



Concours externes 2025 INRAE

Candidature pour un poste de Directeur de Recherche de 2ème classe (DR2)

Rapport Scientifique

Sylvain Foissac, INRAE Toulouse

Table des matières

1	<i>Curriculum Vitae</i>	2
2	Mémoire	4
2.1	Travaux et production scientifique	4
a.	L'annotation génomique	4
	La thèse : annotation de gènes codants chez les plantes	4
	L'épopée ENCODE post-thèse : annotation du génome humain	5
	Annotation génomique des animaux d'élevage	6
b.	La génomique 3D	8
	Contexte scientifique : technologie et données Hi-C	8
	Application de la génomique 3D aux animaux d'élevage	8
	Vers l'analyse différentielle de données Hi-C	10
2.2	Animation scientifique et responsabilités collectives	10
a.	Collectifs scientifiques : contributions et animations	11
b.	Responsabilités collectives et instances	11
c.	Formation à la recherche par l'encadrement	12
3	Projet de recherche. Génomique structurale et fonctionnelle : analyses de données omiques pour mieux comprendre le lien génotype-phénotype.	14
3.1	Contexte et enjeux	14
3.2	Programme scientifique du projet	15
a.	L'annotation génomique pour améliorer la compréhension des génomes animaux	15
	Annotation génomique : exploitation des ressources disponibles	15
	De l'annotation structurale à l'annotation fonctionnelle	15
b.	Analyse de données omiques : vers une génomique fonctionnelle comparative . .	16
	Analyse différentielle de données omiques	16
	Analyse inter-espèces : vers l'épigénomique comparative	16
3.3	Pilotage de la recherche et animation scientifique	17
4	Liste des productions	19
a.	Articles publiés dans des revues internationales à comité de lecture	19
b.	Chapitres d'ouvrage	22
c.	Communications dans des conférences	22
d.	Jeux de données	23
e.	Logiciels	24
	Bibliographie	25

1) Curriculum Vitae

SYLVAIN FOISSAC, PhD, HDR

INRAE GenPhySE, ch. de Borde Rouge, 31326 Castanet-Tolosan, France

<http://genoweb.toulouse.inra.fr/~sfoissac>

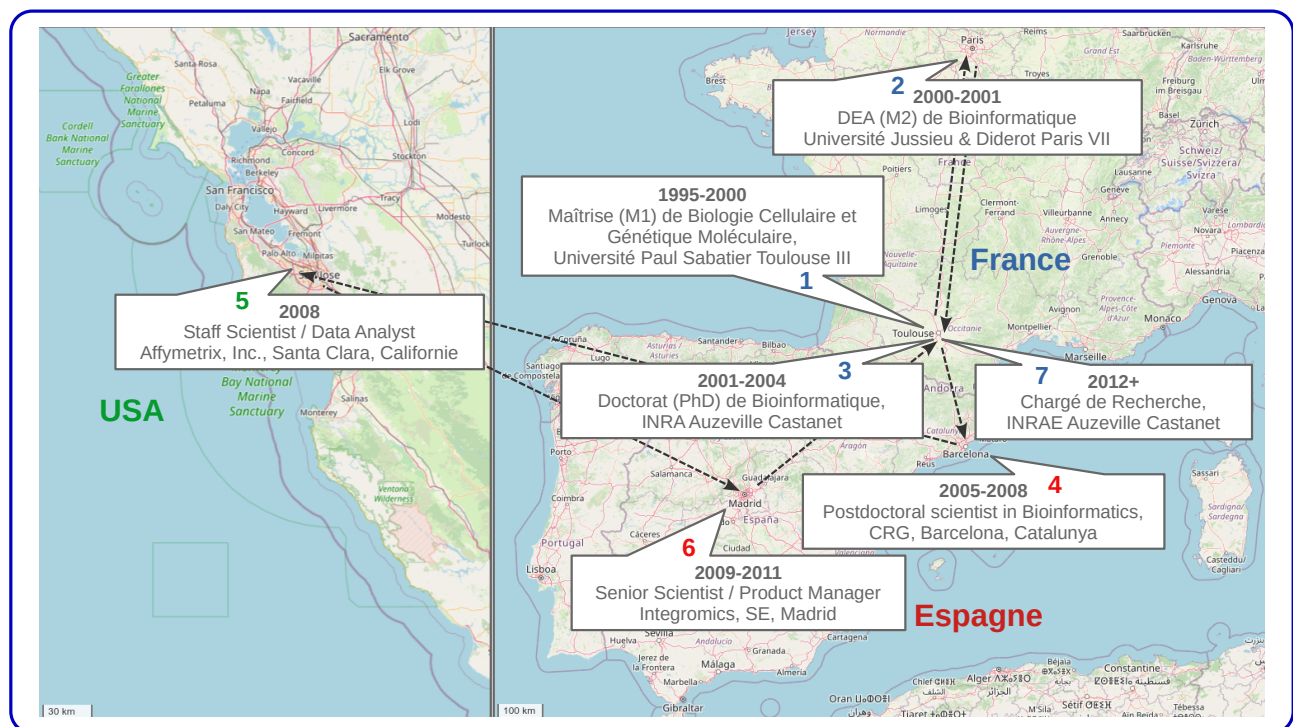
sylvain.foissac@inrae.fr

Summary

- About **20 years of experience in computational biology and omics data analysis**.
- Scientific expertise in **structural and functional genomics**.
- Management experience in **academia and industry**.
- Publication record of **33 peer-reviewed articles** and 3 book chapters.

Experience

- Nov 2012 - present : Research Scientist (“CRCN”) at **INRAE** in Toulouse, France.
- Dec 2010 - Oct 2012 : Product Manager at **Integromics, S.L.** in Madrid, Spain.
- Dec 2008 - Dec 2010 : Senior Research Scientist at **Integromics, S.L.** in Madrid, Spain.
- Apr 2008 - Nov 2008 : Data Analyst at **Affymetrix, Inc.** in Santa Clara, California, USA.
- Apr 2005 - Mar 2008 : PostDoctoral researcher at the **CRG** in Barcelona, Catalunya, Spain.



Education

- 2024 : HDR (“Habilitation à Diriger des Recherches”). Toulouse, France.
- 2001-2004 : PhD thesis in Bioinformatics. Toulouse 3 University, France.

- 2000-2001 : Master degree in Bioinformatics. Paris VI & VII University, France.
- 1995-2000 : Bachelor degree in Cellular Biology. Toulouse 3 University, France.

Publications

- Karami et al (2025). **Environmental Epigenetics**. [10.1093/eep/dvaf009](#).
- Jorge et al (2025). **Briefings in Bioinformatics**. [10.1093/bib/bbaf074](#).
- Degalez et al (2024). **Scientific Reports**. [10.1038/s41598-024-56705-y](#).
- Neuvial et al (2024). **JRSSS Applied Statistics**. [10.1093/jrsssc/qlae011](#).
- Dufour et al (2024). **Genomics**. [10.1016/j.ygeno.2023.110780](#).
- Kurylo et al (2023). **NAR Genomics And Bioinformatics**. [10.1093/nargab/lqad089](#).
- Hoellinger et al (2023). **Frontiers in Bioinformatics**. [10.1093/nar/gkx920](#).
- Jehl et al (2021). **Frontiers in Genetics**. [10.3389/fgene.2021.655707](#).
- Marti-Marimon et al (2021). **Frontiers in Genetics**. [10.3389/fgene.2021.748239](#).
- Jehl et al (2020). **Scientific Reports**. [10.1038/s41598-020-77586-x](#).
- Giuffra et al (2019). **Ann. Rev. of Anim. Biosc.**. [10.1146/annurev-animal-020518-114913](#).
- Foissac et al (2019). **BMC Biology**. [10.1186/s12915-019-0726-5](#).
- Fève et al (2017). **PLoS ONE**. [10.1371/journal.pone.0187617](#).
- David et al (2017). **Epigenomes**. [10.3390/epigenomes1030020](#).
- Muret et al (2017). **Genetics Selection Evolution**. [10.1186/s12711-016-0275-0](#).
- Andersson et al (2015). **Genome Biology**. [10.1186/s13059-015-0622-4](#).
- Rubio-Peña et al (2015). **RNA**. [10.1261/rna.053397.115](#).
- Bonnet et al (2013). **BMC Genomics**. [10.1186/1471-2164-14-904](#).
- Djebali et al (2012). **Nature**. [10.17615/r8aq-6h68](#).
- Djebali et al (2012). **PLoS ONE**. [10.1371/journal.pone.0028213](#).
- The ENCODE Project Consortium (2012). **Nature**. [10.1038/nature11247](#).
- Ozsolak et al (2010). **Cell**. [10.1016/j.cell.2010.11.020](#).
- Kapranov et al (2010). **Nature**. [10.1038/nature09190](#).
- Fejes-Toth et al (2009). **Nature**. [10.1038/nature07759](#).
- Djebali et al (2008). **Nature Methods**. [10.1038/nmeth.1216](#).
- Piqué et al (2008). **Cell**. [10.1016/j.cell.2007.12.038](#).
- Sammeth et al (2008). **PLoS Computational Biology**. [10.1371/journal.pcbi.1000147](#).
- Foissac et al (2008). **Current Bioinformatics**. [10.2174/157489308784340702](#).
- Foissac et al (2007). **Nucleic acids research**. [10.1093/nar/gkm311](#).
- The ENCODE Project Consortium (2007). **Nature**. [10.1038/nature0587](#).
- Denoeud et al (2007). **Genome Research**. [10.1101/gr.5660607](#).
- Foissac et al (2005). **BMC Bioinformatics**. [10.1186/1471-2105-6-25](#).
- Foissac et al (2003). **Nucleic Acids Research**. [10.1093/nar/gkg586](#).

2) Mémoire

2.1 Travaux et production scientifique

PRÉAMBULE

Mes travaux s'inscrivent dans deux thématiques principales de recherche qui constitue mon domaine d'expertise : l'annotation génomique et la génomique tridimensionnelle. L'annotation implique l'ensemble de ma carrière depuis la thèse, la spécialisation en génomique 3D concernant la dernière décennies environ. Cette section présente mes activités de recherche dans ces deux thématiques.

a. L'annotation génomique

Comprendre la structure et la fonction des génomes animaux fait partie des grands objectifs scientifiques du département de Génétique Animale à l'INRAE. L'annotation génomique, qui consiste à identifier, positionner et caractériser les régions fonctionnelles¹ du génome, s'inscrit dans cette dynamique en améliorant la cartographie génétique des espèces d'élevage. Les éléments fonctionnels ciblés par l'annotation sont en priorité (1) les gènes, qu'ils produisent des transcrits longs ou courts, codant ou non pour des protéines, et (2) les éléments régulant leur expression, source potentielle de diversité génétique et phénotypique. Au-delà de contribuer aux connaissances fondamentales en génomique régulationnelle, l'objectif à terme est de mieux caractériser les déterminants génétiques des caractères à intérêt agronomique, d'augmenter la capacité de prédiction "génotype → phénotype", et d'améliorer les schémas de sélection.

Mes activités d'annotation couvrent trois étapes principales : la thèse, l'après-thèse (projet ENCODE surtout, en post-doc et dans le privé), et la période INRAE.

