



**HAL**  
open science

# Les petits ARN non codants du spermatozoïde bovin : de potentiels biomarqueurs de la fertilité mâle ?

Eli Sellem

## ► To cite this version:

Eli Sellem. Les petits ARN non codants du spermatozoïde bovin : de potentiels biomarqueurs de la fertilité mâle?. Biologie de la reproduction. Université Paris-Saclay, 2021. Français. NNT : 2021UPASB008 . tel-03223078

**HAL Id: tel-03223078**

**<https://pastel.hal.science/tel-03223078>**

Submitted on 10 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les petits ARN non codants du  
spermatozoïde bovin : de potentiels  
biomarqueurs de la fertilité mâle ?

*The bovine sperm small non-coding RNAs:  
Potential biomarkers of male fertility?*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n°581 Agriculture, Alimentation, Biologie,  
Environnement et Santé (ABIES)

Spécialité de doctorat : Biologie de la reproduction

Unité de recherche : Université Paris-Saclay, UVSQ, INRAE, BREED,  
78350, Jouy-en-Josas, France.

Référent : AgroParisTech

**Thèse présentée et soutenue à Paris-Saclay,  
le 15/03/2021, par**

**Eliaou SELLEM**

**Composition du Jury**

**Patrice HUMBLLOT**

Professeur SLU Univ. Uppsala (Suède)

Président

**Romain BARRES**

Professeur UCPH Univ. Copenhague (Danemark)

Rapporteur & Examineur

**Valérie GRANDJEAN**

Chargée de Recherche INSERM, Provence-Alpes-Côtes d'Azur

Rapporteur & Examinatrice

**Sébastien BANZET**

Professeur agrégé, IRBA

Examineur

**Direction de la thèse**

**Hélène JAMMES**

Directrice de Recherche, INRAE IdF-Jouy-en-Josas-Antony

Directrice de thèse

**Laurent SCHIBLER**

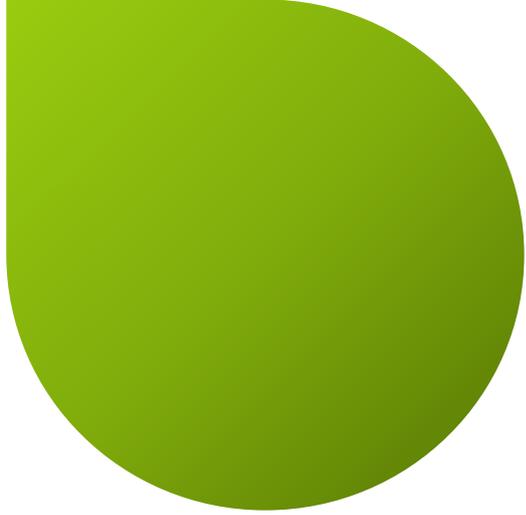
Responsable Développement & Innovation, ALLICE

Co-Directeur de thèse

<b>Remerciements.....</b>	<b>6</b>
<b>Liste des abréviations .....</b>	<b>11</b>
<b>Liste des illustrations .....</b>	<b>17</b>
1) Tables et figures du manuscrit : .....	18
2) Tables et figures du premier article : .....	19
3) Tables et figures du second article : .....	20
<b>Prologue.....</b>	<b>21</b>
<b>Introduction : contexte socio-économique et scientifique .....</b>	<b>23</b>
<b>Chapitre I : Le contexte de l'IA en France .....</b>	<b>24</b>
1) Importance de la maîtrise de la fonction sexuelle du taureau .....	29
1.1) La success-story du spermatozoïde : de sa conception à la fécondation .....	29
1.1.1) La spermatogenèse .....	29
a) Durant la vie fœtale.....	30
b) De la vie périnatale à péri-pubertaire .....	31
1.1.2) La maturation épидидymaire .....	35
a) <i>Les différents rôles de l'épididyme</i> .....	36
b) L'épididyme, un lieu d'échange avec le spermatozoïde .....	39
1.1.3) Le plasma séminal .....	41
1.1.4) Dans les voies femelles et les premières étapes de la fécondation.....	43
1.2) De la production à la mesure de la fertilité : comment sont utilisés les taureaux et leurs gamètes en France ?.....	47
1.2.1) Du testage sur descendance à la sélection génomique .....	47
1.2.2) La production de semence en centre de collecte .....	48
1.2.3) Les indicateurs de la fertilité mâle .....	51
2) La prédiction de la fertilité mâle .....	53
2.1) Les paramètres historiques .....	55
2.2) Les paramètres « cellulaires » via l'essor de nouvelles technologies .....	56
2.3) Les nouveaux paramètres moléculaires.....	58
<b>Chapitre II Épигénétique et fertilité mâle .....</b>	<b>59</b>
1) Des modifications post-traductionnelles des histones...aux protamines .....	60
1.1) Rôles et fonctionnement biologique des histones.....	60
1.2) Des histones aux protamines dans le spermatozoïde.....	61
1.3) Impact sur la fertilité .....	62
2) La méthylation de l'ADN.....	63

