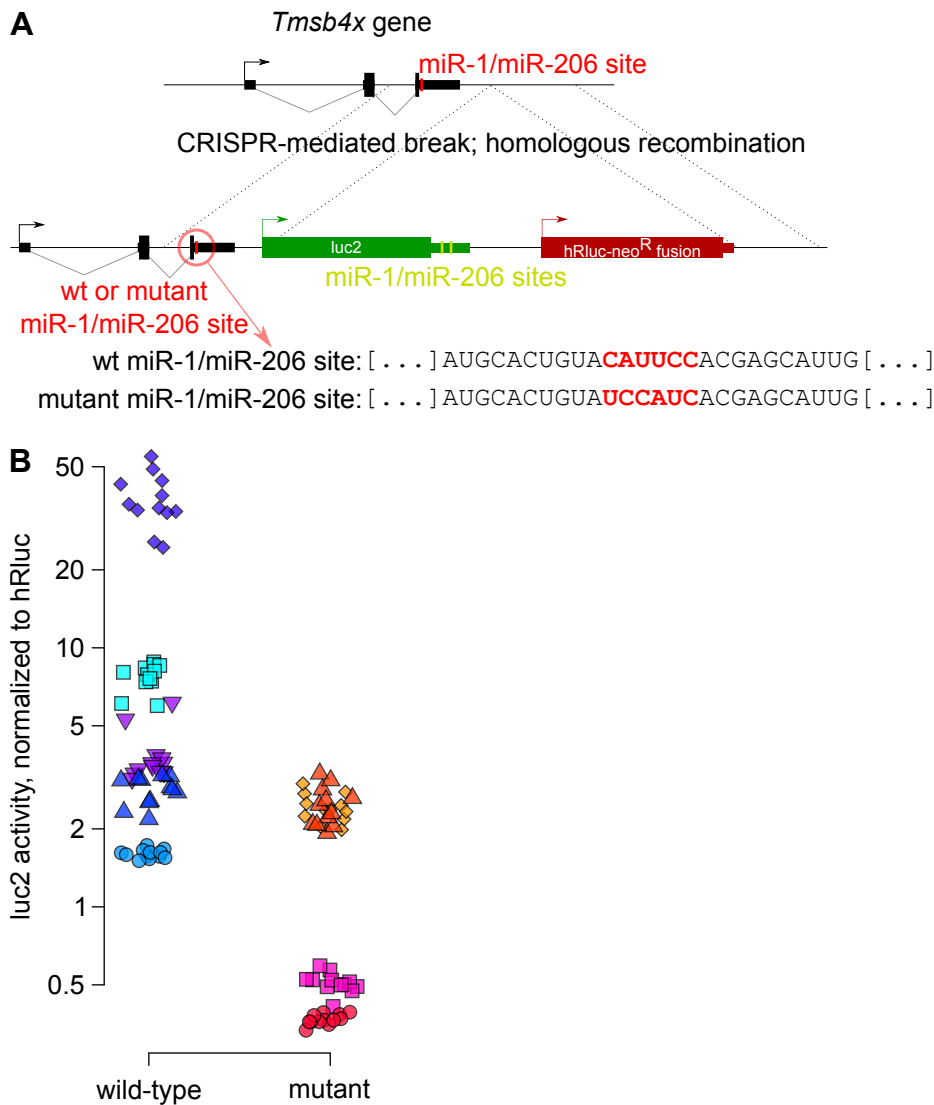


FIGURE 4 – *Tmsb4x* titre efficacement miR-1a/miR-206 dans les cellules C2C12 en différenciation. **A.** Stratégie de mutagenèse. Un rapporteur luciférase et une cassette de résistance au G418 ont été introduits 286 bp en aval du signal de polyadénylation de *Tmsb4x* et la 3' UTR de *Tmsb4x* a été remplacée par une copie où le site de fixation de miR-1a/miR-206 est soit remplacé par lui-même (« wt »), soit par un hexamère qui n'est complémentaire à aucune graine de miARN murin (« mutant »). **B.** L'activité luciférase a été mesurée dans chacune des 5 lignées polyclonales wt et des 4 lignées polyclonales mutantes, après différenciation. Chaque lignée a été analysée en 12 répliquats techniques (les répliquats d'une même lignée sont représentés par le même symbole et la même couleur). Un test *two-way ANOVA* sur le logarithme des données montre qu'à la fois le génotype et l'identité de la lignée cellulaire ont un effet significatif sur l'activité du rapporteur ($p < 2.2 \times 10^{-16}$ pour chaque facteur).

2.4 La titration par les ARNm affine le patron d'expression des miARN

La section précédente montre qu'une demi-douzaine d'ARNm semblent assez abondants pour réguler miR-1a/miR-206 ou miR-133 dans les cellules C2C12. Il est possible que que les autres cibles prédites pour ces miARN ne soient pas capables de les titrer. Alternativement, il est possible que certains de ces autres ARNm soient des titrateurs efficaces, mais dans d'autres types cellulaires, où leur expression serait suffisante.

Pour tester la généralité de ce résultat, nous avons donc décidé d'explorer la relation entre l'abondance des ARNm et la conservation de leurs sites d'interaction aux miARN. Partant du principe que la titration des miARN par les ARNm, quand elle a lieu, joue un rôle de régulateur de miARN, il est probable que cette titration soit phylogénétiquement conservée dans les types cellulaires où elle est bénéfique. Les miARN exprimés fortement dans un type cellulaire (par exemple, miR-122 dans les hépatocytes) jouent probablement un rôle biologique dans ce type cellulaire, et la titration de ces miARN n'aurait un effet bénéfique que dans les autres types cellulaires. Ainsi, outre les ARNm montrés en Figure 3F, il est possible que des titrateurs supplémentaires existent pour miR-1a/miR-206 ou miR-133, qui réguleraient ces miARN seulement dans des types cellulaires bien précis.

Une prédiction testable peut être déduite de ces considérations : si effectivement des ARNm jouent un rôle titrateur bénéfique dans divers types cellulaires, alors dans chacun de ces types cellulaires les titrateurs seraient fortement exprimés. Plus la titration est bénéfique, plus l'interaction entre ARNm et miARN serait conservée : il serait donc prévisible que les ARNm les plus fortement exprimés soient sous la pression de sélection la plus forte pour conserver leurs sites d'interaction aux miARN.

Nous avons donc comparé la conservation des sites de fixation de miARN (mesurée par leur P_{CT} ; Friedman *et al.*, 2009) et l'abondance des ARNm dans des données de transcriptomique publiquement accessibles, pour une variété de tissus de Souris. Nous avons observé qu'effectivement, pour la plupart des familles de miARN, l'abondance des ARNm corrèle positivement avec la conservation de leurs sites d'interaction aux miARN (voir Figure 5A). Notez que le P_{CT} compense pour la conservation globale des 3' UTR (les sites portés par des UTR très conservées doivent être eux-mêmes plus conservés pour atteindre le même P_{CT} que des sites portés par des UTR peu conservées). Cette propriété corrige pour d'éventuels biais de conservation de séquence entre gènes fortement et faiblement exprimés.