La thèse : annotation de gènes codants chez les plantes

FICHE SYNTHÉTIQUE

- **Période et lieu** : 2001-2004, INRA Toulouse Auzeville (thèse)
- **Génome** : plantes, surtout *Arabidopsis thaliana* (arabette) et *Oryza sativa* (riz)
- **Cible** : gènes et transcrits codants
- **Données traitées** : génomes et transcriptomes (séquençage Sanger)
- **Contribution principale** : méthodologie, algorithmique, développement logiciel

1. à l'échelle génomique, on considère qu'une région est "fonctionnelle" quand on peut lui attribuer une activité moléculaire mesurable de façon significative, spécifique et reproductible.

• **Publications** : Foissac et al. 2003 ; Foissac et al. 2005 ; Foissac et al. 2008

Mes premiers travaux concernent l’annotation de gènes codant pour des protéines. L’objectif de la détection de gènes est d’identifier la position et la composition exon-intron de tous les gènes présents dans une séquence génomique donnée. Lors de ma thèse à l’INRA de Toulouse, j’ai travaillé dans le cadre du logiciel EUGÈNE, initialement utilisé pour annoter le génome de la plante modèle *Arabidopsis thaliana* (arabette).

La méthode initialement implémentée dans le logiciel présentait deux limitations majeures, communes aux autres outils du même type : (1) elle ne pouvait produire qu’une seule prédiction par séquence, ne permettant donc pas d’identifier plusieurs transcrits par gène. (2) elle faisait l’hypothèse, en accord avec les connaissances d’alors, que les gènes étaient des entités distinctes et non chevauchantes.

Or, les connaissances produites par l’arrivée massive de nouvelles données “omiques” (ici transcriptomique) remettaient ces choix en questions. J’ai donc conçu, implémenté et appliqué des améliorations méthodologiques apportant des solutions pour dépasser ces limites, avec par exemple une modification du modèle de graphe au coeur de la méthode et la création d’un algorithme de programmation dynamique permettant l’extension du logiciel au cas de l’épissage alternatif (Foissac et al. 2003 ; Foissac 2004 ; Foissac et al. 2005 ; Foissac et al. 2008).

L’épopée ENCODE post-thèse : annotation du génome humain

FICHE SYNTHÉTIQUE

- **Période et lieu** : 2005-2012, CRG (Barcelone, Catalogne), Affymetrix (Santa Clara, USA), Integromics (Madrid, Espagne)
- **Génome** : principalement *Homo sapiens* (humain), mais aussi *Mus musculus* (souris), *Saccharomyces cerevisiae* (levure) et *Xenopus laevis* (batracien).
- **Cible** : gènes et transcrits courts et longs, codants et non codants
- **Données traitées** : génomes, annotations et transcriptomes (données de puces *tiling arrays* puis de séquençage RNA-seq)
- **Contribution principale** : méthodologie, programmation, analyses de données
- **Publications** : The ENCODE Project Consortium 2007 ; Denoeud et al. 2007 ; Foissac et al. 2007 ; Djebali et al. 2008 ; Piqué et al. 2008 ; Fejes-Toth et al. 2009 ; Sammeth et al. 2008 ; Oszolak et al. 2010 ; Kapranov et al. 2010 ; Djebali et al. 2012b ; Foissac et al. 2015

Ma thèse a pris fin au début d’une nouvelle ère de la génomique, marquée par une augmentation considérable de la variété et du volume des données omiques produites par des technologies toujours plus innovantes. Or, ces nouvelles données mettaient à mal la vision du génome alors en vigueur (concept de gènes distincts séparés par de l’“ADN poubelle”). Afin d’améliorer les connaissances dans le domaine, j’ai poursuivi mes travaux de thèse par un post-doc au CRG de Barcelone (2005-08), intégrant ainsi le consortium international ENCODE dédié à l’annotation du génome humain. J’ai pu ainsi avoir accès à des données de qualité et quantité exceptionnelles, construire un large réseau de collaborations et contribuer à la production de connaissances en front de science. J’ai ensuite travaillé dans le privé sur des postes de type CDI, d’abord en tant que *Staff Scientist* chez Affymetrix (2008), une multinationale biotech de la Silicon Valley en Californie, puis comme chercheur (2008-10) et *Product Manager* (2010-12) à Integromics, une entreprise de logiciels bioinformatiques en Espagne. Par le biais

de diverses collaborations internationales, mes travaux de recherche ont ainsi contribué à améliorer la compréhension du génome humain en révélant notamment l'importance de deux phénomènes : l'évanescence des frontières géniques et la prévalence des mécanismes alternatifs.

Évanescence des frontières géniques Avec des collègues du consortium international ENCODE, nous utilisons des puces à ADN de type *tiling arrays* d'Affymetrix sur de l'ADN complémentaire issu d'ARN messagers afin de réaliser des cartes transcriptomiques du génome humain. Cette technologie de pointe à l'époque permettait de sonder l'ensemble du génome et de détecter ainsi toute région témoignant d'une activité de transcription, sans besoin d'annotation préalable informant sur la position des gènes. Or, bien qu'une partie de ces régions correspondait bien aux exons de transcrits connus, la plupart (63%) révélait un signal de transcription dans des régions encore non annotées, augmentant considérablement la portion à considérer comme transcrite du génome ainsi que la complexité du paysage transcriptionnel connu (The ENCODE Project Consortium 2007). Des expériences ultérieures de *RACEarray*, destinées à caractériser les limites des gènes en combinant *Rapid Amplification of cDNA Ends* et *tiling array*, ont révélé le caractère fugitif de ces frontières, au point de remettre en question la validité du concept même de frontière génique (The ENCODE Project Consortium 2007 ; Denoeud et al. 2007 ; Djebali et al. 2008 ; Djebali et al. 2012b).

Prévalence des mécanismes alternatifs et du non codant La multiplicité des nouveaux transcrits découverts soulevait la question de leur biogénèse et de leur fonction. Pour mieux comprendre la complexité des annotations produites, je me suis principalement intéressé au processus de maturation des ARN, source majeure de diversité transcriptomique.

- La plupart des transcrits contenant plusieurs exons, j'ai étudié le mécanisme d'épissage alternatif en analysant les annotations génomiques de plusieurs espèces. Ces travaux ont établi une nomenclature originale permettant une classification exhaustive des événements d'épissage alternatifs (Foissac et al. 2007 ; Sammeth et al. 2008 ; Foissac et al. 2015).
- En parallèle, je me suis penché sur les variations des parties terminales des ARNs et sur leur impact fonctionnel, notamment le rôle des 3'UTR dans la régulation de la traduction (Piqué et al. 2008). Des analyses de données produites par *Helicos Single-Molecule Sequencing*, la première technologie de séquençage direct d'ARN sans amplification (bien avant l'arrivée de PacBio/ONT), ont révélé l'étendue de ces variations (Ozsolak et al. 2010) ainsi qu'un nouveau mécanisme moléculaire de copie d'ARN (Kapranov et al. 2010).
- Enfin, poursuivant l'axe de l'annotation fonctionnelle, j'ai contribué à la caractérisation des transcrits ne codant pas pour des protéines (The ENCODE Project Consortium 2007 ; Djebali et al. 2012b ; The ENCODE Project Consortium 2012 ; Djebali et al. 2012a). En particulier, les analyses que j'ai réalisées sur des séquences Illumina (alors Solexa) issues petits ARN non-codants ont révélé un nouveau mécanisme de maturation post-transcriptionnelle des ARN (Fejes-Toth et al. 2009).

Annotation génomique des animaux d'élevage

FICHE SYNTHÉTIQUE
• Période et lieu : 2012-présent, INRAE Toulouse

- **Génomes** : *B. taurus* (vache), *G. gallus* (poule), *S. scrofa* (porc), *C. hircus* (chèvre)
- **Cible** : gènes et transcrits, éléments régulateurs
- **Données traitées** : RNA-seq, small RNA-seq, ATAC-seq, Hi-C, WGBS, PC-HiC
- **Contributions** : analyses de données, montage et gestion de projet, management
- **Publications** : Andersson et al. 2015 ; Muret et al. 2017 ; Foissac et al. 2019 ; Giuffra et al. 2019 ; Jehl et al. 2020 ; Jehl et al. 2021 ; Kurylo et al. 2023 ; Degalez et al. 2024

Depuis ma prise de fonction à l'INRAE, j'ai mis à disposition mon expertise en analyse de données omiques pour l'annotation génomique d'animaux à intérêt agronomique, prolongement naturel de mes travaux de recherche. Toujours dans une démarche *FAIR* de science ouverte, je privilégie le développement de code *open source*, le dépôt systématique de l'ensemble des données produites dans des bases publiques et les publications *open access*.

- Dans le contexte du consortium international FAANG (*Functional Annotation of ANimal Genomes*), pendant d'ENCODE pour les animaux, j'ai co-monté et co-dirigé le projet pilote français FR-AgENCODE. Ce projet risqué et ambitieux, financé à 300 KE par le métaprogramme INRAE SelgGen, visait à améliorer l'annotation génomique de quatre espèces (vache, poule, porc et chèvre) par le biais d'expériences omiques à haut débit réalisées sur une centaine d'échantillons, impliquant plus de cinquante agents répartis sur une douzaine d'unités et installations expérimentales en France.

Responsable du WP d'analyse de données, j'ai piloté la conception, l'implémentation et la valorisation de toutes les analyses de séquences produites par le projet. Ainsi, nous avons réalisé la première annotation fonctionnelle combinant expression génique (RNA-seq), accessibilité de la chromatine (ATAC-seq) et conformation tridimensionnelle (Hi-C) sur ces génomes (Foissac et al. 2019). En outre, le succès de ce projet pilote a confirmé le positionnement de l'INRAE au coeur de la coordination internationale FAANG (Andersson et al. 2015 ; Foissac et al. 2019 ; Giuffra et al. 2019).

- En parallèle, par le biais de ces travaux et du réseau de collaborations développé lors de cette initiative, j'ai contribué à la caractérisation du transcriptome non codant et de l'épigénome de ces espèces, en particulier chez la poule (Muret et al. 2017 ; Foissac et al. 2019 ; Jehl et al. 2020 ; Jehl et al. 2021 ; Degalez et al. 2024 ; Karami et al. 2025).
- Enfin, dans le cadre notamment du projet EU H2020 GENE-SWitCH, une suite du projet pilote FR-AgENCODE, j'ai co-piloté l'équipe française du WP2 *Bioinformatics and Data Analysis* et coordonné l'analyse des données omiques produites (RNA-seq, sRNA-seq, ATAC-seq, cHi-C) ainsi que leur valorisation et les échanges avec les partenaires (WUR, Roslin, EMBL/EBI). J'ai notamment co-supervisé le développement d'un outil original d'analyse de données massives RNA-seq, premier pipeline bioinformatique open source capable de combiner de façon robuste, portable et reproductible les technologies *short reads* Illumina et *long reads* PacBio pour améliorer des annotations génomiques existantes (Kurylo et al. 2023). L'ensemble de ces analyses, toujours en cours, a déjà produit des centaines d'annotations fonctionnelles du génome de la poule et du porc, permettant d'identifier les régions impliquées dans l'expression génique et sa régulation pendant le développement embryonnaire.

b. La génomique 3D

PRÉAMBULE

Les derniers progrès technologiques qui ont le plus influencé l'orientation de mes recherches concernent l'étude de la conformation tridimensionnelle de la chromatine. En effet, l'arrivée récente de données massives de génomique 3D a apporté une nouvelle dimension à la compréhension du génome et de son fonctionnement. Cette partie présente ce nouveau domaine de recherche que j'ai choisi d'investir.