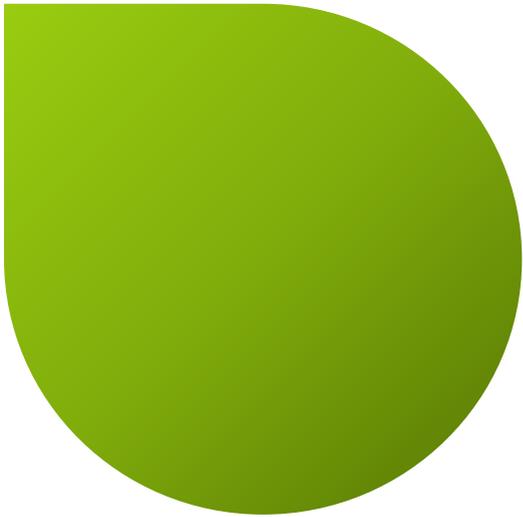
2.1) Rôle et fonctionnement biologique de la méthylation de l'ADN .....	63
2.2) Impact sur la fertilité .....	64
3) Les petits ARN non codants.....	65
3.1) Les microARN (miRNA).....	66
3.1.1) Biogenèse et régulation de l'expression des miRNA.....	67
3.1.2) Mécanismes d'action et rôle des miRNA .....	70
3.2) Les piRNA.....	72
3.2.1) Biogenèse des piRNA.....	72
3.2.2) Rôles et mécanismes d'action.....	74
3.3) Des ARN de transfert aux tRFs .....	75
3.3.1) Biogenèse des tRFs.....	75
3.3.2) Rôles et mécanismes d'action.....	76
3.4) Des ARN ribosomiques aux rsRNA .....	76
3.4.1) Biogenèse des rsRNA.....	76
3.4.2) Rôles et mécanismes d'action.....	77
3.5) Les autres familles de petits ARN non codants .....	78
4) Les petits ARN non codants au cours de la spermatogenèse .....	78
4.1) Rôle des miRNA .....	79
4.2) Rôle des piRNA .....	83
5) Les petits ARN non codants et maturation de la semence .....	84
5.1) Le contenu en sncRNA varie au cours de la maturation .....	85
5.2) Origine des sncRNA du spermatozoïde épидидymaire.....	86
5.3) Rôles potentiels des sncRNA dans la maturation de la semence .....	88
6) Les sncRNA du spermatozoïde support de la transmission entre générations.....	89
6.1) Rôles potentiels des sncRNA dans le développement embryonnaire .....	90
6.2) Rôles potentiels des sncRNA dans la transmission intergénérationnelle .....	92
<b>Chapitre III : Objectifs de la thèse .....</b>	<b>94</b>
1) Objectif 1 : catalogue exhaustif du contenu en sncRNA du spermatozoïde bovin .....	94
2) Objectif 2 : L'origine des sncRNA.....	95
3) Objectif 3 : biomarqueurs de la fertilité mâle .....	95
<b>Résultats .....</b>	<b>96</b>
<b>Chapitre I : le catalogue en sncRNA du spermatozoïde bovin éjaculé .....</b>	<b>97</b>
1) Éléments de contexte.....	97
2) Résumé de l'étude.....	97
3) Article publié dans <i>Epigenetics &amp; Chromatin</i> en 2020 .....	99