Il est intéressant de constater que la corrélation n'est pas seulement forte pour les miARN fortement exprimés dans le tissu analysé (tels que miR-1 et miR-133 dans le muscle), mais aussi pour les miARN spécifiques d'un autre tissu que le tissu analysé (voir par exemple l'analyse du transcriptome du muscle pour miR-122, qui est lui-même spécifique du foie ; Figure 5B). Cette observation suggère que le transcriptome de chaque type cellulaire n'est pas sous une pression de sélection forte pour moduler ses propres miARN. Il titre principalement les miARN spécifiques des autres tissus (dont, potentiellement, l'expression est détrimentale dans ce type cellulaire).

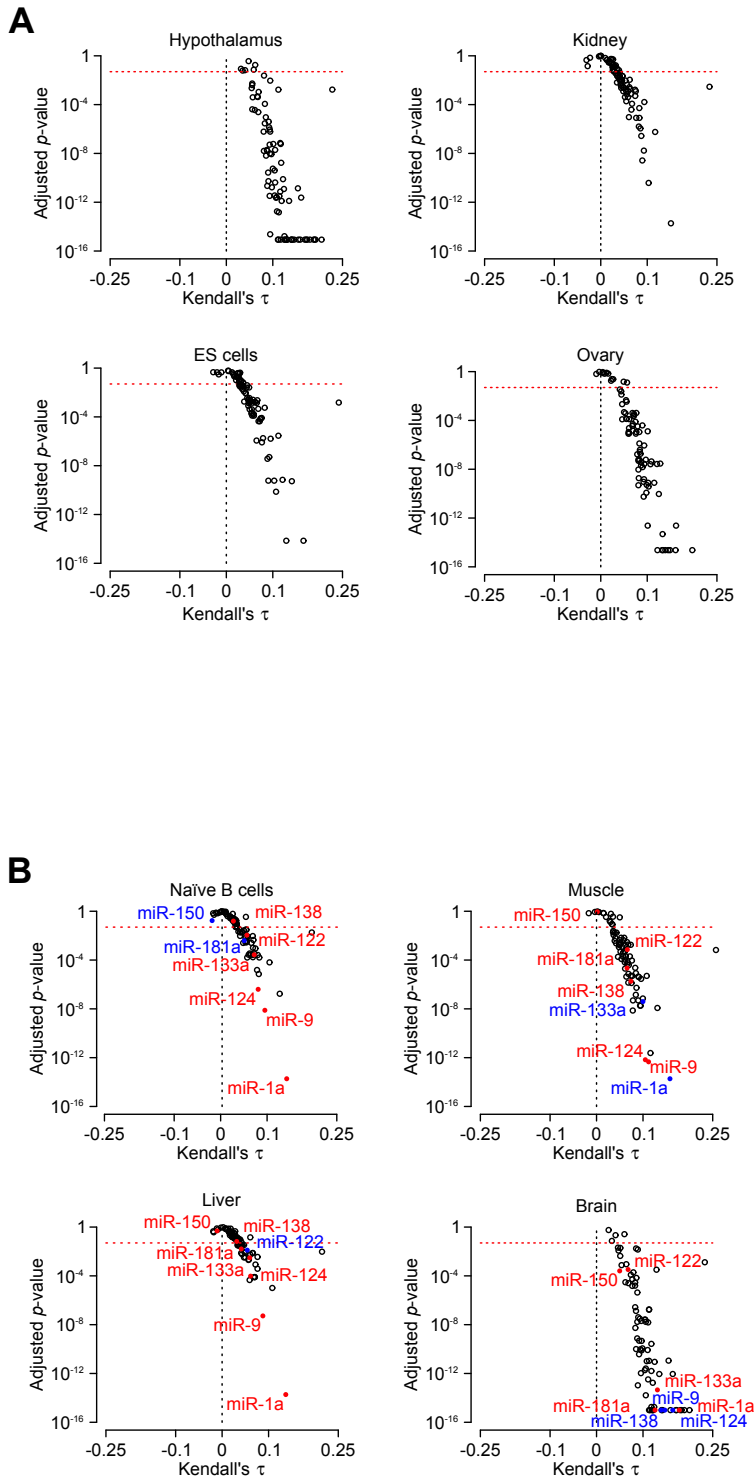


FIGURE 5 – Les gènes les plus exprimés tendent à porter les sites d'interaction aux miARN les plus conservés. **A.** Ces volcano plots montrent les coefficients de corrélation entre les signaux de microarray et le P_{CT} agrégé (Friedman *et al.*, 2009) pour chaque ARNm (axe des abscisses), et leurs p -values (axe des ordonnées). Chaque famille de miARN est représentée par un cercle. Les p -values ont été ajustées avec la correction de Benjamini-Hochberg. La ligne pointillée rouge indique une p -value ajustée de 0,05 et la ligne pointillée noire indique un coefficient de corrélation de 0. Les p -values ajustées inférieures à 2.2×10^{-16} ont été ramenées à 2.2×10^{-16} pour plus de clarté graphique. **B.** Mêmes conventions que dans le panel **A**, mais les familles de miARN à patron d'expression très spécifique sont colorées (en rouge : miARN spécifique d'un autre tissu que le tissu analysé ; en bleu : miARN spécifique du tissu analysé).

2.5 Des faux positifs dans les analyses de génomique comparative

Nos résultats précédents suggèrent que seule une petite fraction des gènes avec des sites de fixation de miARN sont suffisamment sensibles pour être fonctionnellement régulés par les miARN (Figures 1 et 2). Parmi les autres « cibles » connues, certaines pourraient porter un site de fixation conservé en raison de leur activité titratrice de miARN (Figures 3, 4 et 5). Il est difficile de savoir combien de titrateurs existent en tout : pour chaque gène (même les moins exprimés), il est formellement possible qu'il existe un type cellulaire ou un stade développemental où son expression est forte, et où il exercerait une titration bénéfique.

Alternativement, il reste possible que de nombreuses « cibles » connues ne soient ni sensibles aux miARN, ni capables de titrer efficacement des miARN. Dans cette « zone grise », les gènes portent des sites de fixation aux miARN dont la conservation phylogénétique resterait donc à expliquer. Nous avons donc décidé d'explorer la possibilité que les résultats de génomique comparative soient contaminés par des faux positifs.

Il est en effet possible qu'une séquence soit conservée à cause d'une autre raison, indépendante des miARN, tout en étant fortuitement complémentaire à un miARN. La conservation de cette séquence serait attribuée par erreur à la fonctionnalité de l'interaction entre le miARN et l'ARNm. Il est possible d'estimer la fréquence de tels faux positifs en comptant les sites de fixation de miARN dans des espèces qui ne possèdent pas le miARN d'intérêt (les miARN partageant la même séquence-graine appartiennent à la même « famille de miARN »).

En utilisant l'alignement multiple de 100 génomes, centré sur le génome humain, de l'UCSC (Rosenbloom *et al.*, 2015) nous avons ainsi identifié les complémentarités aux graines de miARN, dans les 3' UTR, qui apparaissent plus conservées que la graine de miARN elle-même (voir Figure 6A pour un exemple). De tels sites « sur-conservés » peuvent produire des faux positifs dans les prédictions de cibles de miARN : leur conservation pourrait être due à un autre facteur *trans* (une protéine d'interaction à l'ADN ou à l'ARN par exemple).