Contexte scientifique : technologie et données Hi-C

La longueur cumulée de l'ADN présent dans une seule cellule animale atteint l'ordre du mètre, quand le noyau qui le contient fait moins d'un centième de millimètre. Or, les contraintes s'exerçant sur la conformation spatiale de la chromatine relèvent autant de la structure que de la fonction du génome. En effet, les processus moléculaires fondamentaux du vivant, comme la réplication et la transcription, dépendent pour leur bon fonctionnement de modifications coordonnées de la structure de la chromatine. La structure 3D du génome joue ainsi un rôle crucial dans le fonctionnement de la cellule (Lupiáñez et al. 2015 ; Oudelaar et al. 2020).

Or, si la microscopie fut longtemps un moyen privilégié pour explorer cette structure 3D (Davies et al. 2017), la technologie de séquençage Hi-C (*High-throughput Chromatin Conformation Capture*) a désormais établi un nouveau standard dans le domaine, de par ses avantages en débit et résolution (Davies et al. 2017 ; Kempfer et al. 2019). Pour résumer, le protocole Hi-C consiste à réaliser des interactions entre régions génomiques non ciblées, sur la base de leur proximité spatiale dans le noyau : en gros, plus deux régions sont proches, plus elles risquent de se retrouver en interaction. Ces interactions, typiquement réalisées sur des centaines de milliers de cellules, sont ensuite détectées et quantifiées par séquençage, produisant des paires de lectures associant des régions génomiques en interaction.

Après traitement des séquences obtenues, les données Hi-C se présentent donc sous la forme d'une matrice d'interactions (ou matrice Hi-C), indiquant pour chaque paire de positions génomiques le nombre d'interactions détectées les impliquant. En prenant ces comptages comme une estimation (*proxy*) de la distance spatiale entre régions, on peut considérer la matrice Hi-C comme une projection en 2D d'une information 3D. En général, les matrices sont symétriques, creuses, et montrent un signal dont l'intensité (1) se concentre sur la diagonale, du fait de la nécessaire proximité spatiale entre régions adjacentes le long du génome, et (2) diminue en s'éloignant de la diagonale, faisant de la distance génomique entre régions le déterminant principal du nombre d'interactions (Fig. 2.1 et 2.2).

L'analyse de ces données peut permettre d'identifier divers types de structures connues de la chromatine qui jouent un rôle essentiel dans le fonctionnement du génome : compartiments chromatiniens, domaines topologiques TADs (*Topologically Associating Domains*) et boucles de chromatine. Pour chaque type de structure, des méthodes et outils ont été développés pour permettre leur détection à partir de données Hi-C, avec plus ou moins de succès (Dali et al. 2017 ; Zufferey et al. 2018 ; Marti-Marimon et al. 2021 ; Liu et al. 2023).

Application de la génomique 3D aux animaux d'élevage

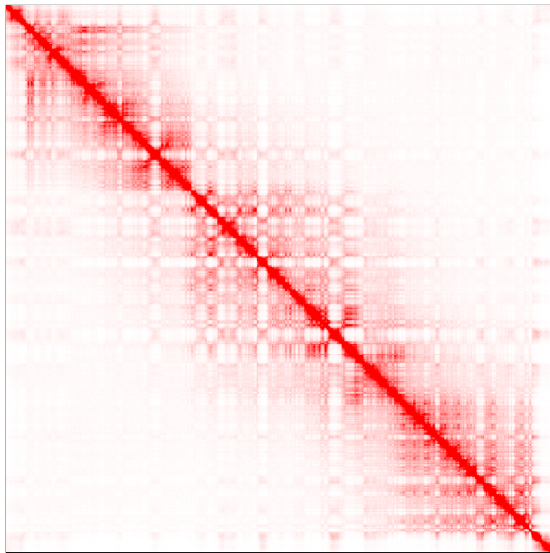


FIGURE 2.1 – Carte Hi-C montrant la fréquence d’interaction entre régions du chromosome 1 (en x et y) dans un foie de porc (Foissac et al. 2019).

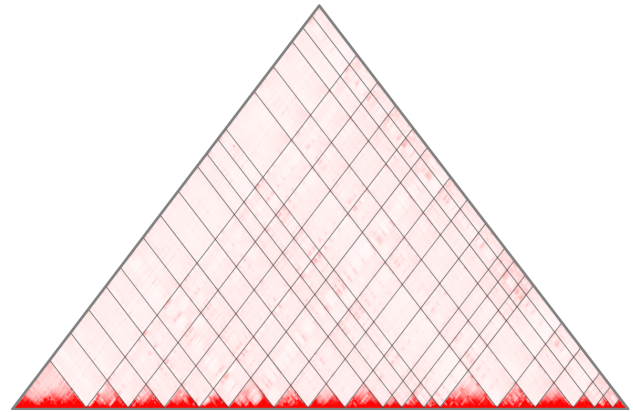


FIGURE 2.2 – Carte Hi-C tout génome réalisée à partir d’échantillons de muscle porcin en développement, avec les 18 autosomes en abscisse. Les triangles plus foncés indiquent la prévalence des interactions intra-chromosomiques (Marti-Marimon et al. 2021).

FICHE SYNTHÉTIQUE

- **Période et lieu** : 2014-présent, INRAE Toulouse
- **Génomes** : *G. gallus* (poule), *S. scrofa* (porc), *C. hircus* (chèvre)
- **Cible** : génome, structure spatiale de la chromatine
- **Données traitées** : Hi-C, capture-Hi-C
- **Contributions** : montage et gestion de projet, animation, analyses de données
- **Publications** : Foissac et al. 2019 ; Marti-Marimon et al. 2021

La problématique de la structure chromatinienne ayant été abordée lors du projet ENCODE humain, c’est assez naturellement que, suite à mon intégration à l’INRAE, j’ai souhaité apporter cette nouvelle dimension à l’annotation des génomes animaux. En particulier, lors du montage du projet pilote FR-AgENCODE en 2014 pour l’annotation de quatre espèces (porc, poule, vache et chèvre), nous avons intégré un volet de caractérisation structurale par Hi-C, bien que le nombre de laboratoires réalisant ce type d’expérience dans le monde se comptait sur les doigts de la main à l’époque. C’est donc dans le cadre de ce projet qu’ont été réalisées les premières expériences Hi-C à l’INRAE, expériences qui se font désormais en routine à la plateforme de séquençage. Ces travaux ont permis la première caractérisation de la structure 3D du génome de ces espèces dans le cadre du consortium international FAANG (Foissac et al. 2019).

Puis, dans le cadre de la première thèse que j’ai co-encadrée à l’INRAE (2016-2019), nous avons entrepris d’apporter une dimension “tout-génome” à la caractérisation d’interactions 3D par des expériences Hi-C sur des échantillons de muscle foetal porcin. Ces travaux ont apporté de nouvelles connaissances sur le rôle de la conformation chromatinienne pendant la maturation, étape du développement embryonnaire porcin qui joue un rôle critique dans la mortalité néonatale des porcelets (Marti-Marimon

et al. 2021). Cette étude a également permis de lancer une dynamique d'animation scientifique autour du thème de la génomique 3D, notamment par le biais d'un réseau de collaborations interdisciplinaire que j'anime depuis plus de 5 ans (réseau [Chrocogen](#)).

Vers l'analyse différentielle de données Hi-C

FICHE SYNTHÉTIQUE

- **Période et lieu** : 2018-présent, INRAE Toulouse
- **Génomes** : *S. scrofa* (porc), *G. gallus* (poule), *H. sapiens* (humain), *M. musculus* (souris)
- **Cible** : génome, structure spatiale de la chromatine
- **Données traitées** : Hi-C, capture-Hi-C
- **Contributions** : développements méthodologiques et logiciels, analyses, management
- **Publications** : Randriamihamison 2021 ; Neuvial et al. 2024 ; Jorge et al. 2025

Afin d'identifier les liens entre structure et fonction du génome, une approche classique consiste à comparer les structures présentes dans des échantillons cellulaires venant de groupes qui diffèrent par un facteur d'intérêt : tissu d'origine, état de santé, condition expérimentale, etc. L'objectif est d'associer à ce facteur une différence significative de conformation de la chromatine, comme on le fait dans le cas des analyses différentielles d'expression génique par exemple. Dans le cas de la génomique 3D, les méthodes d'analyses dépendent du type de structure chromatinienne considérée, qui font intervenir différents niveaux d'échelle. A cette complexité s'ajoute le faible nombre de répliques généralement disponibles à cause du coût élevé des expériences. Pour ces raisons, les méthodes et outils manquent dans le domaine.

Je me suis donc investi ces dernières années dans le développement de méthodes différentielles à divers niveaux d'échelle : au niveau des TADs, des compartiments de la chromatine et de chaque pixel de la matrice d'interaction. Afin de mettre en oeuvre ces développements méthodologiques et de les implémenter dans des solutions logicielles concrètes et applicables sur des données réelles, j'ai mobilisé les compétences de collègues de différents champs disciplinaires (biologie, bioinformatique, informatique, statistique), créé et animé un réseau dédié à l'échelle nationale, et obtenus les financements nécessaires (infrastructure informatique, missions, bourses de thèse).

Les résultats obtenus comprennent des méthodes et logiciels de traitement de données Hi-C développées dans le cadre de thèses ou de stages de Master : la méthode TreeDiff pour la comparaison de matrices sous formes d'arbres hiérarchiques (Randriamihamison 2021 ; Neuvial et al. 2024) et son package R associé (<https://cran.r-project.org/web/packages/treediff/index.html>), la méthode HiCDOC pour la détection de différences de compartiments chromatinien, en cours de publication et disponible sous forme d'un package R Bioconductor (<https://bioconductor.org/packages/release/bioc/html/HiCDOC.html>), et la méthode HiCream (<https://cran.r-project.org/web/packages/hicream/index.html>) qui fait l'objet d'une thèse en cours (Jorge et al. 2025).

2.2 Animation scientifique et responsabilités collectives

PRÉAMBULE

En complément de mes activités de recherche, je me suis toujours investi dans des actions d’animation scientifique, de prise de responsabilités collectives et de formation à la recherche par l’encadrement. Ces engagements poursuivent plusieurs objectifs : entretenir des liens avec des communautés scientifiques variées, assurer une visibilité auprès de partenaires locaux, nationaux et internationaux, ouvrir l’accès à de nouvelles collaborations et sources de financement, et maintenir un haut niveau de connaissance prospective au service des projets collectifs dont j’ai la charge. Les sections suivantes présentent quelques exemples de ces activités, sélectionnés en fonction de leur pertinence vis-à-vis de mon positionnement scientifique et de mon projet.

a. Collectifs scientifiques : contributions et animations

Mes recherches se déploient dans des collectifs variés, allant de projets collaboratifs à des réseaux scientifiques. Ces collectifs constituent autant d’opportunités de structuration, d’échanges et de mises en synergie entre disciplines.