<b>Chapitre II : L'origine des sncRNA du spermatozoïde bovin éjaculé .....</b>	<b>152</b>
1) Éléments de contexte.....	152
2) Résumé de l'étude.....	152
3) Article en cours de soumission dans <i>Epigenetics &amp; Chromatin</i> (janvier 2021).....	155
<b>Chapitre III : Les sncRNA comme biomarqueur de la fertilité mâle ?.....</b>	<b>196</b>
1) Éléments de contexte.....	196
2) Plan expérimental.....	197
2.1) Constitution de la bio-banque d'éjaculats.....	197
2.2) Analyse des paramètres fonctionnels de la semence.....	198
2.3) Extraction d'ARN total et séquençage des sncRNA.....	198
2.4) Les différentes approches statistiques.....	199
3) Résultats.....	200
3.1) Fertilité et paramètres spermatiques des éjaculats utilisés.....	200
3.2) Données obtenues par séquençage.....	202
3.3) sncRNA biomarqueurs de fertilité mâle.....	203
3.4) Prédiction de fertilité en race Montbéliarde.....	209
3.5) Relation entre sncRNA et la qualité de la semence.....	210
4) Discussion.....	212
4.1) La construction du dispositif expérimental.....	212
4.2) Les sncRNA biomarqueurs de qualité de semence.....	213
4.3) La prédiction de la fertilité mâle par l'expression des sncRNA spermatique ?.....	218
<b>Discussion globale et perspectives .....</b>	<b>220</b>
1) Les sncRNA sont-ils de bons biomarqueurs de la fertilité mâle ?.....	221
2) Perspectives.....	225
2.1) Impact sur le développement embryonnaire.....	225
2.2) La prédiction de fertilité : Quel outil terrain ?.....	227
2.3) Exploitation des vésicules extracellulaires ?.....	230
2.4) Le spermatozoïde peut-il être une sentinelle ?.....	231
<b>Références bibliographiques .....</b>	<b>234</b>
<b>Annexes.....</b>	<b>248</b>
1) Figures supplémentaires du premier article.....	249
<b>Communications.....</b>	<b>258</b>
1) Publications hors « article(s) de thèse ».....	259
2) Présentations orales lors de congrès internationaux.....	259



*« Tout est changement, tout évolue, tout est en devenir,  
non pour ne plus être, mais pour devenir ce qui n'est pas  
encore »*

Epictète (50-125 après J.C.)



# Remerciements

*« ...Ta thèse concrétisera tout le travail que tu mènes depuis toutes ces années...c'est un des rares moments où tu pourras te recentrer sur ton travail, revenir vers des expériences passées et finalement prendre suffisamment de recul pour envisager le meilleur chemin à prendre pour la suite... »*

Tu avais raison [Hélène](#), la thèse est une démarche très personnelle, qui m'a demandé une introspection importante... Mais je me suis très vite rendu compte que c'est une aventure que l'on doit vivre à plusieurs. Sans l'équipe qui m'a entouré, que ce soit celle d'ALLICE ou celle de l'INRAE, j'aurais eu beaucoup plus de mal à atteindre mes objectifs. Je tenais donc naturellement à tous vous remercier.

Tout d'abord, merci à [Valérie Grandjean](#) et [Romain Barres](#) qui ont accepté d'être les rapporteurs de la thèse. Merci également à [Sébastien Banzet](#) et [Patrice Humblot](#) d'avoir accepté d'être examinateur dans ce jury. Je me rends compte de la charge supplémentaire de travail que cela représente. J'ai hâte de pouvoir discuter avec vous tous des perspectives de ces travaux. [Patrice](#), j'en profite pour te dire que ta présence ici est très symbolique pour moi. Merci !

Ensuite, je tenais à remercier tout particulièrement deux personnes qui ont véritablement enrichi cette aventure et j'espère bien plus encore : [Brigitte Boizet](#) et [Joël Drevet](#). Nos discussions durant les comités de suivi ont été très importantes pour moi ! Merci.

Merci également à ABIES et AgroParisTech, sans qui tout cela n'aurait pas été possible. Merci [Alexandre Pery](#) et [Pierre Larraufie](#) pour les échanges et les conseils que vous m'avez donnés.

Je tenais également à remercier [Corinne Cotinot](#), [Pascale Chavatte-Palmer](#) et [Olivier Sandra](#) pour leur soutien et la confiance qu'ils m'ont témoignée dans cette longue démarche !

Au démarrage de la thèse, j'avais comme velléité d'acquérir suffisamment de compétences en bio-informatique pour être en mesure d'analyser de façon autonome les quantités importantes de données omiques que l'on génère aujourd'hui dans les labos. Je me suis vite rendu compte qu'il me faudrait revoir à la baisse cette ambition et, à défaut d'être bio-informaticien, devenir au moins un utilisateur averti ! Après trois ans, je pense avoir atteint juste le stade « utilisateur » ... En revanche, je suis sûr d'avoir beaucoup usé les nerfs de [Sylvain Marthey](#) ! [Sylvain](#), sans toi j'en serai encore à essayer de comprendre comment accéder à Genotoul ! Merci d'avoir passé tant d'heures à traiter nos données et à comprendre nos

problématiques. Notre relation prouve s'il en était besoin que seule la collaboration permet d'avancer efficacement et durablement. Un proverbe africain résume exactement ce constat : « *Si tu veux aller vite, marche seul, mais si tu veux aller loin, marchons ensemble* ». Merci [Sylvain](#) !

Deux autres personnes ont été importantes pour le projet. Elles ont même réussi à augmenter mon intérêt pour les approches statistiques. Merci beaucoup pour le temps que vous avez patiemment passé, [Andrea Rau](#) et [Luc Jouneau](#), à m'expliquer et réexpliquer le principe de DeSeq2 ou de CART !