Les programmes de prédiction de cibles de miARN utilisent d'autres critères (outre la conservation de la complémentarité à la graine) : richesse locale en A/U, complémentarité au 3' du miARN, *etc.*, et ces critères supplémentaires pourraient aider à diminuer le taux de faux positifs. Effectivement, les sites prédits par les programmes les plus populaires tendent à être moins souvent sur-conservés que les simples sites de complémentarité dans les 3' UTR (comparer la Figure 6B à la Figure 6C). Pourtant, même ces programmes sont lourdement contaminés par des candidats dont les sites de complémentarité sont sur-conservés (de $\approx 30\%$ à $\approx 70\%$ pour les familles de miARN spécifiques des Hominidés).

La fréquence des sites sur-conservés montre une dépendance frappante à la profondeur phylogénétique du clade d'intérêt : les clades peu profonds (comme les Hominidés) ont des taux de sur-conservation supérieurs à ceux des clades profonds (comme les Eutélostomiens). Nous avons pu imaginer plusieurs interprétations à ce phénomène : **(i)** l'annotation des miARN serait moins fiable pour les miARN peu conservés, résultant en une contamination des miARN spécifiques des Hominidés par des petits ARN qui ne sont pas d'authentiques miARN ; **(ii)** un site de complémentarité à une graine de miARN est moins susceptible d'être conservé à l'extérieur d'un clade profond que d'un clade peu profond, parce qu'il a plus de chances d'être perdu lors d'une longue divergence évolutive ; **(iii)** il y a plus d'espèces externes à un clade peu profond qu'à un clade profond, ce qui augmente les chances qu'une espèce externe possède la complémentarité au miARN.

Pour tester ces possibilités, nous avons mesuré la proportion de complémentarités aux graines de miARN qui sont conservées au-delà de quatre clades de profondeur variable, et pour des graines de miARN de conservation variable (voir Figure 6D). Le résultat montre que le nombre de sites sur-conservés diminue quand on recherche la conservation au-delà des clades les plus profonds, et cette diminution est similaire pour tous les quatre types de graines (spécifiques des Hominidés, des Catarrhini, des Boreo euthériens, ou des Eutélostomiens). Cette observation indique que la diminution n'est pas due à une différence de qualité d'annotation des miARN entre miARN faiblement et fortement conservés. C'est simplement une conséquence de la structure arborescente de l'évolution (interprétations **(ii)** et **(iii)** ci-dessus). D'ailleurs, les hexamères qui ne sont complémentaires d'aucune graine de miARN de Vertébré montrent la même décroissance (en noir dans la Figure 6D), confirmant qu'elle est due à la tendance des éléments de séquence à être plus facilement conservés au-delà des clades peu profonds que des clades profonds.

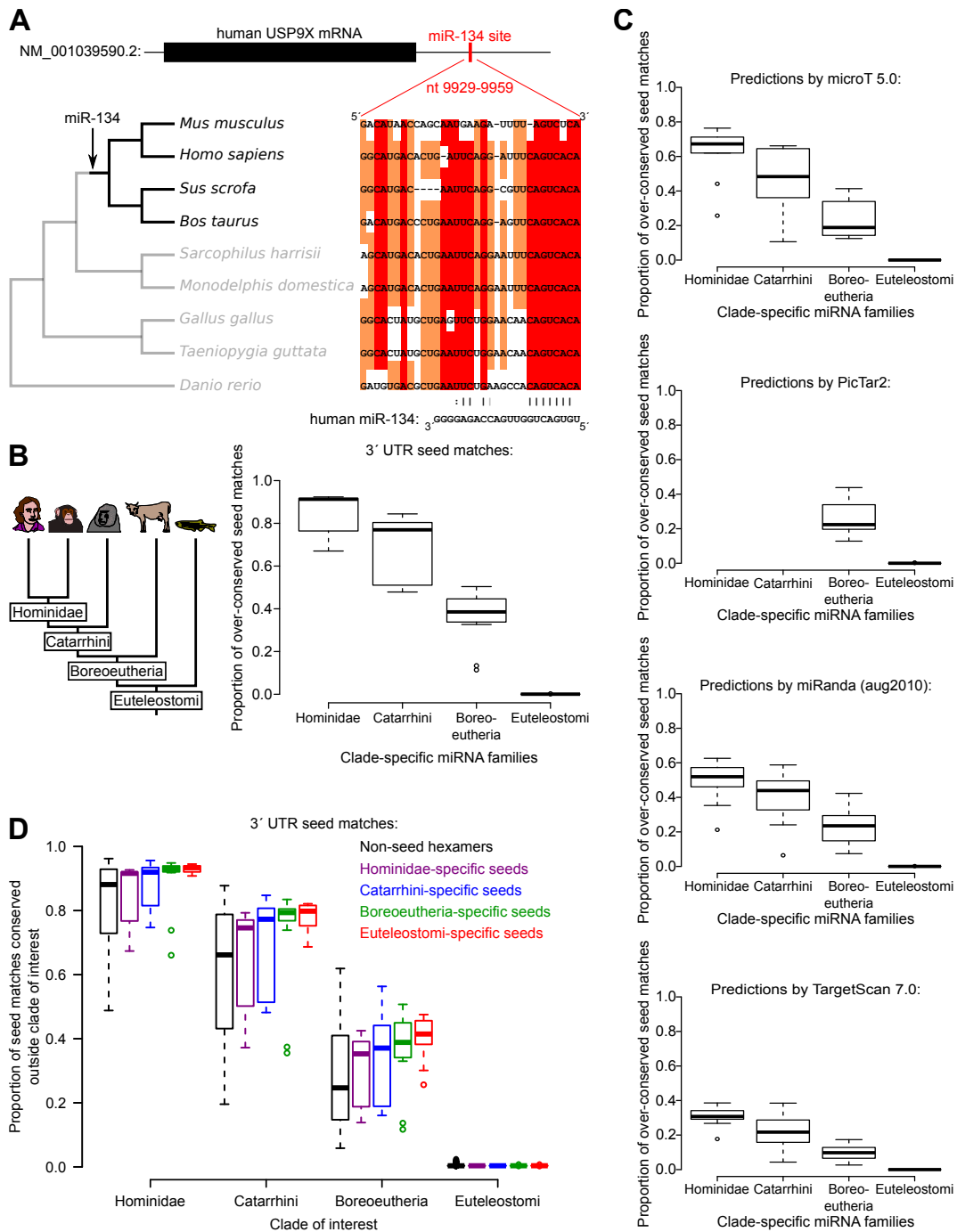


FIGURE 6 – Les complémentarités aux graines de miARN identifiées informatiquement sont fréquemment plus conservées que les graines elles-mêmes. **A.** La famille miR-134 est spécifique des Mammifères placentaires, mais son site de fixation dans la 3' UTR de *USP9X* est plus profondément conservée. **B.** Quatre clades de Vertébrés ont suffisamment de familles de miARN qui leur sont spécifiques pour pouvoir mener une analyse statistique détaillée (10 familles spécifiques des Hominidés, 14 familles spécifiques des Catarrhini, 14 familles spécifiques des Boreoeuthériens, 10 familles spécifiques des Eutélostomiens). Chaque point dans les *boxplots* représente une famille de miARN. La proportion de sites sur-conservés dans les 3' UTR est définie comme la fraction de sites qui sont conservés dans au moins une espèce extérieure au clade d'intérêt. **C.** La proportion de sites sur-conservés parmi les prédictions de plusieurs programmes de prédiction de cibles de miARN. Notez que PicTar2 ignore les miARN spécifiques des Hominidés ou des Catarrhini. Pour comparer de manière juste tous les programmes, les analyses ont été restreintes aux complémentarités parfaites aux graines de miARN, situées dans les 3' UTR, et en excluant les sites qui chevauchent des jonctions exon-exon. **D.** Les complémentarités situées dans les 3' UTR ont été analysées comme dans le panel **B**, mais la sur-conservation a été mesurée pour chaque groupe de graines spécifiques d'un clade, au-delà de chacun des quatre clades. Les hexamères qui ne constituent la graine d'aucun miARN de Vertébré (dans miRBase 21) ont été analysés de manière identique (*boxplots* noires).