Ainsi, j’ai participé à divers degrés d’implication à plusieurs projets d’envergure (financés à plusieurs centaines de KE). J’ai notamment co-monté et co-piloté le projet SelGen FR-AgENCODE (2015–2018, métaprogramme INRA) où j’étais aussi responsable du WP2 “Bioinformatics & Data Analysis”. Plus récemment, j’ai assuré le rôle de WP leader dans l’ANR Plus4Pigs (2019–2023, WP1 sur les analyses transcriptomiques single-cell), et dans le cadre du projet européen H2020 GENE-SWitCH, je coordonne depuis 2019 la plupart des analyses bioinformatiques du WP2 (RNA-seq, sRNA-seq, ATAC-seq et PC-HiC). Ces responsabilités m’ont permis de renforcer mes compétences en gestion de projets interdisciplinaires, tout en développant des collaborations structurantes pour mes recherches.

En parallèle, j’ai pris l’initiative de fédérer une communauté autour de la génomique 3D en créant en 2019 le groupe de travail CHROCOGEN (Chromatin Conformation and Gene Expression). Ce réseau transdisciplinaire, qui compte aujourd’hui une soixantaine de membres issus d’INRAE, CNRS, INSERM, CEA et universités (Toulouse, Paris, Bordeaux, Lyon), propose régulièrement des séminaires et des journal clubs². Depuis 2023, grâce à un financement du métaprogramme INRAE DIGIT-BIO, j’ai élargi l’animation du réseau par des déplacements sur sites, des bourses de co-encadrement de masters et l’organisation d’un atelier de type “mini-symposium” lors de la conférence de bioinformatique JOBIM 2024. Ces activités renforcent la visibilité nationale de la thématique et favorisent des échanges interinstitutionnels bénéfiques à mes propres recherches.

À l’échelle internationale, j’ai assuré entre 2015 et 2018 l’animation du sous-comité “Chromatin Structure and Chromosome Conformation” au sein du collectif “Bioinformatics & Data Analysis” du consortium FAANG. Cette responsabilité m’a permis d’élargir mon réseau de collaborations au-delà du cadre national, de faciliter les échanges au sein du consortium FAANG et de renforcer ma visibilité dans le domaine.

b. Responsabilités collectives et instances

Au-delà de l’animation scientifique, je considère que contribuer au fonctionnement des instances collectives fait pleinement partie de mes missions de chercheur. Ces engagements, locaux ou nationaux, m’enrichissent en retour en m’apportant une meilleure compréhension du fonctionnement de mon institut et en consolidant mes compétences d’animation et d’évaluation.

2. Programme complet et inscription à Chrocogen : <https://groupes.renater.fr/sympa/info/chrocogen>

Depuis mai 2024, j'assure l'animation du nouveau Pôle "Génomique Structurale et Fonctionnelle" de l'unité GenPhySE, créé dans le cadre de la préparation à l'évaluation Hcéres. Ce Pôle réunit trois équipes (25 agents et 7 non permanents) et vise à dépasser la juxtaposition de thématiques d'équipes pour construire une dynamique commune. À ce titre, je suis missionné par ma Direction d'Unité pour organiser la réflexion prospective, animer les échanges scientifiques, contribuer à la définition de profils de recrutement, favoriser les interactions avec les autres Pôles de l'unité et assurer la rédaction et la présentation du projet lors de l'évaluation Hcéres.

J'ai ainsi lancé des séances de type "journal club", organisé un premier mini-symposium hors site et co-rédigé le projet de Pôle. Ces activités renforcent non seulement la cohésion interne, mais elles me permettent aussi de mieux inscrire mes propres recherches dans une dynamique collective et interdisciplinaire.

À l'échelle nationale, je participe depuis plusieurs années aux instances représentatives et d'évaluation. J'ai siégé comme représentant du personnel en Commission Administrative Paritaire (CAP) des CR INRAE (2016–2022, élu sur deux mandatures), puis en Commission Consultative Paritaire des Contractuels (CCPC) des contractuels depuis 2023 comme mandaté syndical. Ces contributions ont renforcé ma compréhension du fonctionnement de l'institut et de la fonction publique en général, utile pour les activités d'encadrement et de management.

Depuis 2020, je suis également membre de la Commission Scientifique Spécialisée "GVA – Génétique Végétale et Animale", où j'ai d'abord été nommé puis élu. Cette expérience m'apporte un regard nouveau sur la diversité des parcours scientifiques et sur les dynamiques de carrière, enrichissant mes compétences en évaluation scientifique et en construction de projet. De manière ponctuelle, je contribue aussi à des jurys de thèse (examinateur pour S. Espeso Gil, Univ. P. Fabra, Barcelone en 2016, et L. Lima, Univ. Lyon 1 en 2019), des commissions de recrutement et jury professionnel (CRCN-TH INRAE 2019–2020) des jury d'Ecole Doctorale et de Master (Toulouse) et à des Comités de Suivi de thèse, prolongeant ainsi mon implication dans l'évaluation et la formation à la recherche.

Dans l'ensemble, ces responsabilités collectives locales et nationales s'articulent bien avec mes activités de recherche. Elles développent ma capacité à fédérer, à anticiper les évolutions scientifiques et à inscrire mes propres travaux dans une dynamique collaborative au sein de mon collectif.

c. Formation à la recherche par l'encadrement

J'ai co-encadré 5 thèses en lien avec mes thématiques scientifiques principales autour de l'annotation génomique et de la génomique 3D :

- **3D genome conformation and gene expression in fetal pig muscle at late gestation**, de Maria Marti Marimon (2016–9/11/2018), avec Martine Yerle à GenPhySE, INRAE (<https://theses.fr/2018INPT0099>). Situation actuelle : en "CDI" au CRG de Barcelone.
- **Analyse de la méthylation de l'ADN par séquençage haut-débit chez la Poule**, de Marjorie Mersch (2016–30/10/2018), avec Frédérique Pitel à GenPhySE, INRAE (<https://theses.fr/2018INPT0099>). Situation actuelle : inconnue.
- **Classification Ascendante Hiérarchique sous Contrainte de Contiguïté pour l'Analyse de données Hi-C**, de Nathanaël Randriamihamison (2018–27/10/2021), avec Nathalie Vialaneix (INRAE), Marie Chavent (INRIA) et Pierre Neuvial (CNRS) à MIAT, INRAE (<https://theses.fr/2021TOU30108>). Situation actuelle : enseignant en mathématiques (Carcassonne).
- **Analyse comparative de données de génomique 3D**, d'Elise Jorge (2023–présent), avec

Nathalie Vialaneix (INRAE) et Pierre Neuval (CNRS) à GenPhySE, INRAE (<https://theses.fr/s415694>). Cette thèse est la première pour laquelle je suis déclaré comme encadrant principal.

- **Evolution des paysages régulateurs de l'expression des gènes en lien avec les réarrangements génomiques chez les vertébrés** de Victor Lefebvre (2024–présent), avec Anamaria Necselea (CNRS) Sarah Djebali (INSERM) au LBBE de Lyon (<https://theses.fr/s409540>).

J'ai aussi co-encadré plusieurs stages de Master :

- 2015 : M. Mersch (M2 Toulouse) à GenPhySE, INRAE, avec F. Pitel.
- 2017 : M. Smart (M1 Curtin, Australia) à GenPhySE, INRAE, avec S. Djebali et K. Munyard.
- 2017 : R. Taris (M2 Toulouse) à GenPhySE, INRAE, avec C. Noirot et F. Pitel.
- 2018 : C. Mestre (M2 Toulouse) à GenPhySE, INRAE, avec T. Faraut et S. Djebali.
- 2019 : C. Kurylo (M2 Toulouse) à MIAT, INRAE, avec M. Zytnicki.
- 2022 : T. Azevedo (M1 INSA Toulouse) à GenPhySE, INRAE, avec S. Djebali et C. Guyomar.
- 2023 : G. Cardenas (M2 Rennes) à MIAT, INRAE, avec N. Vialaneix.
- 2024 : V. Lefebvre (M2 Lyon) au LBBE, CNRS, avec A. Necselea et S. Djebali.

3) Projet de recherche. Génomique structurale et fonctionnelle : analyses de données omiques pour mieux comprendre le lien génotype-phénotype.

3.1 Contexte et enjeux

Le changement climatique constitue l'un des principaux défis pour l'agriculture du XXI^e siècle. L'augmentation de la fréquence et de l'intensité des crises environnementales, conjuguée à la pression accrue des bioagresseurs, menace la durabilité des filières et met à l'épreuve la compétitivité de l'agriculture française et européenne. Face à ces bouleversements, l'amélioration de la robustesse et de la résilience des espèces agricoles représente une priorité stratégique, tant pour la sécurité alimentaire que pour la transition vers des pratiques agricoles plus durables. Dans ce contexte, l'INRAE joue un rôle central en produisant les connaissances et outils nécessaires à l'adaptation de l'agriculture et à l'accompagnement des politiques publiques.

Un levier essentiel pour relever ces défis réside dans la compréhension des mécanismes par lesquels les espèces agricoles perçoivent les contraintes, mobilisent des processus de régulation à diverses échelles et activent des mécanismes de défense et d'adaptation. Les progrès récents en génomique, transcriptomique et épigénomique permettent désormais d'explorer avec une résolution inédite la diversité génétique et fonctionnelle des espèces d'élevage. La mobilisation conjointe des biotechnologies de séquençage pour la production de données omiques et de la science des données pour l'analyse des systèmes complexes rend possible la caractérisation fine des mécanismes d'adaptation et la valorisation de la diversité des ressources génomiques. Ces avancées ouvrent la voie à l'identification de mécanismes de tolérance et de résistance, ainsi qu'à leur intégration dans les programmes de sélection et de gestion de la biodiversité. C'est en particulier l'objet du Défi Recherche et Innovation "Caractérisation et mobilisation des ressources génétiques" d'INRAE 2030, qui vise à mieux caractériser et valoriser les ressources génétiques existantes, ainsi que du Grand Objectif Scientifique "Comprendre et modéliser la dynamique et la régulation de l'expression des génomes" du Département de Génétique Animale, qui a notamment pour ambition de développer l'annotation fonctionnelle des génomes.

C'est dans ce contexte que s'inscrit mon projet de recherche, dont l'objectif est d'améliorer la compréhension des mécanismes reliant génotype et phénotype chez les espèces animales agricoles. Le projet s'appuie principalement sur mon cœur d'expertise, l'annotation fonctionnelle et la structure 3D de la chromatine, en continuité avec les projets internationaux auxquels je contribue (en particulier le projet H2020 GENE-SWitCH). Sur le plan scientifique, le projet se décline en plusieurs parties, détaillées dans la section suivante.

Au niveau institutionnel, le projet s'inscrit dans le cadre de mon équipe actuelle REGLISS ("Regulatory Genomics for Livestock Species"), en cohérence avec les objectifs du Pôle de Recherche n°5 "Génomique Structurale et Fonctionnelle" de l'unité GenPhySE, pôle dont j'assure la responsabilité depuis 2024. Le départ prochain, dans le cadre d'une mobilité, de l'actuel responsable d'équipe pour-

rait m'amener à en assumer également la responsabilité, en fonction des discussions à suivre au sein du collectif et avec la Direction d'Unité.

En termes de collaborations, le projet mobilise en priorité mes partenaires réguliers. À l'échelle nationale, il s'agit des départements INRAE GA et MathNum (unités GenPhySE et MIAT à Toulouse, PEGASE à Rennes et GABI à Jouy-en-Josas), de l'INSERM (IRD Toulouse) et du CNRS (IMT Toulouse, LBBE Lyon). À l'échelle internationale (essentiellement européenne), mes principaux partenaires sont issus du consortium FAANG, avec notamment le WUR (Pays-Bas), l'Institut Roslin (Écosse), l'EMBL/EBI (Angleterre), le FLI/FBN (Allemagne), l'Université d'Évora (Portugal) et le CRG/Seqera (Catalogne, Espagne). De par son caractère inter- et transdisciplinaire, le projet mobilise des compétences variées et vise un large spectre d'applications sur plusieurs espèces animales.