Je tiens tout particulièrement à remercier mes proches collègues de l'équipe MECP<sup>2</sup> et autres, qui pendant 5 ans ont enduré mes blagues incroyables et mes fabuleux jeux de mots ! Merci [Mélanie Jouin](#), [Charline Pontlevoy](#), [Anne Gabory](#), [Aurélie Chaulot-Talmon](#), [Jean-Philippe Perrier](#), [Audrey Prezelin](#), [Valentin Costes](#) et [Hélène Kiefer](#) de m'avoir demandé quasiment à chaque fois une explication ! Merci de m'avoir composé mon « compte rond – ma chanson sur Hotel California », de m'avoir si bien intégré dans l'équipe, de m'avoir fait découvrir le célèbre beurre salé caramel mou d'[Anne](#) ou encore le lieu de stockage des poubelles CMR ! Albert Einstein expliquait sa théorie de la relativité avec cette citation « *Placez votre main sur un poêle une minute et ça vous semble durer une heure. Asseyez-vous auprès d'une jolie fille une heure et ça vous semble durer une minute. C'est ça la relativité* ». J'oserai presque à dire que ces 5 ans sont passés à la vitesse de la lumière ! J'espère que l'on continuera à échanger et à rigoler ensemble pendant encore longtemps !

La fin de la thèse coïncide également avec mon 20<sup>ième</sup> anniversaire au sein d'ALLICE. Je suis tombé amoureux des idées véhiculées par cette fédération et par le monde agricole. En vingt ans, j'ai également eu le temps de nouer des relations très fortes avec mes collègues, tout aussi dinosaures que moi ! Je tiens particulièrement à remercier [Maurice Barbezant](#), [Xavier David](#) et [Pascale Parisot](#), pour la confiance qu'ils m'ont témoigné toutes ces années au sein d'ALLICE. [Olivier Gérard](#), merci de m'avoir transmis ta sagesse et fait profiter de ton expérience. J'ai énormément appris à tes côtés. J'aime bien l'une des citations d'un des maîtres fondateurs de l'Aïkido, Morihei Ueshiba, qui résume ce que parfois tu me disais : « *Les progrès viennent à ceux qui s'entraînent encore et encore, vous reposer sur des techniques secrètes ne vous conduira nulle part* ». Comme je n'avais aucune technique secrète, je n'ai pas eu le choix ! Tu restes aujourd'hui un des piliers dans ma vie professionnelle. Merci également

à Laure Menard, Stephane Devillers, Mathieu Diribarne, Pascal Salvetti, Aurélie Bonnet, Marie-Christine Deloche, Julie Gatien, Daniel Le Bourhis, Zoé Mortimore, Cécile Quetglas, Sebastien Fritz, Chris Hoze et tous mes autres collègues d’Alice, pour votre aide précieuse dans l’élaboration de ces travaux ou de ce document. Vous ne vous en rendez peut-être pas compte, mais toutes les conversations que nous avons eues, ont façonné mon esprit. C’est également une part de vous qui est retranscrit dans ce document. Je citerai bien Henry Ford pour illustrer notre équipe, et promis rien à voir avec la mustang : « *Se réunir est un début ; rester ensemble est un progrès ; travailler ensemble est la réussite* ».

Chrystelle Le Danvic, tu as le droit aussi à une citation ! Elle vient du Dalai Lama et résume très bien les innombrables discussions que nous avons sur l’utilité de notre travail : « *Si vous pensez que vous êtes trop petit pour changer quelque chose, essayez donc de dormir avec un moustique dans votre chambre et vous verrez lequel des deux empêche l’autre de dormir* ». Merci pour les centaines d’heures passées ensemble à refaire le monde agricole, à me faire rire avec tes expressions et j’espère que tu continueras encore longtemps à m’étonner avec tes idées, ta disponibilité, ta ténacité et surtout tes goûts pour la musique Allemande !

Hélène Jammes, tu m’as accueilli il y a maintenant un peu plus de 5 ans. Après 14 ans passé à étudier la fonctionnalité de la semence, je pensais en avoir fait le tour. J’avais décidément bien tort ! L’environnement que tu m’as proposé m’a énormément stimulé et enrichi. Tu as même réussi à me faire aimer la recherche fondamentale ! Tu sais comme j’adore les devises Shaddock, voici celle que j’ai expérimentalement validée la première année : « *En essayant continuellement on finit par réussir. Donc plus ça rate, plus on a de chance que ça marche.* » Merci d’avoir été présente pour moi, merci d’avoir acceptée d’être ma Directrice de thèse et j’espère qu’on se fera encore de très nombreux Apéro-Falcon !