Ainsi le taux de faux positifs pour les graines de miARN en général est probablement plus proche de celui mesuré pour les graines spécifiques des Hominidés, que pour les graines spécifiques des Eutélostomiens : les graines spécifiques des Eutélostomiens ont des taux de sur-conservation plus bas simplement à cause de la propriété générale des sites sur-conservés à être plus rares dans les clades profonds. Mais la propension des programmes de prédiction de cibles à capturer des faux positifs est probablement similaire à celle des graines spécifiques des Hominidés, parce qu'ils utilisent les mêmes critères de prédiction pour les graines fortement et faiblement conservées, et qu'ils sont probablement soumis aux mêmes contaminations par des sites conservés pour d'autres raisons (protéines d'interaction à l'ADN ou à l'ARN, *etc.*). Le taux de faux positifs dans les programmes de prédiction les plus populaires semble donc être de l'ordre de plusieurs dizaines de pourcents, de $\approx 30\%$ (avec TargetScan 7.0) à $\approx 70\%$ (avec microT 5.0).

L'ensemble des travaux présentés ci-dessus (Figures 1 à 6) est en considération pour publication par *Genome Research* (soumission initiale le 4 février 2016, révision soumise le 22 août 2016).

2.6 Projet terminé : évolution du mode de répression par les miARN chez les Animaux

Dans le cadre d'une collaboration avec le laboratoire d'U. Technau (université de Vienne, en Autriche), celui de F. Rentzsch (SARS Center, Norvège) et mon laboratoire post-doctoral (celui de P. D. Zamore, University of Massachusetts Medical School, États-Unis), nous avons mené une analyse à haut débit de la population de petits ARN au cours du développement de l'Anémone de mer *Nematostella vectensis*. Cet animal est un Cnidaire, dont le dernier ancêtre commun avec les animaux à symétrie bilatérale est vieux d'au moins 700 millions d'années. Nos analyses ont montré que les miARN de *Nematostella* présentaient fréquemment de longues complémentarités avec les ARNm, ce qui suggérait que ces cibles pourraient être dégradées par clivage endonucléolytique (à la manière du RNAi). Notre analyse des produits de clivage du complexe effecteur des miARN a confirmé que les cibles de miARN, chez cet animal, sont fréquemment dégradées par clivage endonucléolytique, de manière similaire au mode de répression par les miARN les mieux caractérisés chez les Plantes, et à la différence du mode de répression principal chez les autres Animaux étudiés.

Nos travaux suggèrent donc que le mode de répression ancestral par les miARN était le clivage endonucléolytique (à la manière du RNAi), et que cette caractéristique a été surtout préservée chez les Plantes et chez ce Cnidaire. La répression principalement par dégradation exonucléolytique et répression traductionnelle serait donc une innovation des animaux à symétrie bilatérale.

Depuis la publication de ces travaux (Moran *et al.*, 2014), ce projet est terminé et n'est plus traité au laboratoire.

2.7 Collaborations diverses

miARN et répression méiotique des chromosomes sexuels

Nos collaborateurs dans le laboratoire de J. Turner (Crick Institute, Royaume-Uni) étudient la *meiotic sex chromosome inactivation*. Il s'agit d'une répression des hétérosomes chez les Mammifères mâles au cours de la méiose : alors que les autosomes s'apparient (par paires de chromosomes homologues) en début de méiose, les chromosomes X et Y ne peuvent s'apparier que sur la région pseudo-autosomale. Le reste de la chromatine des chromosomes sexuels, non-appariée, est alors réprimée par une le dépôt de marques épigénétiques répressives.

Un autre laboratoire avait proposé que des gènes de miARN situés sur le chromosome X échappaient à ce phénomène (Song *et al.*, 2009). Les travaux de nos collaborateurs suggèrent qu'il s'agit essentiellement d'un artefact : les expériences de Song *et al.*, 2009 détectaient les miARN matures, qui pouvaient avoir été hérités des stades développementaux précédents. Nos collaborateurs ont pris soin de détecter spécifiquement les transcrits primaires des gènes de miARN (les « pri-miARN »), et ont montré que ces transcrits, aussi bien que les autres gènes hétérosomaux, sont bien soumis à la *meiotic sex chromosome inactivation*. Nous les avons aidés à analyser l'expression des miARN par microarray, et à conduire les tests statistiques sur les phénotypes des souris transgéniques qui miment la perte de la répression de ces miARN.

Ces travaux ont été publiés l'an dernier (Royo *et al.*, 2015).

Réorganisation des origines de réplication au cours du développement du Nématode

Nos collaborateurs dans le laboratoire de M. Méchali (dans notre institut, l'IGH, à Montpellier) ont détecté expérimentalement, par pavage par microarray pan-génomique, les origines de réplication chez des embryons précoces (2 à 40 cellules) de Nématode, et dans une population contenant à la fois embryons précoces et tardifs. Nous les avons aidés à analyser ces données : à partir des pics de signal le long du génome, nous avons extrait les positions des origines de réplication chez les embryons précoces, puis (par soustraction de l'échantillon contenant embryons précoces et tardifs) celles des origines de réplication chez les embryons tardifs.

Nous avons ensuite croisé ces données avec de nombreux jeux de données génomiques publiquement disponibles (marques de chromatine, loci de petits ARN, occupation par l'ARN polymérase II, annotation d'éléments régulateurs, ...). Ces travaux nous ont permis de caractériser les déterminants spécifiques des origines de réplication et leur dynamique au cours du développement précoce du Nématode.

Le manuscrit résumant ces travaux est en cours d'évaluation par *Nature Structural and Molecular Biology* (soumis le 26 août 2016).

Histoire évolutive des miARN

Nos collaborateurs dans le laboratoire de M. Cock (station biologique de Roscoff) avaient identifié, par small RNA-Seq, des candidats miARN chez l'Algue brune *Ectocarpus silicosus*. Cet Eucaryote basal (aussi éloigné des Plantes que des Animaux ou des Champignons) pourrait donc exprimer des miARN, ce qui préciserait davantage la date de leur apparition dans le règne vivant.

Nous les avons aidés à confirmer l'existence de ces miARN en les détectant par une autre méthode (le *Northern blot*), validant ainsi leurs conclusions sur l'abondance et la diversité des miARN chez cette Algue brune (Tarver *et al.*, 2015). Nos travaux ont également permis de revisiter des candidats miARN, précédemment publiés par un autre laboratoire (Billoud *et al.*, 2014), et qui se trouvent être principalement des artefacts de qPCR (une méthode moins spécifique que le *Northern blot*, et incapable de distinguer les miARN matures, de taille stéréotypée, des produits de dégradation du transcriptome).