3.2 Programme scientifique du projet

a. L'annotation génomique pour améliorer la compréhension des génomes animaux

Annotation génomique : exploitation des ressources disponibles

Pour valoriser au mieux les ressources génomiques produites par l'INRAE et les autres acteurs internationaux, il est essentiel de disposer d'une cartographie fiable et précise des séquences génomiques. L'annotation génomique, en identifiant les éléments fonctionnels du génome, fournit les informations nécessaires à la compréhension de sa structure et de son fonctionnement. Elle constitue ainsi un socle pour de nombreuses applications : établir le lien génotype-phénotype, prédire l'impact de perturbations environnementales, anticiper la réponse immunitaire face à un pathogène, ou encore modéliser les interactions hôte-microbiote. Or, les annotations de référence disponibles présentent encore des lacunes importantes, notamment pour les espèces d'intérêt agronomique.

La première étape de mon projet vise donc à améliorer l'annotation génomique des espèces animales agricoles. Je m'appuierai sur les données de l'initiative internationale FAANG, en priorité celles que nous avons produites dans les projets FR-AgENCODE (métaprogramme INRAE) et GENE-SWitCH (H2020), dont le financement est achevé. Dans un second temps, l'objectif est d'intégrer l'ensemble des ressources mondiales disponibles. En continuité avec mes travaux, je ciblerai deux types d'éléments fonctionnels : (i) les gènes, longs ou courts, codants ou non-codants, et (ii) les éléments régulateurs de l'expression génique. Pour cela, nous disposons déjà de données RNA-seq, ATAC-seq, ChIP-seq, WGBS/RRBS, Hi-C et PC-HiC, acquises sur des centaines d'échantillons de porc, poule, vache et chèvre, ainsi que des pipelines et des infrastructures adaptées et éprouvées.

De l'annotation structurale à l'annotation fonctionnelle

Selon l'état d'avancement des jeux de données, il s'agira soit de finaliser l'annotation d'un type particulier (ex. : transcriptome de petits ARNs ou régions accessibles de la chromatine), soit d'intégrer les différentes couches d'annotation pour atteindre un niveau d'information supérieur : associer activateurs et gènes cibles, reconstituer des réseaux de régulation, identifier des cofacteurs au sein de mêmes voies métaboliques. Ces analyses intégratives visent à passer d'une annotation essentiellement structurale à une annotation véritablement fonctionnelle, en révélant les mécanismes complexes qui régulent l'expression génique.

Fort de l'expérience acquise comme WP leader du projet pilote FR-AgENCODE (Foissac et al. 2019), je souhaite étendre cette approche aux données plus riches et variées du projet H2020 GENE-

SWitCH, afin d’approfondir la connaissance du développement embryonnaire chez le porc et la poule.

Un axe original du projet consiste à intégrer la génomique 3D, facteur clé du lien entre structure et fonction. La conformation tridimensionnelle de la chromatine joue en effet un rôle déterminant dans l’activation ou la répression des gènes. Mon expertise combinée en annotation et en génomique 3D chez les animaux d’élevage (Foissac et al. 2019 ; Marti-Marimon et al. 2021 ; Neuvial et al. 2024 ; Jorge et al. 2025 ; Kurylo et al. 2023), ainsi que l’animation d’une communauté interdisciplinaire sur le sujet (réseau Chrocogen), me place en position de proposer et d’implémenter cette intégration innovante.

b. Analyse de données omiques : vers une génomique fonctionnelle comparative

Analyse différentielle de données omiques

Une autre stratégie clé de mes travaux consiste à développer et systématiser les approches omiques différentielles. Comparer différents états d’un système est en effet au cœur de la démarche scientifique, et les données omiques n’y font pas exception. Au-delà de la description des états constitutifs de l’annotation structurale, leur comparaison permet d’identifier les variations significatives, potentiellement associées à des phénotypes d’intérêt. Ainsi, la confrontation d’annotations issues d’échantillons biologiques distincts (tissus, stades de développement, conditions physiologiques, etc.) fournit des informations fonctionnelles précieuses.

Si les méthodes d’analyse différentielle sont aujourd’hui robustes pour la plupart des technologies, des lacunes persistent pour les approches plus récentes telles qu’en Hi-C, single-cell ou séquençage ONT. Dans la continuité de mes travaux, je développerai des méthodologies permettant des analyses différentielles correctes et efficaces, avec une priorité donnée à la génomique 3D. Pour cela, je m’appuierai sur un réseau pluridisciplinaire de collaborateurs à l’interface entre biologie, bioinformatique et mathématiques.

Analyse inter-espèces : vers l’épigénomique comparative

Enfin, le projet intègre une dimension comparative inter-espèces, déjà amorcée dans FR-AgENCODE (Foissac et al. 2019). Ce volet s’intéresse aux aspects évolutifs de la génomique fonctionnelle. La comparaison des profils omiques entre espèces permet d’envisager une véritable épigénomique comparative, révélant la dynamique évolutive des mécanismes de régulation de l’expression génique. Je co-encadre d’ailleurs depuis peu une thèse sur l’évolution de la régulation dans le cadre des gènes dupliqués.

Un enjeu majeur de l’exploitation des données génomiques en élevage est la quantité souvent limitée de données disponibles pour chaque espèce, qui restreint l’application des approches statistiques classiques et des dernières évolutions en Intelligence Artificielle (comme l’apprentissage profond et les grands modèles génératifs). Dans ce contexte, l’approche d’épigénomique comparative que je propose de développer apporte une solution prometteuse. Elle consiste à mobiliser le transfert de connaissances entre espèces, en s’appuyant sur leurs similarités évolutives et fonctionnelles. Cette approche enrichit l’interprétation des données pour les espèces moins étudiées et ouvre la voie à l’identification de mécanismes de tolérance et de résilience potentiellement transférables d’une espèce à l’autre.

3.3 Pilotage de la recherche et animation scientifique

PRÉAMBULE

Cette dernière partie présente la politique de gestion de recherche que je souhaite mettre en œuvre pour mener à bien ce projet scientifique, en tenant compte du contexte actuel et anticipé de mon environnement.

Jusqu'à récemment, mes responsabilités managériales concernaient principalement l'animation scientifique en appui à mon expertise propre (notamment via le réseau Chrocogen). Depuis un an, mon investissement s'est élargi vers l'animation de collectifs de recherche au sein d'INRAE. Ainsi, lors de la restructuration de l'unité GenPhySE en vue de l'évaluation Hcéres, j'ai pris une part active dans la construction du nouveau Pôle scientifique "Génomique structurale et fonctionnelle", que j'anime depuis. Ce Pôle regroupe 3 équipes de recherche et 25 agents titulaires. Dans le même contexte, j'ai également créé l'une des équipes de ce pôle, l'équipe "Génomique régulationnelle pour les espèces d'élevage" (REGLISS) en rassemblant et fédérant autour d'un projet scientifique commun diverses expertises en transcriptomique, épigénétique et bioinformatique auparavant dispersées dans 4 équipes. Pour limiter le cumul de responsabilités, j'ai accompagné la transition de la responsabilité d'équipe vers un collègue et concentré mon action sur l'animation du Pôle.

Cette responsabilité comporte plusieurs enjeux. Il a fallu élaborer un projet commun, dépassant la simple juxtaposition des intérêts des trois équipes ("Organisation tridimensionnelle du génome", "Génomique régulationnelle pour les espèces animales", "Génomique, évolution, environnement et structure de l'abeille"). Pour cela, j'ai organisé une co-construction à travers des ateliers prospectifs sur les projets déposés ou en cours de construction, et sur des sessions thématiques transverses (interactions entre transcription et organisation 3D du génome, IA et génomique, édition de l'épigénome avec CRISPR/Cas9). J'ai ensuite coordonné la rédaction de la partie du rapport Hcéres relative au Pôle. La stratégie définie vise à aborder le lien génotype-phénotype au plus près du génome, à travers des phénotypes dits intermédiaires ou moléculaires, en combinant approches expérimentales et analyses *in silico* à l'échelle de la cellule. À noter que le Pôle ne compte actuellement qu'un seul DR, proche de la retraite, ce qui confère à mon action un rôle déterminant en matière d'animation et de prospective.

L'objectif est désormais d'assurer la cohérence scientifique à long terme, en renforçant les liens inter-équipes et en favorisant l'émergence de projets fédérateurs. Pour cela, je prévois deux types d'animations régulières : des présentations de type "journal club", inspirées de l'expérience menée avec succès dans Chrocogen depuis 7 ans, et des ateliers thématiques d'une demi-journée à une journée organisés une à deux fois par semestre, comme j'ai fait pour construire le projet Hcéres. À ces dispositifs s'ajoutent des actions concrètes pour encourager les projets inter-équipes : intégration croisée de membres dans les comités de suivi de thèse, organisation des répétitions de soutenances de thèse et Master au niveau du Pôle, co-encadrement de stages, ou encore proposition de financements collaboratifs d'appui lors des discussions autour des demandes de poste. À peine initiées, certaines de ces actions portent déjà leurs fruits : un stage M2 co-encadré par REGLISS et GENOME3D débutera en 2026 suite à des échanges inter-équipes lors d'un CSI, et un axe de génomique 3D a été intégré dans un projet PEPR AgroDiv déposé par l'équipe BEEGEES.

En outre, mes activités d'animation sont amenées à se diversifier et à s'étendre au-delà de l'unité. Au niveau national, à la suite de sollicitations lors des conférences JOBIM, j'ai accepté de m'impliquer progressivement dans la Société Française de Bioinformatique (SFBI) : aide ponctuelle aux aspects

administratifs, puis candidature au bureau prévue après 2026, une fois l'animation de Pôle stabilisée. À l'international, j'ai été mandaté par mon Département pour représenter INRAE et promouvoir les intérêts de la Génétique Animale au sein du focus group "Domestic Animals Genome and Phenome" de l'infrastructure européenne ELIXIR, qui réunit plusieurs de mes partenaires FAANG actuels (H2020 GENE-SWitCH, nf-core Special Interest Group "Animal Genomics"). Mon objectif est d'y assurer un positionnement stratégique fort de l'Institut et du Département, et de faciliter les échanges entre ELIXIR et le département GA d'INRAE.

Enfin, je souhaite maintenir un investissement actif dans l'animation scientifique sur mes thématiques d'expertise, en particulier à travers le réseau Chrocogen dédié à la génomique 3D.