Laurent Schibler, ça fait trois ans que je cherche une citation pour toi ! Je l’ai trouvé, on la doit à Albert Einstein : « *La théorie, c’est quand on sait tout et que rien ne fonctionne. La pratique, c’est quand tout fonctionne et que personne ne sait pourquoi. Ici, nous avons réuni théorie et pratique : Rien ne fonctionne... et personne ne sait pourquoi !* ». Je ne sais pas comment tu fais, mais si Albert Einstein me le permet, j’agrémenterai bien sa phrase, en ajoutant : « *...et personne ne sait pourquoi...enfin avant l’arrivée de Laurent !* ». Je te dois beaucoup pour la réalisation de cette thèse. Tu as été « une lumière au bout du tunnel » et j’espère qu’à force

de travailler à tes côtés, au moins l'une des tes qualités, comme ton esprit de synthèse, ta pertinence ou ta pugnacité, a déteint sur moi !



# Liste des abréviations

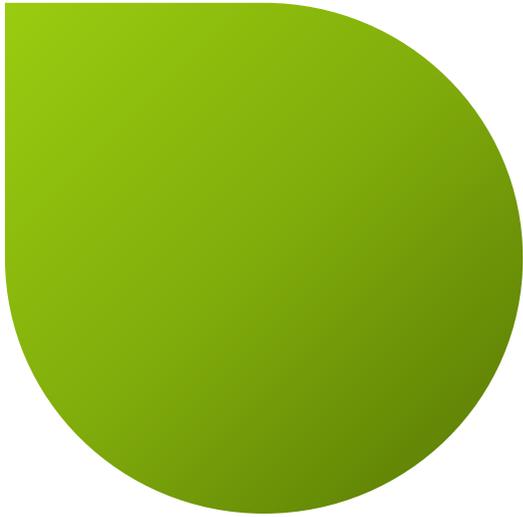
ACP/C	Système cellulaire de dégradation protéique
ADAM	A Disintegrin And Metalloproteinase
ADAR	Adenosine Deaminase RNA Specific
ADN ou DNA	Acide Désoxyribonucléique
AGO	Argonaute RISC Component
AKAP	A-Kinase Anchoring Protein
AKT	AKT Serine/Threonine Kinase
ALH	Amplitude Lateral Head $\mu\text{m}$
AMH	Anti-Müllerian Hormone
AMPc	Adénosine monophosphate cyclique
ANR	Agence Nationale pour la Recherche
APOBEC	Apolipoprotein B mRNA Editing Enzyme Catalytic
ARN, RNA ou RNAs	Acide Ribonucléique
ARNm, mRNA ou mRNAs	ARN messenger
ARNr, rRNA ou rRNAs	ARN ribosomique
ARNt, tRNA ou tRNAs	ARN de transfert
ATP	Adénosine tri phosphate
AWE	Association Wallonne pour l'Elevage
BAZ2A	Bromodomain Adjacent to Zinc Finger Domain 2A
BCA	Between Class analysis
BDR	Biologie du Développement et de la Reproduction
BMPR1A	Bone Morphogenetic Protein Receptor Type 1A
BosTau9	9e version de l'assemblage du génome Bos Taurus
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase
BREED	Biologie de la Reproduction, Environnement, Epigénétique, et Développement
BSP	Bovine Seminal Plasma
CAF1	deadenylase
CaMKII	Calcium/Calmodulin Dependent Protein Kinase II
CASA	Computer aided sperm analysis
CCT2	Chaperonin Containing TCP1 Subunit 2
CDH1	Cadherin 1
CENP-A	Variant centromérique d'Histone H3
cGMP-PKG	KEGG pathway: cGMP-dependent protein kinase
CIA	Centre d'insémination animale
CMH	Complexe majeur d'histocompatibilité
CpG	Cytosine et Guanine consécutives
C-ROS	Tyrosine kinase receptor
DAZL	Deleted In Azoospermia Like
D-Box	Domaine des PIWI protéines
DDX4	DEAD-Box Helicase 4
DGCR8	DiGeorge Syndrome Critical Region Gene 8
DHEA	DeHydroEpiAndrostérone
DNMT	DNA Methyltransferase
DROSHA	Drosha Ribonuclease
EIF4EBP1	Eukaryotic Translation Initiation Factor 4E Binding Protein 1
ELSPBP1	Epididymal sperm binding protein 1