Un variant génomique qui fait apparaître un site de fixation de miARN chez l'Homme

Nos collaborateurs dans le laboratoire de G. Rappold (université d'Heidelberg, Allemagne) nous ont contactés lorsqu'ils se sont aperçus qu'une mutation humaine, fréquemment associée à une pathologie cardiaque (la fibrillation atriale), se trouvait dans la 3' UTR du gène *SHOX2*. Nous nous sommes aperçus que cette mutation faisait apparaître un site de fixation pour le miARN miR-92b-5p, susceptible de provoquer une répression de *SHOX2* qui pourrait participer au phénotype pathologique.

Nos collaborateurs ont ensuite mesuré l'expression de *SHOX2* *in vivo* chez des sujets portant ou non la mutation, et ont observé que ce gène est effectivement réprimé chez les individus mutés. Il est donc possible que l'acquisition de ce site pour miR-92b-5p contribue aux symptômes de la maladie (Hoffmann *et al.*, 2016).

Prédiction des cibles de miARN chez le Puceron

Nos collaborateurs dans le laboratoire de D. Tagu (INRA de Rennes) s'intéressent aux mécanismes épigénétiques qui contrôlent la plasticité phénotypique chez le Puceron. En fonction des stimuli extérieurs, la morphologie et le mode de reproduction du Puceron peuvent changer profondément, alors que leur information génétique reste inchangée (tous les pucerons d'une colonie sont clones, issus de la reproduction par parthénogenèse d'une femelle fondatrice). Les analyses génomiques et transcriptomiques conduites dans le laboratoire de D. Tagu avaient suggéré que la machinerie des miARN pourrait être impliquée dans ces changements phénotypiques, parce que les niveaux d'expression de plusieurs acteurs de la voie des miARN sont altérés au cours des transitions phénotypiques.

J'étais partenaire sur le financement ANR obtenu par D. Tagu (projet « MiRNAdapt », ANR-11-BSV6-01701), et qui couvrait la période de 2012 à 2015. J'ai aidé nos collaborateurs à annoter les candidats miARN qu'ils ont détectés, puis j'ai participé au choix des critères de prédiction bio-informatiques de cibles pour ces candidats miARN. J'ai aidé à choisir les méthodes de validation expérimentale. Le manuscrit présentant ces travaux est en cours de préparation.

Les participants à ce projet étaient :

- Coordinateur : Denis Tagu (INRA Rennes) ;
- Partenaire : Anne Siegel (IRISA Rennes) ;
- Partenaire : moi-même (CNRS Montpellier).

2.8 Références

- Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., et Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, 455(7209) : 64–71.
- Bauernfeind, F., Rieger, A., Schildberg, F. A., Knolle, P. A., Schmid-Burgk, J. L., et Hornung, V. (2012). NLRP3 inflammasome activity is negatively controlled by miR-223. *J Immunol*, 189(8) : 4175–4181.
- Billoud, B., Nehr, Z., Le Bail, A., et Charrier, B. (2014). Computational prediction and experimental validation of microRNAs in the brown alga *Ectocarpus siliculosus*. *Nucleic Acids Res*, 42(1) : 417–429.
- Bosson, A. D., Zamudio, J. R., et Sharp, P. A. (2014). Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol Cell*, 56(3) : 347–359.
- Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., et Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*, 147(2) : 358–369.
- Chen, C. Z., Li, L., Lodish, H. F., et Bartel, D. P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. *Science*, 303(5654) : 83–86.
- Chen, J. F., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T. E., Hammond, S. M., Conlon, F. L., et Wang, D. Z. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet*, 38(2) : 228–233.
- Dang, V. T., Kassahn, K. S., Marcos, A. E., et Ragan, M. A. (2008). Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *Eur J Hum Genet*, 16(11) : 1350–1357.
- Denzler, R., Agarwal, V., Stefano, J., Bartel, D. P., et Stoffel, M. (2014). Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol Cell*, 54(5) : 766–776.
- Ebert, M. S. et Sharp, P. A. (2010). Emerging roles for natural microRNA sponges. *Curr Biol*, 20(19) : R858–R861.
- Friedman, R. C., Farh, K. K., Burge, C. B., et Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*, 19(1) : 92–105.
- Garcia, D. M., Baek, D., Shin, C., Bell, G. W., Grimson, A., et Bartel, D. P. (2011). Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lscy-6* and other microRNAs. *Nat Struct Mol Biol*, 18(10) : 1139–1146.
- Goljanek-Whysall, K., Pais, H., Rathjen, T., Sweetman, D., Dalmay, T., et Münsterberg, A. (2012). Regulation of multiple target genes by miR-1 and miR-206 is pivotal for C2C12 myoblast differentiation. *J Cell Sci*, 125(Pt 15) : 3590–3600.
- Haneklaus, M., Gerlic, M., Kurowska-Stolarska, M., Rainey, A. A., Pich, D., McInnes, I. B., Hammerschmidt, W., O'Neill, L. A., et Masters, S. L. (2012). Cutting edge : miR-223 and EBV miR-BART15 regulate the NLRP3 inflammasome and IL-1 β production. *J Immunol*, 189(8) : 3795–3799.
- Hoffmann, S., Clauss, S., Berger, I. M., Weiß, B., Montalbano, A., Röth, R., Bucher, M., Klier, I., Wakili, R., Seitz, H., Schulze-Bahr, E., Katus, H. A., , Flachsbarth, F., Nebel, A., Guenther, S. P., Bagaev, E., Rottbauer, W., Kääb, S., Just, S., et Rappold, G. (2016). Coding and non-coding variants in the *SHOX2* gene in patients with early-onset atrial fibrillation. *Basic Res Cardiol*, 111(3) : 36.
- Huang, N., Lee, I., Marcotte, E. M., et Hurles, M. E. (2010). Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet*, 6(10) : e1001154.
- Jens, M. et Rajewsky, N. (2015). Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat Rev Genet*, 16(2) : 113–126.
- Johnnidis, J. B., Harris, M. H., Wheeler, R. T., Stehling-Sun, S., Lam, M. H., Kirak, O., Brummelkamp, T. R., Fleming, M. D., et Camargo, F. D. (2008). Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature*, 451(7182) : 1125–1129.