4) Liste des productions

FICHE SYNTHÉTIQUE

- **Total publication** : 33 *peer-reviewed* articles, 3 chapitres de livres.
- **Authorship** : 7 articles signés en (co-)premier auteur, 3 en dernier ou *corresponding* auteur (les 3 dans les cinq dernières années).
- **Domaine** : essentiellement biologie moléculaire/cellulaire et bioinformatique.
- **Journaux les plus fréquents** : 5 articles dans *Nature*, 2 dans *Cell*, *Nucleic Acids Research*, *Scientific Reports*, *Frontiers in Genetics*, *PLoS ONE*, 1 dans *Nature Methods*, *Genome Research*, *PLoS Computational Biology*, *Journal of the Royal Statistical Society*, etc.
- **Liens et accès pdf** : <http://genoweb.toulouse.inra.fr/~sfoissac/website/publications.html>.

a. Articles publiés dans des revues internationales à comité de lecture

- Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, Casas E, Cheng HH, Clarke L, Couldrey C et al. (2015). “Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project”. In : *Genome Biology* 16.1, p. 1-6. ISSN : 1474760X. DOI : [10.1186/s13059-015-0622-4](https://doi.org/10.1186/s13059-015-0622-4).
- Bonnet A, Cabau C, Bouchez O, Sarry J, Marsaud N, Foissac S, Woloszyn F, Mulsant P et Mandon-Pepin B (2013). “An overview of gene expression dynamics during early ovarian folliculogenesis : Specificity of follicular compartments and bi-directional dialog”. In : *BMC Genomics* 14.1, p. 1-19. ISSN : 14712164. DOI : [10.1186/1471-2164-14-904](https://doi.org/10.1186/1471-2164-14-904).
- David SA, Mersch M, Foissac S, Collin A, Pitel F et Coustham V (2017). “Genome-wide epigenetic studies in chicken : A review”. In : *Epigenomes* 1.3, p. 20. ISSN : 20754655. DOI : [10.3390/epigenomes1030020](https://doi.org/10.3390/epigenomes1030020).
- Degalez F, Charles M, Foissac S, Zhou H, Guan D, Fang L, Klopp C, Allain C, Lagoutte L, Lecerf F et al. (2024). “Enriched atlas of lncRNA and protein-coding genes for the GRCg7b chicken assembly and its functional annotation across 47 tissues”. In : *Scientific Reports* 14.1, p. 6588. DOI : [10.1038/s41598-024-56705-y](https://doi.org/10.1038/s41598-024-56705-y).
- Denoeud F, Kapranov P, Ucla C, Frankish A, Castelo R, Drenkow J, Lagarde J, Alioto T, Manzano C, Chrast J et al. (2007). “Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions”. In : *Genome Research* 17.6, p. 746-759. ISSN : 10889051. DOI : [10.1101/gr.5660607](https://doi.org/10.1101/gr.5660607).
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F et al. (2012a). “Landscape of transcription in human cells”. In : *Nature* 489.7414, p. 101-108. ISSN : 1476-4687. DOI : [10.1038/nature11233](https://doi.org/10.1038/nature11233).

- Djebali S, Kapranov P, Foissac S, Lagarde J, Reymond A, Ucla C, Wyss C, Drenkow J, Dumais E, Murray RR et al. (2008). “Efficient targeted transcript discovery via array-based normalization of RACE libraries”. In : *Nature Methods* 5.7, p. 629-635. ISSN : 15487091. DOI : [10.1038/nmeth.1216](https://doi.org/10.1038/nmeth.1216).
- Djebali S, Lagarde J, Kapranov P, Lacroix V, Borel C, Mudge JM, Howald C, Foissac S, Ucla C, Chrast J et al. (2012b). “Evidence for transcript networks composed of chimeric rnas in human cells”. In : *PLoS ONE* 7.1, e28213. ISSN : 19326203. DOI : [10.1371/journal.pone.0028213](https://doi.org/10.1371/journal.pone.0028213).
- Dufour A, Kurylo C, Stöckl JB, Laloë D, Bailly Y, Manceau P, Martins F, Turhan AG, Ferchaud S, Pain B et al. (2024). “Cell specification and functional interactions in the pig blastocyst inferred from single-cell transcriptomics and uterine fluids proteomics”. In : *Genomics* 116.2, p. 110780. ISSN : 0888-7543. DOI : [10.1016/j.ygeno.2023.110780](https://doi.org/10.1016/j.ygeno.2023.110780).
- Fejes-Toth K, Sotirova V, Sachidanandam R, Assaf G, Hannon GJ, Kapranov P, Foissac S, Willingham AT, Duttagupta R, Dumais E et al. (2009). “Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs”. In : *Nature* 457.7232, p. 1028-1032. ISSN : 00280836. DOI : [10.1038/nature07759](https://doi.org/10.1038/nature07759).
- Fève K, Foissac S, Pinton A, Mompert F, Esquerré D, Faraut T, Yerle M et Riquet J (2017). “Identification of a t(3;4)(p1.3;q1.5) translocation breakpoint in pigs using somatic cell hybrid mapping and high-resolution mate-pair sequencing”. In : *PLoS ONE* 12.11, p. 1-17. ISSN : 19326203. DOI : [10.1371/journal.pone.0187617](https://doi.org/10.1371/journal.pone.0187617).
- Foissac S, Bardou P, Moisan A, Cros MJ et Schiex T (2003). “EUGÈNE’HOM : A generic similarity-based gene finder using multiple homologous sequences”. In : *Nucleic Acids Research* 31.13, p. 3742-3745. ISSN : 03051048. DOI : [10.1093/nar/gkg586](https://doi.org/10.1093/nar/gkg586).
- Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, Esquerré D, Zytnicki M, Derrien T, Bardou P et al. (2019a). “Multi-species annotation of transcriptome and chromatin structure in domesticated animals”. In : *BMC Biology* 17.1, p. 1-25. ISSN : 17417007. DOI : [10.1186/s12915-019-0726-5](https://doi.org/10.1186/s12915-019-0726-5).
- Foissac S, Gouzy J, Rombauts S, Mathe C, Amselem J, Sterck L, Peer Y de, Rouze P et Schiex T (2008). “Genome Annotation in Plants and Fungi : EuGene as a Model Platform”. In : *Current Bioinformatics* 3.2, p. 87-97. ISSN : 15748936. DOI : [10.2174/157489308784340702](https://doi.org/10.2174/157489308784340702).
- Foissac S et Sammeth M (2007). “ASTALAVISTA : dynamic and flexible analysis of alternative splicing events in custom gene datasets”. In : *Nucleic acids research* 35.suppl_2, W297-W299. DOI : [10.1093/nar/gkm311](https://doi.org/10.1093/nar/gkm311).
- Foissac S et Schiex T (2005). “Integrating alternative splicing detection into gene prediction”. In : *BMC Bioinformatics* 6.1, p. 1-10. ISSN : 14712105. DOI : [10.1186/1471-2105-6-25](https://doi.org/10.1186/1471-2105-6-25).
- Giuffra E, Tuggle CK et the FAANG Consortium (2019). “Functional Annotation of Animal Genomes (FAANG) : Current Achievements and Roadmap”. In : *Annual Review of Animal Biosciences* 7, p. 65-88. ISSN : 21658110. DOI : [10.1146/annurev-animal-020518-114913](https://doi.org/10.1146/annurev-animal-020518-114913).
- Hoellinger T, Mestre C, Aschard H, Le Goff W, Foissac S, Faraut T et Djebali S (2023). “Enhancer/gene relationships : need for more reliable genome-wide reference sets”. In : *Frontiers in Bioinformatics* 3, p. 1092853. DOI : [10.1093/nar/gkx920](https://doi.org/10.1093/nar/gkx920).
- Jehl F, Degalez F, Bernard M, Lecerf F, Lagoutte L, Désert C, Coulée M, Bouchez O, Leroux S, Abasht B et al. (2021). “RNA-Seq Data for Reliable SNP Detection and Genotype Calling : Interest for Coding Variant Characterization and Cis-Regulation Analysis by Allele-Specific Expression in Livestock Species”. In : *Frontiers in Genetics* 12, p. 1104. ISSN : 16648021. DOI : [10.3389/fgene.2021.655707](https://doi.org/10.3389/fgene.2021.655707).

- Jehl F, Muret K, Bernard M, Boutin M, Lagoutte L, Désert C, Dehais P, Esquerré D, Acloque H, Giuffra E et al. (2020). “An integrative atlas of chicken long non-coding genes and their annotations across 25 tissues”. In : *Scientific reports* 10.1, p. 20457. DOI : [10.1038/s41598-020-77586-x](https://doi.org/10.1038/s41598-020-77586-x).
- Jorge É, Foissac S, Neuviat P, Zytynicki M et Vialaneix N (2025). “A comprehensive review and benchmark of differential analysis tools for Hi-C data”. In : *Briefings in Bioinformatics* 26.2, bbaf074. ISSN : 1477-4054. DOI : [10.1093/bib/bbaf074](https://doi.org/10.1093/bib/bbaf074).
- Kapranov P, Ozsolak F, Kim SW, Foissac S, Lipson D, Hart C, Roels S, Borel C, Antonarakis SE, Monaghan AP et al. (2010a). “New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism”. In : *Nature* 466.7306, p. 642-646. ISSN : 14764687. DOI : [10.1038/nature09190](https://doi.org/10.1038/nature09190).
- Karami K, Sabban J, Cerutti C, Devailly G, Foissac S, Gourichon D, Hubert A, Hubert JN, Leroux S, Zerjal T et al. (2025). “Molecular responses of chicken embryos to maternal heat stress through DNA methylation and gene expression : a pilot study”. In : *Environmental Epigenetics* 11.1. ISSN : 2058-5888. DOI : [10.1093/eep/dvaf009](https://doi.org/10.1093/eep/dvaf009).
- Kurylo C, Guyomar C, Foissac S et Djebali S (2023a). “TAGADA : a scalable pipeline to improve genome annotations with RNA-seq data”. In : *NAR Genomics And Bioinformatics* 5.4, lqad089. ISSN : 2631-9268. DOI : [10.1093/nargab/lqad089](https://doi.org/10.1093/nargab/lqad089).
- Marti-Marimon M, Vialaneix N, Lahbib-Mansais Y, Zytynicki M, Camut S, Robelin D, Yerle-Bouissou M et Foissac S (2021). “Major Reorganization of Chromosome Conformation During Muscle Development in Pig”. In : *Frontiers in Genetics* 12, p. 1895. ISSN : 16648021. DOI : [10.3389/fgene.2021.748239](https://doi.org/10.3389/fgene.2021.748239).
- Muret K, Klopp C, Wucher V, Esquerré D, Legeai F, Lecerf F, Désert C, Boutin M, Jehl F, Acloque H et al. (2017). “Long noncoding RNA repertoire in chicken liver and adipose tissue”. In : *Genetics Selection Evolution* 49.1, p. 1-17. ISSN : 12979686. DOI : [10.1186/s12711-016-0275-0](https://doi.org/10.1186/s12711-016-0275-0).
- Neuviat P, Randriamihison N, Chavent M, Foissac S et Vialaneix N (2024). “A two-sample tree-based test for hierarchically organized genomic signals”. In : *Journal of the Royal Statistical Society Series C : Applied Statistics*, qlae011. DOI : [10.1093/jrsssc/qlae011](https://doi.org/10.1093/jrsssc/qlae011).
- Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B et Milos PM (2010). “Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation”. In : *Cell* 143.6, p. 1018-1029. ISSN : 00928674. DOI : [10.1016/j.cell.2010.11.020](https://doi.org/10.1016/j.cell.2010.11.020).
- Piqué M, López JM, Foissac S, Guigó R et Méndez R (2008). “A Combinatorial Code for CPE-Mediated Translational Control”. In : *Cell* 132.3, p. 434-448. ISSN : 00928674. DOI : [10.1016/j.cell.2007.12.038](https://doi.org/10.1016/j.cell.2007.12.038).
- Rubio-Peña K, Fontrodona L, Aristizábal-Corrales D, Torres S, Cornes E, García-Rodríguez FJ, Serrat X, González-Knowles D, Foissac S, Porta-De-La-Riva M et al. (2015). “Modeling of autosomal-dominant retinitis pigmentosa in *Caenorhabditis elegans* uncovers a nexus between global impaired functioning of certain splicing factors and cell type-specific apoptosis”. In : *RNA* 21.12, p. 2119-2131. ISSN : 14699001. DOI : [10.1261/rna.053397.115](https://doi.org/10.1261/rna.053397.115).
- Sammeth M, Foissac S et Guigó R (2008). “A general definition and nomenclature for alternative splicing events”. In : *PLoS Computational Biology* 4.8, e1000147. ISSN : 1553734X. DOI : [10.1371/journal.pcbi.1000147](https://doi.org/10.1371/journal.pcbi.1000147).
- The ENCODE Project Consortium (2007). “Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project”. In : *Nature* 447.7146, p. 799-816. ISSN : 0028-0836. DOI : [10.1038/nature0587](https://doi.org/10.1038/nature0587).