ENA	European Nucleotide Archive
ENO1	Enolase 1
ENVA	Ecole Nationale Vétérinaire d'Alfort
ESR1	Estrogen Receptor 1
ETV3	ETS Variant Transcription Factor 3
FGF9	Fibroblast Growth Factor 9
FIV	Fécondation in vitro
FN1	Fibronectin 1
FOLH1	Folate Hydrolase 1
FSH	Follicle Stimulating Hormone
GABI	Génétique Animale et Biologie Intégrative
GASZ	Cofacteur de l'endonucléase ZUC (piRNA)
GATA4	GATA Binding Protein 4
GBB2	guanine nucleotide-binding protein G(1)/G(S)/G(T) subunit beta-2
Glu	Glutamine
Gly	Glycine
GNAS	GNAS Complex Locus
GnRH	Gonadotropin Releasing Hormone
GPX5	Glutathione Peroxidase 5
GtRNAdb	genomic tRNA database ( <a href="http://gtrnadb.ucsc.edu/">http://gtrnadb.ucsc.edu/</a> )
H2A	Histone 2A
HCO3	Hydrogénocarbonate
HDL	High Density Lipoprotein
HIP1	Huntingtin Interacting Protein 1
HIWI	PIWI protéines
HNF4A	Hepatocyte Nuclear Factor 4 Alpha
IA ou AI	Insémination animale
ICM	Inner cell mass
ICSI	Intra cytoplasmique sperm injection
IGF1	Insuline growth factor
IGF2R	Insulin Like Growth Factor 2 Receptor
IGV	Integrative Genomics viewer
IL-10	Interleukine 10
IMV	IMV Technologies
INRA ou INRAE	Institut national de recherche pour l'agriculture, l'alimentation et l'environnement
IPS1	ARN non codant chez les plantes IPS1
IZUMO1 ou JUNO	Izumo Sperm-Egg Fusion 1
KDa	Kilo dalton
KDM2B	Lysine Demethylase 2B
KEGG	Kyoto Encyclopedia of genes and genomes
KO	Knock-Out
K-RAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
LabCom	Laboratoire Commun
LFNG	Lunatic Fringe Homolog (Drosophila)
LH	Luteinizing Hormone
LHX9	LIM Homeobox 9

LIN28	Lin-28 Homolog A
LINE1	Elément transposable
LNCR	Laboratoire National du Contrôle des Reproducteurs
LSU	Large subunit ; partie du ribosome
LTF	Lactotransferrin
LTR	Elément transposable
LUMA	Luminometric Methylation assay
MAEL	Maelstrom Spermatogenic Transposon Silencer
MAPK1	Mitogen-Activated Protein Kinase 1
MECP	Mécanismes épigénétiques et construction - prédiction des phénotypes
MED1	Mediator Complex Subunit 1
MERVL	Element transposable
MEST	Mesoderm Specific Transcript
MID	Domaine des PIWI protéines
MILI	PIWI protéines
MINO	Cofacteur de l'endonucléase ZUC (piRNA)
MIWI	PIWI protéines
MIWI2	PIWI protéines
MT	Mitochondria
mTOR	Mechanistic Target of Rapamycin Kinase
MYH	Moyenne d'intensité de fluorescence sur l'axe des Y pour la population Hmmp
NANOG	Nanog Homeobox
NANOS2	Nanos C2HC-Type Zinc Finger
NGS	Next generation sequencing
OGP	Oviduct-specific glycoprotein
OTX2	Orthodenticle Homeobox 2
PAPD4	nucléotidyl transferases
PaterLeg	Nom programme ANR (Paternal Legacy)
PAZ	Domaine des PIWI protéines
PBS	Phosphate Buffer Saline
PCR	Polymerase chain reaction
PDPK1	3-Phosphoinositide Dependent Protein Kinase 1
PEG	Paternally Expressed
PGC ou PGCs	Primordial germ cells
PIK3R2	Phosphoinositide-3-Kinase Regulatory Subunit 2
PIWI	P-element induced wimpsy testis
PRCE	Appel à projet ANR contenant une entreprise privée comme partenaire
PRDM1	PR Domain Containing 1, With ZNF Domain
PRDX5	Peroxiredoxin 5
PRM1	Protamine 1
PSGEN	Équipe Populations, Statistique, et Génome
PSMA	Prostate-specific membrane antigen
RFAM	Rfam database ( <a href="http://rfam.xfam.org/">http://rfam.xfam.org/</a> )
RISC	RNA-induced silencing complex
RLT	Tampon de lyse cellulaire
RMN	Résonance magnétique nucléaire