- Kim, D. H., Villeneuve, L. M., Morris, K. V., et Rossi, J. J. (2006). Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol*, 13(9) : 793–797.
- Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., et Joung, J. K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529(7587) : 490–495.
- Moran, Y., Fredman, D., Praher, D., Li, X. Z., Wee, L. M., Rentzsch, F., Zamore, P. D., Technau, U., et Seitz, H. (2014). Cnidarian microRNAs frequently regulate targets by cleavage. *Genome Res*, 24(4) : 651–663.
- Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., et Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, 465(7301) : 1033–1038.
- Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T. R., Fujita, P. A., Guruvadoo, L., Haeussler, M., Harte, R. A., Heitner, S., Hickey, G., Hinrichs, A. S., Hubley, R., Karolchik, D., Learned, K., Lee, B. T., Li, C. H., Miga, K. H., Nguyen, N., Paten, B., Raney, B. J., Smit, A. F., Speir, M. L., Zweig, A. S., Haussler, D., Kuhn, R. M., et Kent, W. J. (2015). The UCSC Genome Browser database : 2015 update. *Nucleic Acids Res*, 43(Database issue) : D670–681.
- Royo, H., Seitz, H., ElInati, E., Peters, A. H., Stadler, M. B., et Turner, J. M. (2015). Silencing of X-Linked MicroRNAs by Meiotic Sex Chromosome Inactivation. *PLoS Genet*, 11(10) : e1005461.
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., et Pandolfi, P. P. (2011). A ceRNA hypothesis : the Rosetta Stone of a hidden RNA language? *Cell*, 146(3) : 353–358.
- Seitz, H. (2009). Redefining microRNA targets. *Curr Biol*, 19(10) : 870–873.
- Song, R., Ro, S., Michaels, J. D., Park, C., McCarrey, J. R., et Yan, W. (2009). Many X-linked microRNAs escape meiotic sex chromosome inactivation. *Nat Genet*, 41(4) : 488–493.
- Tarver, J. E., Cormier, A., Pinzón, N., Taylor, R. S., Carré, W., Strittmatter, M., Seitz, H., Coelho, S. M., et Cock, J. M. (2015). microRNAs and the evolution of complex multicellularity : identification of a large, diverse complement of microRNAs in the brown alga *Ectocarpus*. *Nucleic Acids Res*, 43(13) : 6384–6398.
- Wee, L. M., Flores-Jasso, C. F., Salomon, W. E., et Zamore, P. D. (2012). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell*, 151(5) : 1055–1067.

2.9 Place de ma recherche dans l'unité

Mon équipe (« Impact systémique des petits ARN régulateurs ») est une « équipe junior » à l'IGH (UPR 1142 du CNRS ; composé de 20 équipes, dont actuellement 4 équipes junior). Les équipes junior sont recrutées pour 5 ans, puis évaluées à l'issue des cinq ans : l'équipe est alors entérinée, ou dissoute, ou éventuellement la période « junior » peut être prolongée pour permettre à l'équipe de mener à bien les projets qui ont pris du retard. On attend des équipes junior d'avoir publié des articles importants, et d'assurer la viabilité de l'équipe à long terme (en terme de personnel permanent et de financement).

Mon équipe sera évaluée au printemps 2017. Je dois remettre, à l'automne 2016, à la direction de l'institut (qui le fera suivre à des évaluateurs externes et aux membres du *scientific advisory board*) le rapport d'activité de l'équipe. Je bénéficie, depuis l'installation de l'équipe à l'IGH en octobre 2011, des conseils d'un mentor externe, Javier Martinez (IMBA, Vienne, Autriche) : nous nous rencontrons au moins une fois par an pour discuter des aspects scientifiques, administratifs, financiers et humains de la gestion de l'équipe.

Au sein de l'IGH, mon équipe appartient au département de « Génétique et développement » (un département qui compte 6 équipes, dont 2 équipes junior). Nous participons activement à la vie du département, comme à celle de l'institut de manière générale (contribution régulière aux séminaires internes, invitations de conférenciers externes, ...). L'organigramme de l'IGH compte trois départements scientifiques, de tailles similaires (« Bases moléculaires de pathologies humaines », « Dynamique du génome » et « Génétique et développement »).

À titre personnel, j'ai donné à mes collègues de l'IGH plusieurs formations sur l'utilisation d'Unix et des outils simples de bio-informatique, sur les statistiques en biologie et sur l'usage du programme R (voir rubrique 1.5 « Autres activités professionnelles, activités d'expertise », en page 3), qui ont été très appréciées.

2.10 Distinctions scientifiques

Promu CR1 par le CNRS en 2013.

Prime d'excellence scientifique attribuée en novembre 2014.

2.11 Publications scientifiques sur les 5 derniers semestres

(numérotées selon l'ordre d'importance que je leur accorde ; par importance décroissante)

Revue à comité de lecture

1. Moran, Y., Fredman, D., Praher, D., Li, X. Z., Wee, L. M., Rentzsch, F., Zamore, P. D., Technau, U., and Seitz, H. (*) (2014) Cnidarian microRNAs frequently regulate targets by cleavage. *Genome Res*, 24(4) : 651–663. PDF de l'article (*) : **auteur correspondant**
2. Royo, H., Seitz, H., Ellnati, E., Peters, A., Stadler, M. and Turner, J. (2015) Silencing of X-linked microRNAs by meiotic sex chromosome inactivation. *PLoS Genetics*, 11(10) : e1005461. PDF de l'article
3. Chambeyron, S. and Seitz, H. (*) (2014) Insect small non-coding RNA involved in epigenetic regulations. *Curr Opin Insect Sci*, 1 : 1–9. PDF de l'article (*) : **auteur correspondant**
4. Hoffmann, S., Clauss, S., Berger, I.M., Weiß, B., Montalbano, A., Röth, R., Bucher, M., Klier, I., Wakili, R., Seitz, H., Schulze-Bahr, E., Katus, H.A., Flachsbarth, F., Nebel, A., Guenther, S.P., Bagaev, E., Rottbauer, W., Kääh, S., Just, S. and Rappold, G.A. (2016) Coding and non-coding variants in the SHOX2 gene in patients with early-onset atrial fibrillation. *Basic Res Cardiol*, 111(3) : 36. PDF de l'article
5. Tarver, J. E., Cormier, A., Pinzón, N., Taylor, R. S., Carré, W., Strittmatter, M., Seitz, H., Coelho, S. M. and Cock, M. (2015) microRNAs and the evolution of complex multicellularity : identification of a large, diverse complement of microRNAs in the brown alga *Ectocarpus*. *Nucleic Acids Res*, 43(13) : 6384–6398. PDF de l'article

Conférences invitées dans des congrès

1. Le 12 février 2016, au congrès *Statistical Methods for Post Genomic Data* à Lille (**keynote speaker**).
2. Le 27 juin 2016, à l'*International Symposium "Clues to Non-coding RNA Taxonomy"* à Tōkyō (Japon).
3. Le 3 mars 2016, au *ECETOC Noncoding RNAs and risk assessment science workshop* à Málaga (Espagne).
4. Le 4 juin 2015, à l'*ESPE science school "Non-coding RNAs in paediatric endocrinology"* à Annecy.
5. Le 9 juin 2014, au *Taiwan-France mini-symposium on regulatory RNAs* à Taipei (Taïwan).

Communications à des congrès, symposiums

1. Le 2 février 2014, au *Keystone Symposium "RNA Silencing (A9)"* à Seattle (États-Unis).

Séminaires, workshops

1. Le 4 janvier 2016, à l'École normale supérieure (Paris), sur invitation d'É. Meyer.
2. Le 31 octobre 2014, au Max Planck Institute for Infection Biology (Berlin) sur invitation de L. Lampe.
3. Le 1er décembre 2015, à l'IBMC (Strasbourg), sur invitation de S. Blandin.
4. Le 9 décembre 2014, à Medesis Pharma S.A., sur invitation de l'entreprise.