The ENCODE Project Consortium (2012). “An integrated encyclopedia of DNA elements in the human genome”. In : *Nature* 489.7414, p. 57-74. ISSN : 14764687. DOI : [10.1038/nature11247](https://doi.org/10.1038/nature11247).

b. Chapitres d’ouvrage

Djebali S, Wucher V, Foissac S, Hitte C, Corre E et Derrien T (2017). “Bioinformatics pipeline for transcriptome sequencing analysis”. In : *Methods in Molecular Biology*. T. 1468. Springer Protocols, p. 201-219. DOI : [10.1007/978-1-4939-4035-6_14](https://doi.org/10.1007/978-1-4939-4035-6_14).

Foissac S et Sammeth M (2015). “Analysis of alternative splicing events in custom gene datasets by AStalavista”. In : *Methods in Molecular Biology*. T. 1269. Springer Protocols, p. 379-392. DOI : [10.1007/978-1-4939-2291-8_24](https://doi.org/10.1007/978-1-4939-2291-8_24).

Neuvial P, Foissac S et Vialaneix N (2023). “Comprendre l’organisation spatiale de l’ADN à l’aide de la statistique”. In : *L’Interdisciplinarité*. T. 1. CNRS Editions, p. 172-179.

c. Communications dans des conférences

Quelques exemples de communications orales et/ou poster.

Acloque H, Harrison P, Lakhal W, Martin F, Archibald A, Beinat M, Davey M, Djebali S, Foissac S, Guizard S et al. (2022). “Extensive functional genomics information from early developmental time points for pig and chicken (Talk)”. In : *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*. Wageningen Academic Publishers, p. 2281-2284. DOI : [10.3920/978-90-8686-940-4_550](https://doi.org/10.3920/978-90-8686-940-4_550).

Degalez F et al. (2023). “A lncRNA gene-enriched atlas for GRCg7b chicken genome and its functional annotation across 47 tissues (Poster)”. In : *ISAG*. Cape Town, South Africa.

Djebali S, Foissac S, Vialaneix N, Munyard K, Rau A, Faraut T, Lagarrigue S, Acloque H et Giuffra E (2019). “Chromatin accessibility conservation across four livestock species (Talk)”. In : *Inproceedings of the International Society for Animal Genetics (ISAG 2019)* (7-12 juill. 2019). Llieda, Spain.

Foissac S (2019). “Functional annotation of livestock genomes : chromatin structure and regulation of gene expression (Invited talk)”. In : *Journal of Animal Science (Proceedings of the ASAS/ASDS Midwest Joint Meeting)*. T. 97. Suppl. 2. Omaha, NE, USA, p. 15-16. DOI : [10.1093/jas/skz122.028](https://doi.org/10.1093/jas/skz122.028).

— (2022). “TAGADA : Transcripts and Genes Assembly, Deconvolution, Analysis ; a Nextflow pipeline to improve genome annotations with RNA-seq data (Invited talk)”. In : PRBB-CRG. Barcelona, Catalunya.

Foissac S, Djebali S, Vialaneix N, Zytnicki M, Rau A, Lagarrigue S, Acloque H et Giuffra E (2019b). “Multi-level conservation of chromosome conformation across livestock species reveals evolutionary links between genome structure and function (Talk)”. In : *Inproceedings of the International Society for Animal Genetics (ISAG 2019)* (7-12 juill. 2019). Llieda, Spain.

Kurylo C, Maigné E, Foissac S et Zytnicki M (2023b). “Prediction and differential analysis of chromatin compartments from Hi-C data (Poster)”. In : *ISMB/ECCB*. Lyon, France.

Lahbib-Mansais Y, Marti-Marimon M, Vialaneix N, Foissac S, Bouissou-Matet M et Liaubet L (2019). “Organisation nucléaire et expression génique lors du développement chez le porc”. In : *Séminaire du réseau EpiPHASE* (26-27 juin 2019). Castanet-Tolosan, France.

Randriamihamison N, Chavent M, Foissac S, Vialaneix N et Neuvial P (2020). “Classification ascendante hiérarchique sous contrainte de contigüité pour l’analyse différentielle de données Hi-C”. In : *Journées de Statistique de la SFdS (volume exceptionnel)*.

d. Jeux de données

Je travaille le plus possible dans un esprit de science ouverte, dans le cadre de projets qui produisent des données de type FAIR. Pour les données omiques, le principe est de rendre les séquences publiques au plus tôt après séquençage, avec comme seule restriction d’usage de ne pas publier avant le groupe producteur des données.

Par exemple, j’ai contribué à la production de centaines d’annotations génomiques dans le cadre du consortium international FAANG et du projet H2020 GENE-SWitCH en particulier : identification de gènes, transcrits (longs, courts, codants ou non) et de régions actives de la chromatine par analyses de données RNA-seq, sRNA-seq, ATAC-seq et capture Hi-C chez le porc et la poule. Ces résultats sont déposés dans des registres publics structurés, riches en métadonnées, accessibles sur le portail du *FAANG Data Center* à l’EMBL/EBI :

- *Gallus gallus*
 - GENE-SWitCH chicken promoter capture Hi-C data analysis (21 fichiers déposés, ENA ID : PRJEB53986)
 - GENE-SWitCH chicken gene expression profiling by RNA-seq (89 fichiers déposés, ENA ID : PRJEB42025)
 - GENE-SWitCH Chicken transcriptome and gene expression atlas (smallRNA-seq) (420 fichiers déposés, ENA ID : PRJEB42041)
 - GENE-SWitCH Chicken chromatin accessibility profiling by ATAC-seq (106 fichiers déposés, ENA ID : PRJEB45945)
- *Sus scrofa*
 - GENE-SWitCH pig gene expression profiling by RNA-seq (89 fichiers déposés, ENA ID : PRJEB41970)
 - GENE-SWitCH Pig chromatin accessibility profiling by ATAC-seq (106 fichiers déposés, ENA ID : PRJEB44468)
 - GENE-SWitCH Pig chromatin accessibility profiling by ATAC-seq (333 fichiers déposés, ENA ID : PRJEB53440)
 - GENE-SWitCH Pig transcriptome and gene expression atlas (smallRNA-seq) (420 fichiers déposés, ENA ID : PRJEB42001)

Pour plus d’information sur la politique FAIR de données du consortium FAANG : <https://www.faang.org/data-share-principle> et <https://data.faang.org/home>.

Pour le partage et la diffusion de données et résultats, j’utilise également des ressources nationales au besoin, comme par exemple le repository INRAE Omics Dataverse, hébergé par recherche.data.gouv.fr :

- Companion dataverse pour “Major reorganization of chromosome conformation during muscle development in pig” : <https://doi.org/10.15454/DOMEHB>
- Companion dataverse pour “TAGADA : a scalable pipeline to improve genome annotations with RNA-seq data” : <https://doi.org/10.57745/3UGLXW>

Ces dépôts “dataverse” diffusent en libre accès (FAIR) des informations liées à un projet et/ou une publication : données de type *Supplementary Files* ou ne pouvant être hébergées par l’éditeur, lien vers toutes les données relatives au projet (séquences, métadata, résultats), preprint éventuel de l’article, scripts non disponibles ailleurs.

Eventuellement, d'autres informations peuvent être mises à disposition par d'autres canaux, comme par exemple par le site fragencode <https://www.fragencode.org/>.

e. Logiciels

En terme de développement logiciel, mes contributions ont considérablement évolué au cours de ma carrière. Pendant ma thèse il y a vingt ans, j'ai beaucoup contribué au code du logiciel EuGène (<http://eugene.toulouse.inra.fr/>). Depuis, mes contributions relèvent de deux formes : soit le modelage artisanal de scripts “à l'ancienne” parfois monolignes et difficilement diffusables pour les analyses exploratoires (bash, awk, perl), soit en accompagnant ou supervisant le développement de logiciels codés par des collègues. Mis à part mon passage dans le privé, où j'étais devenu responsable de plusieurs produits logiciels commerciaux, je ne me suis entièrement consacré à des productions *open-source*. En particulier, depuis mon recrutement à l'INRAE, j'ai contribué au développements de plusieurs logiciels d'analyse de données omiques :

- TAGADA : un logiciel d'analyse de données RNA-seq pour l'annotation génomique sous forme d'un pipeline open source encapsulé (Kurylo et al. 2023). <https://github.com/FAANG/analysis-TAGADA#readme>
- HiCDOC : un package Bioconductor (R) pour l'analyse de données de génomique 3D de type Hi-C visant à identifier des différences significatives de compartimentation chromatinienne entre conditions biologiques d'intérêt. <https://bioconductor.org/packages/release/bioc/html/HiCDOC.html>
- TREEDIFF : un package R du CRAN pour l'analyse de données de génomique 3D de type Hi-C visant à détecter des différences significatives entre groupes de dendrogrammes associés à des conditions biologiques d'intérêt (Neuvial et al. 2024). <https://cran.r-project.org/web/packages/treediff/index.html>
- HiCREAM : un package R du CRAN pour l'analyse différentielle entre groupes de matrices d'interaction Hi-C avec garanties statistiques sur les résultats du test. <https://cran.r-project.org/web/packages/hicream/index.html>