RNAi	Acide Ribonucléique interférence
RNA-seq	Séquençage des ARN
RNF8	Ring Finger Protein 8
ROS	Reactive oxygen species
RPS6KA3	Ribosomal Protein S6 Kinase A3
RRBS	Reduced representation bisulfite sequencing
RT	Room Temperature
RT-qPCR	Reverse Transcriptase quantitative Polymerase chain reaction
SED1 ou MFGE8	Milk Fat Globule EGF And Factor V/VIII Domain Containing
SeQuaMol	Semence Qualité Moléculaire
SERPINA1	Serpin Family A Member 1
SF1	Splicing Factor 1
SINE	Elément transposable
SLERT	Long ARN non codant
SMAD	SMAD Family
SMARCA5	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin, Subfamily A, Member 5
sncRNA	small non coding RNA, petits ARN non codants
SNP ou SNPs	Single Nucleotide Polymorphisme
SOX2	SRY-Box Transcription Factor 2
SPOCD1	POC Domain Containing 1
SRP	Signal Recognition Particle
SRY	Sex Determining Region Y
SSU	SamII subunit ; partie du ribosome
STAR	Steroidogenic Acute Regulatory Protein
STAT3	Signal Transducer and Activator of Transcription 3
STK11	Serine/Threonine Kinase 11
STRA8	Stimulated By Retinoic Acid 8
TAF5	TATA-Box Binding Protein Associated Factor 5
TANC2	Tetratricopeptide Repeat, Ankyrin Repeat and Coiled-Coil Containing 2
TET	Tet Methylcytosine Dioxygenase
TGFβ	Transforming Growth Factor Beta
TGM4	Transglutaminase 4
TH2A	Variant d'histones spécifique des gemètes mâles
TH2B	Variant d'histones spécifique des gemètes mâles
TIMP-2	TIMP Metallopeptidase Inhibitor 2
TNP2	Transition Protein 2
TNR (28, 56, 90 ou 282)	Taux de non-retour (en chaleur) à 28, 56, 90 ou 282 jours post IA
TRBP	TARBP2 Subunit of RISC Loading Complex
TSS	Transcription start site
TTS	Transcription termination site
UCP2	Uncoupling Protein 2
UCSC	University of California Santa Cruz
ULK1	Unc-51 Like Autophagy Activating Kinase 1
UMI	Unique moleculare identifiers
UMR	Unité mixte de recherche

UNCEIA	Union Nationale des Coopératives agricoles d'Elevage et d'Insémination Animale
VCL	Curvilinear velocity $\mu\text{m/s}$
VRET	Cofacteur de l'endonucléase ZUC (piRNA)
WAGO	Protéine de la famille des Argonautes
WGBS	Whole Genome Bisulfite Sequencing
WT1	WT1 Transcription Factor
ZUC	Endonucléase
miRNA, miR, miRs ou isomiR	microARN
piRNA, piRNAs ou isopiR	Piwi ARN
rsRNA, rRF ou rRFs	Fragments d'ARNr
tsRNA, tRF ou tRFs	Fragments d'ARNt