3 Enseignement, formation et diffusion de la culture scientifique

3.1 Thèses et post-doctorats

J'encadre le post-doctorat de Mlle. Natalia Pinzón Restrepo depuis mars 2012. Mlle. Pinzón Restrepo a mesuré l'abondance de microARN et des ARNm (en mesure absolue : elle a déterminé le nombre de leurs molécules par cellule) au cours de la différenciation des cellules C2C12 (une lignée cellulaire murine, dont on peut induire la différenciation en myotubes en culture) (voir Figure 3, page 8). Elle a également muté (par le système CRISPR/Cas9, puis recombinaison homologue) le site de reconnaissance d'un microARN dans le gène *Tmsb4x* dans cette lignée (voir Figure 4, page 9). Elle a également été impliquée dans deux collaborations de notre équipe (avec le laboratoire de M. Méchali, à Montpellier, et avec le laboratoire de M. Cock, à Roscoff), où ses travaux ont consisté respectivement à analyser des données génomiques issues d'expériences à haut débit chez le Nématode, et à confirmer expérimentalement l'existence de petits ARN chez l'Algue brune *Ectocarpus siliculosus*. Mlle. Pinzón Restrepo était financée par une allocation de la Ligue contre le cancer de mars 2013 à février 2016, et son contrat a été prolongé jusqu'au 15 septembre 2016 par des fonds de mon unité, l'IGH (UPR 1142). La qualité du travail de Mlle. Pinzón Restrepo lui a valu de présenter ses résultats oralement au *Cold Spring Harbor Laboratory Meeting "Regulatory and non-coding RNAs"* en août et septembre 2012, puis au congrès SifrARN de la Société française de biochimie et de biologie moléculaire en novembre 2013, et au *Keystone Symposium "Small RNA Silencing : Little Guides, Big Biology (A6)"* en janvier 2016. Mlle. Pinzón Restrepo a candidaté à des postes d'ingénieur de recherche (au laboratoire « Évolution et diversité biologique » à Toulouse) et de chercheuse (au CIRAD, à Montpellier), pour lesquelles la procédure de sélection est encore en cours.

De novembre 2013 à juin 2016, j'ai co-encadré le post-doctorat de M. Blaise Li (co-encadré avec ma collègue Séverine Chambeyron, qui dirige une autre équipe de l'IGH). Dans le cadre de son travail pour mon équipe, M. Li a vérifié, par des implémentations indépendantes, les résultats présentés en Figure 5, page 11. Il a également comparé la conservation des sites de complémentarité aux microARN (situés dans des ARNm) à la conservation des microARN eux-mêmes (voir Figure 6, page 13). Il a été impliqué dans la collaboration entre mon laboratoire et celui d'A. Spradling (Carnegie Institution, Baltimore, États-Unis), en analysant des jeux de données de Small RNA-Seq d'ovaires de *Drosophile*. M. Li est, depuis juin 2016, post-doctorant à l'Institut Pasteur à Paris, dans le laboratoire de G. Cecere.

3.2 Participation à l'enseignement

J'ai dispensé des cours sur les petits ARN régulateurs eucaryotiques (miRNA, siRNA et piRNA) dans divers établissements au cours des 5 derniers semestres :

- Deux cours de trois heures chacun, pour les étudiants de L3 de l'École normale supérieure, à Paris (titres : « Petits ARN régulateurs chez les Animaux » et « Revisiter la notion de cible des miARN ») (21 février 2014).
- Deux cours d'une heure chacun, pour les étudiants de master et doctorat de la National Taiwan University, à Taipei (titres : "Small regulatory RNAs in eukaryotes" et "Revisiting the notion of microRNA target") (4 juin 2014).
- Cours d'une heure sur les microARN et leur impact physiologique, pour les étudiants en thèse du Max Planck Institute for Infection Biology de Berlin (titre : "On the robustness of microRNA-mediated regulation : a new understanding of the nature of mRNA/microRNA interaction") (31 octobre 2014).
- Cours de deux heures sur les petits ARN régulateurs, pour les étudiants de M2 de l'université de Montpellier (titre : « Petits ARN régulateurs chez les Animaux ») (3 décembre 2014).
- Deux cours de trois heures chacun, pour les étudiants de L3 de l'École normale supérieure, à Paris (titres : « Petits ARN régulateurs chez les Animaux » et « Revisiter la notion de cible des miARN ») (27 février 2015).
- Cours de trois heures, pour les étudiants de M1 de l'École normale supérieure, à Paris (titre : « miRNA in animals ») (5 janvier 2016).
- Deux cours de trois heures chacun, pour les étudiants de L3 de l'École normale supérieure, à Paris (titres : « Petits ARN régulateurs chez les Animaux » et « Revisiter la notion de cible des miARN ») (13 mai 2016).

Suite au cours du 3 décembre 2014, j'ai été chargé de proposer un sujet de dissertation aux étudiants du

M2 de l'université de Montpellier, et j'ai corrigé les copies.

Au cours des 5 derniers semestres, j'ai été rapporteur de dix thèses :

- Amélie Juanchich (directeur de thèse : Julien Bobe, INRA, Rennes ; 23 janvier 2014) ;
- Thomas Clément (directeurs de thèse : Bruno Charpentier et Jean-Yves Jouzeau, CNRS et université de Lorraine, Vandoeuvre-lès-Nancy ; 10 septembre 2014) ;
- Quentin Carradec (directeurs de thèse : Simone Marker, université de Saarland, Allemagne, et Éric Meyer, École normale supérieure, Paris ; 29 septembre 2014) ;
- Susan Higashi (directeurs de thèse : Marie-France Sagot, Christian Gautier et Stefano Colella, INRIA Lyon ; 26 novembre 2014) ;
- Cécile Quéré (directeur de thèse : Denis Dupuy, Inserm, Bordeaux ; 17 décembre 2014) ;
- Andrew Calcino (directeur de thèse : Bernard Degnan, *University of Queensland*, Australie ; janvier 2015) ;
- Catherine Hermant (directeur de thèse : Stéphane Ronsseray, CNRS et UPMV ; 28 janvier 2015) ;
- Marius van den Beek (directeur de thèse : Christophe Antoniewski, CNRS et UPMC ; 9 février 2015) ;
- Anaïs Vittu (directrice de thèse : Stéphanie Blandin, Inserm ; 1er décembre 2015) ;
- Chloé Mengardi (directeur de thèse : Théophile Ohlmann, Inserm ; 22 janvier 2016).

J'ai également été examinateur d'une thèse :

- Adrien Le Thomas (directeurs de thèse : Déborah Bouch'his, Institut Curie, Paris, et Alexei Aravin, *California Institute of Technology*, Pasadena, États-Unis ; 11 septembre 2014).

3.3 Diffusion de l'information scientifique

Alors que la génétique prend de plus en plus de place dans les thèmes de société (à propos des OGM, du clonage, des actions caritatives à visée médicale, ...), les citoyens et électeurs sont très insuffisamment informés (la génération des *baby-boomers* n'a pas entendu parler de chromosome ou d'ADN au lycée ; seuls ceux qui ont fait des études supérieures en biologie savent à peu près ce qu'est cette double hélice qu'on leur montre dans les médias pour illustrer toutes sortes de mauvaises nouvelles). Quant aux jeunes générations, elles entendront parler de génétique pendant leurs études secondaires, mais elles sont déjà confrontées à ces informations – et elles sont curieuses de les comprendre. Le public est donc demandeur d'informations sur ce que nous faisons, et il est du devoir des chercheurs publics, employés par la collectivité, de partager avec elle les résultats de leur travail. Je tâche donc, autant que possible, de contribuer à la diffusion de l'information scientifique en direction du grand public.