Bibliographie

- Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, Casas E, Cheng HH, Clarke L, Couldrey C et al. (2015). “Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project”. In : *Genome Biology* 16.1, p. 1-6. ISSN : 1474760X. DOI : [10.1186/s13059-015-0622-4](https://doi.org/10.1186/s13059-015-0622-4).
- Dali R et Blanchette M (2017). “A critical assessment of topologically associating domain prediction tools”. In : *Nucleic Acids Research* 45.6, p. 2994-3005. ISSN : 1362-4962. DOI : [10.1093/nar/gkx145](https://doi.org/10.1093/nar/gkx145).
- Davies JOJ, Oudelaar AM, Higgs DR et Hughes JR (2017). “How best to identify chromosomal interactions : a comparison of approaches”. In : *Nature Methods* 14.2, p. 125-134. ISSN : 1548-7105. DOI : [10.1038/nmeth.4146](https://doi.org/10.1038/nmeth.4146).
- Degalez F, Charles M, Foissac S, Zhou H, Guan D, Fang L, Klopp C, Allain C, Lagoutte L, Lecerf F et al. (2024). “Enriched atlas of lncRNA and protein-coding genes for the GRCg7b chicken assembly and its functional annotation across 47 tissues”. In : *Scientific Reports* 14.1, p. 6588. DOI : [10.1038/s41598-024-56705-y](https://doi.org/10.1038/s41598-024-56705-y).
- Denoeud F, Kapranov P, Ucla C, Frankish A, Castelo R, Drenkow J, Lagarde J, Alioto T, Manzano C, Chrast J et al. (2007). “Prominent use of distal 5’ transcription start sites and discovery of a large number of additional exons in ENCODE regions”. In : *Genome Research* 17.6, p. 746-759. ISSN : 10889051. DOI : [10.1101/gr.5660607](https://doi.org/10.1101/gr.5660607).
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F et al. (2012a). “Landscape of transcription in human cells”. In : *Nature* 489.7414, p. 101-108. ISSN : 1476-4687. DOI : [10.1038/nature11233](https://doi.org/10.1038/nature11233).
- Djebali S, Kapranov P, Foissac S, Lagarde J, Reymond A, Ucla C, Wyss C, Drenkow J, Dumais E, Murray RR et al. (2008). “Efficient targeted transcript discovery via array-based normalization of RACE libraries”. In : *Nature Methods* 5.7, p. 629-635. ISSN : 15487091. DOI : [10.1038/nmeth.1216](https://doi.org/10.1038/nmeth.1216).
- Djebali S, Lagarde J, Kapranov P, Lacroix V, Borel C, Mudge JM, Howald C, Foissac S, Ucla C, Chrast J et al. (2012b). “Evidence for transcript networks composed of chimeric rnas in human cells”. In : *PLoS ONE* 7.1, e28213. ISSN : 19326203. DOI : [10.1371/journal.pone.0028213](https://doi.org/10.1371/journal.pone.0028213).
- Fejes-Toth K, Sotirova V, Sachidanandam R, Assaf G, Hannon GJ, Kapranov P, Foissac S, Willingham AT, Duttagupta R, Dumais E et al. (2009). “Post-transcriptional processing generates a diversity of 5’-modified long and short RNAs”. In : *Nature* 457.7232, p. 1028-1032. ISSN : 00280836. DOI : [10.1038/nature07759](https://doi.org/10.1038/nature07759).
- Foissac S (2004). “Localisation de gènes et variants par intégration d’informations”. Thèse de doct. Toulouse, France : Université Paul Sabatier Toulouse III.
- Foissac S, Bardou P, Moisan A, Cros MJ et Schiex T (2003). “EUGÈNE’HOM : A generic similarity-based gene finder using multiple homologous sequences”. In : *Nucleic Acids Research* 31.13, p. 3742-3745. ISSN : 03051048. DOI : [10.1093/nar/gkg586](https://doi.org/10.1093/nar/gkg586).
- Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, Esquerré D, Zytynicki M, Derrien T, Bardou P et al. (2019). “Multi-species annotation of transcriptome and chromatin structure in domesticated animals”. In : *BMC Biology* 17.1, p. 1-25. ISSN : 17417007. DOI : [10.1186/s12915-019-0726-5](https://doi.org/10.1186/s12915-019-0726-5).
- Foissac S, Gouzy J, Rombauts S, Mathe C, Amselem J, Sterck L, Peer Y de, Rouze P et Schiex T (2008). “Genome Annotation in Plants and Fungi : EuGene as a Model Platform”. In : *Current Bioinformatics* 3.2, p. 87-97. ISSN : 15748936. DOI : [10.2174/157489308784340702](https://doi.org/10.2174/157489308784340702).
- Foissac S et Sammeth M (2007). “ASTALAVISTA : dynamic and flexible analysis of alternative splicing events in custom gene datasets”. In : *Nucleic acids research* 35.suppl_2, W297-W299. DOI : [10.1093/nar/gkm311](https://doi.org/10.1093/nar/gkm311).

- Foissac S et Sammeth M (2015). “Analysis of alternative splicing events in custom gene datasets by AStalavista”. In : *Methods in Molecular Biology*. T. 1269. Springer Protocols, p. 379-392. DOI : [10.1007/978-1-4939-2291-8_24](https://doi.org/10.1007/978-1-4939-2291-8_24).
- Foissac S et Schiex T (2005). “Integrating alternative splicing detection into gene prediction”. In : *BMC Bioinformatics* 6.1, p. 1-10. ISSN : 14712105. DOI : [10.1186/1471-2105-6-25](https://doi.org/10.1186/1471-2105-6-25).
- Giuffra E, Tuggle CK et the FAANG Consortium (2019). “Functional Annotation of Animal Genomes (FAANG) : Current Achievements and Roadmap”. In : *Annual Review of Animal Biosciences* 7, p. 65-88. ISSN : 21658110. DOI : [10.1146/annurev-animal-020518-114913](https://doi.org/10.1146/annurev-animal-020518-114913).
- Jehl F, Degalez F, Bernard M, Lecerf F, Lagoutte L, Désert C, Coulée M, Bouchez O, Leroux S, Abasht B et al. (2021). “RNA-Seq Data for Reliable SNP Detection and Genotype Calling : Interest for Coding Variant Characterization and Cis-Regulation Analysis by Allele-Specific Expression in Livestock Species”. In : *Frontiers in Genetics* 12, p. 1104. ISSN : 16648021. DOI : [10.3389/fgene.2021.655707](https://doi.org/10.3389/fgene.2021.655707).
- Jehl F, Muret K, Bernard M, Boutin M, Lagoutte L, Désert C, Dehais P, Esquerré D, Acloque H, Giuffra E et al. (2020). “An integrative atlas of chicken long non-coding genes and their annotations across 25 tissues”. In : *Scientific reports* 10.1, p. 20457. DOI : [10.1038/s41598-020-77586-x](https://doi.org/10.1038/s41598-020-77586-x).
- Jorge É, Foissac S, Neuvial P, Zytnecki M et Vialaneix N (2025). “A comprehensive review and benchmark of differential analysis tools for Hi-C data”. In : *Briefings in Bioinformatics* 26.2, bbaf074. ISSN : 1477-4054. DOI : [10.1093/bib/bbaf074](https://doi.org/10.1093/bib/bbaf074).
- Kapranov P, Ozsolak F, Kim SW, Foissac S, Lipson D, Hart C, Roels S, Borel C, Antonarakis SE, Monaghan AP et al. (2010). “New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism”. In : *Nature* 466.7306, p. 642-646. ISSN : 14764687. DOI : [10.1038/nature09190](https://doi.org/10.1038/nature09190).
- Karami K, Sabban J, Cerutti C, Devailly G, Foissac S, Gourichon D, Hubert A, Hubert JN, Leroux S, Zerjal T et al. (2025). “Molecular responses of chicken embryos to maternal heat stress through DNA methylation and gene expression : a pilot study”. In : *Environmental Epigenetics* 11.1. ISSN : 2058-5888. DOI : [10.1093/eep/dvaf009](https://doi.org/10.1093/eep/dvaf009).
- Kempfer R et Pombo A (2019). “Methods for mapping 3D chromosome architecture”. In : *Nature Reviews Genetics* 21.4, p. 207-226. ISSN : 1471-0064. DOI : [10.1038/s41576-019-0195-2](https://doi.org/10.1038/s41576-019-0195-2).
- Kurylo C, Guyomar C, Foissac S et Djebali S (2023). “TAGADA : a scalable pipeline to improve genome annotations with RNA-seq data”. In : *NAR Genomics And Bioinformatics* 5.4, lqad089. ISSN : 2631-9268. DOI : [10.1093/nargab/lqad089](https://doi.org/10.1093/nargab/lqad089).
- Liu L, Han K, Sun H, Han L, Gao D, Xi Q, Zhang L et Lin H (2023). “A comprehensive review of bioinformatics tools for chromatin loop calling”. In : *Briefings in Bioinformatics* 24.2. ISSN : 1477-4054. DOI : [10.1093/bib/bbad072](https://doi.org/10.1093/bib/bbad072).
- Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R et al. (2015). “Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions”. In : *Cell* 161.5, p. 1012-1025. ISSN : 0092-8674. DOI : [10.1016/j.cell.2015.04.004](https://doi.org/10.1016/j.cell.2015.04.004).
- Marti-Marimon M, Vialaneix N, Lahbib-Mansais Y, Zytnecki M, Camut S, Robelin D, Yerle-Bouissou M et Foissac S (2021). “Major Reorganization of Chromosome Conformation During Muscle Development in Pig”. In : *Frontiers in Genetics* 12, p. 1895. ISSN : 16648021. DOI : [10.3389/fgene.2021.748239](https://doi.org/10.3389/fgene.2021.748239).
- Muret K, Klopp C, Wucher V, Esquerré D, Legeai F, Lecerf F, Désert C, Boutin M, Jehl F, Acloque H et al. (2017). “Long noncoding RNA repertoire in chicken liver and adipose tissue”. In : *Genetics Selection Evolution* 49.1, p. 1-17. ISSN : 12979686. DOI : [10.1186/s12711-016-0275-0](https://doi.org/10.1186/s12711-016-0275-0).
- Neuvial P, Randriamihamison N, Chavent M, Foissac S et Vialaneix N (2024). “A two-sample tree-based test for hierarchically organized genomic signals”. In : *Journal of the Royal Statistical Society Series C : Applied Statistics*, qlae011. DOI : [10.1093/jrssc/qlae011](https://doi.org/10.1093/jrssc/qlae011).
- Oudelaar AM et Higgs DR (2020). “The relationship between genome structure and function”. In : *Nature Reviews Genetics* 22.3, p. 154-168. ISSN : 1471-0064. DOI : [10.1038/s41576-020-00303-x](https://doi.org/10.1038/s41576-020-00303-x).
- Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B et Milos PM (2010). “Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation”. In : *Cell* 143.6, p. 1018-1029. ISSN : 00928674. DOI : [10.1016/j.cell.2010.11.020](https://doi.org/10.1016/j.cell.2010.11.020).

- Piqué M, López JM, Foissac S, Guigó R et Méndez R (2008). “A Combinatorial Code for CPE-Mediated Translational Control”. In : *Cell* 132.3, p. 434-448. ISSN : 00928674. DOI : [10.1016/j.cell.2007.12.038](https://doi.org/10.1016/j.cell.2007.12.038).
- Randriamihamison N (2021). “Classification Ascendante Hiérarchique sous Contrainte de Contiguïté pour l’Analyse de données Hi-C”. Thèse de doct.
- Sammeth M, Foissac S et Guigó R (2008). “A general definition and nomenclature for alternative splicing events”. In : *PLoS Computational Biology* 4.8, e1000147. ISSN : 1553734X. DOI : [10.1371/journal.pcbi.1000147](https://doi.org/10.1371/journal.pcbi.1000147).
- The ENCODE Project Consortium (2007). “Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project”. In : *Nature* 447.7146, p. 799-816. ISSN : 0028-0836. DOI : [10.1038/nature0587](https://doi.org/10.1038/nature0587).
- (2012). “An integrated encyclopedia of DNA elements in the human genome”. In : *Nature* 489.7414, p. 57-74. ISSN : 14764687. DOI : [10.1038/nature11247](https://doi.org/10.1038/nature11247).
- Zufferey M, Tavernari D, Oricchio E et Ciriello G (2018). “Comparison of computational methods for the identification of topologically associating domains”. In : *Genome Biology* 19.1. ISSN : 1474-760X. DOI : [10.1186/s13059-018-1596-9](https://doi.org/10.1186/s13059-018-1596-9).