Dans le cadre de la Fête de la Science, je donne chaque année une conférence intitulée « Qu'est-ce que la génétique ? » (durée : 45 minutes) à destination du grand public et des scolaires (donnée 2 fois en 2014, 6 fois en 2015).

J'ai également donné cette conférence à l'Université du Tiers Temps à Montpellier (série de conférences à destination du grand public, essentiellement retraité), le 8 décembre 2014.

J'ai donné deux conférences pour le grand public, dans le cadre de l'association « Rando loisirs culture » de Grabels (34) : conférences d'une heure chacune, suivie d'une heure de discussion avec l'audience (« Qu'est-ce que la génétique ? » le 4 mars 2015 ; « Comment l'ADN détermine le vivant » le 13 janvier 2016).

J'ai donné une conférence d'une heure devant le club « Sciences » du collège Clémence Royer (Montpellier) le 24 mai 2016 (titre : « Utilisation de la génétique de la *Drosophile* pour explorer les mécanismes biologiques »).

3.4 Interventions dans la presse écrite et audiovisuelle

Interview par David Larousserie pour un article du *Monde* au printemps 2015 à propos de la probité dans la recherche en biologie.

Interventions dans les magazines en ligne *Néon* (une contribution en 2014) et <http://www.atlantico.fr> (une contribution en 2015, quatre en 2016).

3.5 Participation à des travaux d'expertise

Évaluateur pour des journaux dans le domaine de l'ARN, de la bio-informatique et de la génétique moléculaire (*BMC Genomics*, *Current Biology*, *Nucleic Acids Research*, ...).

Évaluateur pour l'ANR (2014 : panel CE10 ; 2016 : panel CE12), pour l>IDEX « Super » de Sorbonne Universités (2014) et pour l'Institut universitaire de France (2015).

Participation à des comités d'expertise :

- Depuis 2016 : membre du comité de sélection des projets de la plateforme ABIC de Montpellier.
- Depuis 2016 : membre de la “Governance Leadership Team” du Cefic Long-range research initiative pour les domaines “omics, epigenetics and non-coding RNAs” pour l'amélioration des procédures d'évaluation des risques toxicologiques.

4 Transfert technologique, relations industrielles et valorisation

J'ai participé, au cours de cette période, à deux contrats de recherche :

- ATIP-Avenir du CNRS et de l'Inserm, avec financement complémentaire de la part de Sanofi (partenaire du programme ATIP-Avenir) : de mars 2012 à février 2015 (+ extension jusqu'au 31 décembre 2016). Rôle : coordonnateur du projet. Partenaire : aucun. Montant : 120 000 € par an, pendant les trois ans. Travaux effectués : mesure de la robustesse des cibles prédites de miR-223 dans les neutrophiles de Souris, vis-à-vis de la répression guidée par ce miARN. Mesure de l'effet titrateur de miARN par un ARNm, par mutagenèse ciblée de son site d'interaction. Comparaison de l'abondance de miARN et de leurs cibles prédites au cours de la différenciation des cellules C2C12. Mesure du taux de faux positifs dans les approches de prédiction de cibles de miARN par génomique comparative. Ces résultats suggèrent que la fonction biologique des miARN a été très largement sur-estimée, et que les outils expérimentaux et bio-informatiques couramment utilisés sont lourdement biaisés.
- Projet ANR « miRNAdapt » (coordonnateur : Denis Tagu, INRA). Rôle : partenaire. Montant : 21 000 €. Mes travaux ont consisté à participer aux prédictions de cibles de miARN chez le Puceiron *Acyrtosiphon pisum*, et à développer des critères de sélection des cibles les plus probablement fonctionnellement ciblées par les miARN.

J'ai été invité, en tant qu'expert des petits ARN non-codants, à un atelier organisé par l'ECETOC, un consortium européen financé par l'industrie chimique et qui cherche à améliorer les procédures d'évaluation de la toxicité des nouveaux produits. Les microARN circulants (facilement accessibles, ils peuvent être dosés sans sacrifier l'animal) pourraient leur permettre d'améliorer les analyses toxicologiques, en réduisant la durée des essais, leur coût, et la souffrance animale. J'ai donc donné une conférence sur les microARN et participé aux réflexions de l'atelier organisé à Málaga les 3 et 4 mars 2016. À la demande de la direction d'ECETOC, j'ai ensuite accepté de participer à la "Governance Leadership Team" pour améliorer l'utilisation des données à haut débit, des données d'épigénétique, et des données d'expression des ARN non-codants, dans la caractérisation de la toxicité *in vivo*.

À la demande de l'Anses, j'ai corrigé un bulletin de veille scientifique qu'elle allait publier concernant deux études récentes, et qui traitent de l'utilité des nouveaux biomarqueurs (ARN non-codants, marques épigénétiques) dans les analyses toxicologiques.

5 Encadrement, animation et management de la recherche

Je dirige depuis octobre 2011 une équipe de recherche à l'IGH (UPR 1142 du CNRS). L'évolution de l'effectif de l'équipe est schématisée en Figure 7 : l'équipe actuelle compte 3 membres (une post-doctorante, une chargée de recherche, et moi-même), outre les étudiants stagiaires à court terme.

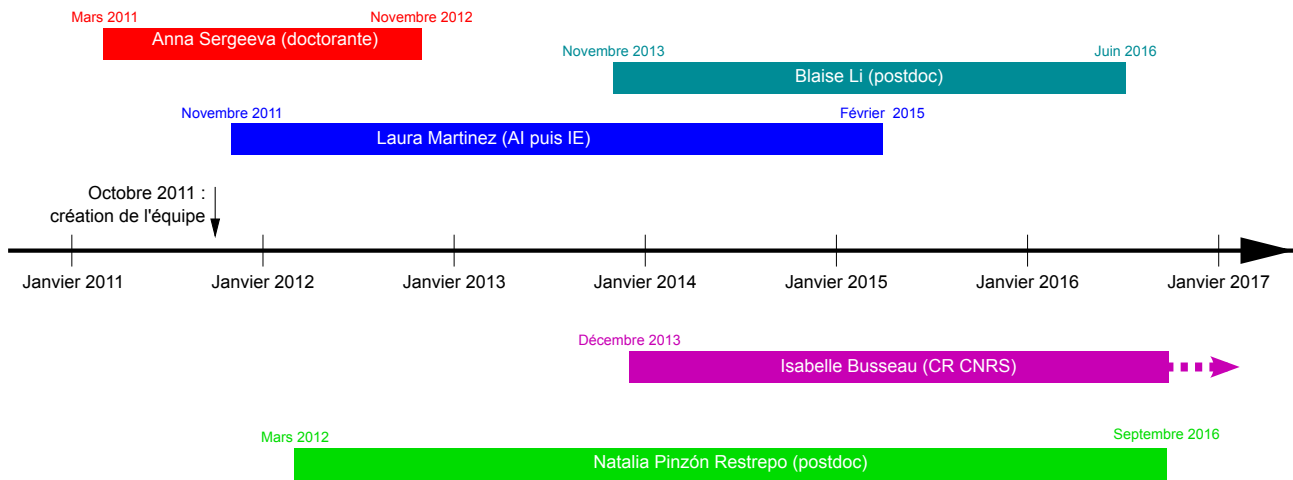


FIGURE 7 – **Effectifs de l'équipe.** Les étudiants de niveau master ou inférieur, et le chef d'équipe, ne sont pas représentés.