## Liste des illustrations

## 1) Tables et figures du manuscrit :

Figure 1 : Dates clés de l'insémination animale. ©Alice .....	24
Figure 2 : Lazzaro Spallanzani.....	25
Figure 3 : Doses d'inséminations animale.....	25
Figure 4 : Première page du compte-rendu de l'Assemblée Générale Constitutive de l'UNCEIA. ....	27
Figure 5 : Alice, une fédération atypique. ....	28
Figure 6 : La spermatogenèse. ....	34
Figure 7 : Représentation des différentes parties de l'épididyme © Alice .....	36
Figure 8 : Différentes anomalies morphologiques du spermatozoïde bovin éjaculé. ....	37
Figure 9 : Gouttelette cytoplasmique du spermatozoïde bovin éjaculé.....	38
Figure 10 : Impact de l'extinction de différents gènes sur l'environnement luminal de l'épididyme de souris <sup>30</sup> .....	40
Figure 11 : Rôle des ROS dans l'induction de la capacitation et hyperactivation du spermatozoïde <sup>52</sup> . 45	45
Figure 12 : De la récolte à la l'insémination. ....	49
Figure 13 : Évolution technologique du contrôle qualité de la semence.....	50
Figure 14 : Le résultat de l'IA est multifactoriel. ....	52
Figure 15 : Prédiction de la fertilité du taureau avec les données de fonctionnalité de la semence. ...	55
Figure 16 : Analyse de la motilité de la semence. ....	56
Figure 17 : Analyse de la motilité des spermatozoïdes avec la technologie « CASA ».....	57
Figure 18 : Analyse de la fonctionnalité des spermatozoïdes avec le cytomètre de flux. ....	57
Figure 19 : Les principaux acteurs des mécanismes épigénétiques.....	60
Figure 20 : Nombre de publications contenant l'occurrence « miRNA ».....	66
Figure 21 : Exemple d'organisation génomique en « cluster » de certains miRNA positionnés sur le chromosome bovin 21.....	68
Figure 22 : Les miRNA de la forme native à la forme mature. ....	68
Figure 23 : Les miRNA, de leur biogenèse à leur fonction. ....	69
Figure 24 : Les différentes éditions d'un miRNA canonique : les isomiR. ....	70
Figure 25 : Perturbation important du système pileux chez la souris mutante du gène Dicer <sup>131</sup> .....	71
Figure 26 : Biogénèse des piRNA <sup>136</sup> . ....	73
Figure 27 : Biogenèse des fragments d'ARNt.....	75
Figure 28 : Biogenèse des fragments d'ARNr <sup>157</sup> . ....	77
Figure 29 : Diversité des sncRNA dans les cellules germinales mâles murines.....	79
Figure 30 : Liste de miRNA impliqués dans la spermatogenèse chez la souris (figure adaptée <sup>17</sup> ).....	80
Figure 31 : Anomalies morphologiques du spermatozoïde murin délété du gène Dicer <sup>168</sup> . ....	81
Figure 32 : Perturbation de la spermatogenèse chez les souris KO en miR-34bc et miR-449a <sup>76</sup> .....	81
Figure 33 : Différence significative d'expression de certains miRNA en fonction de la section de l'épididyme <sup>209</sup> . ....	87
Figure 34 : Retard du développement embryonnaire murins KO pour le gène Dicer <sup>226</sup> .....	91
Figure 35 : Localisation des centres de production de semence. ....	197
Figure 36 : Groupes de fertilité taureaux Holstein et Montbéliard. ....	201
Figure 37 : Fonctionnalité des éjaculats utilisés.....	201
Figure 38 : Diversité des séquences contenues dans le spermatozoïde éjaculé. ....	202
Figure 39 : Diversité des sncRNA selon leur niveau d'expression. ....	203
Figure 40 : BCA des groupes de fertilité des taureaux Holstein et Montbéliard. ....	204
Table 1 : sncRNA différenciellement exprimés selon le niveau de fertilité.....	205
Figure 41 : Séquences différenciellement les plus exprimées entre les groupes de fertilité.....	205

Table 2 : Détail des tsRNA et rsRNA différemment exprimés selon le niveau de fertilité. ....	206
Figure 42 : Processus biologiques significativement enrichis par les gènes ciblés du bta-miR-146. ..	207
Table 3 : Origine tissulaire des sncRNA différemment exprimés selon le niveau de fertilité. ....	208
Figure 43 : Matrices de confusions du méta-classifieur.....	210
Figure 44 : Caractéristiques des tsRNA et rsRNA corrélés avec MYH. ....	211
Figure 45 : Voie de signalisation du processus biologique mTOR.....	212

## 2) Tables et figures du premier article :

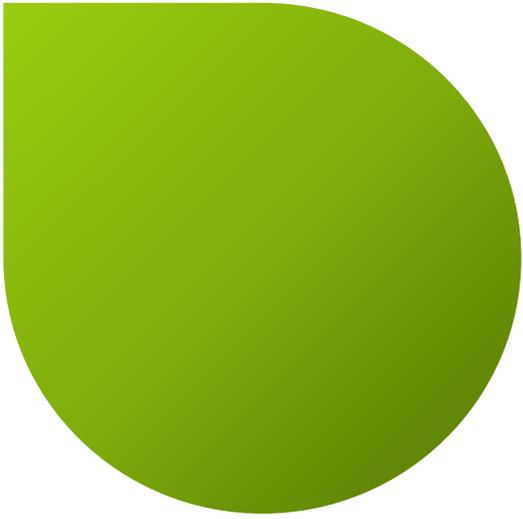
### A comprehensive overview of bull sperm-borne small non-coding RNAs and their diversity across breeds

Table 1. Mapping statistics across the 40 libraries.....	103
Figure 1. sRNA content of bull sperm.....	104
Figure 2. Bull sperm rRNA content.....	105
Figure 3. Bull sperm piRNA content.....	107
Figure 4. Bull sperm tsRNA content.....	108
Figure 5. Bull sperm miRNA content.....	111
Figure 6. Circos plot depicting correlations between miRNAs.....	113
Table 2. Genomic clusters of miRNAs having strong expression correlations.....	114
Figure 7. Correlation networks of positively correlated miRNAs.....	117
Figure 8. Correlation networks of negatively correlated miRNAs.....	118
Figure 9. The diversity of isomiRs in bull sperm.....	120
Figure 10. Frequency of end variations affecting isomiR sequences.....	120
Figure 11. Frequency of substitutions affecting isomiR sequences.....	122
Figure 12. Differential expression of sncRNAs across breeds.....	123
Figure 13: Hierarchical clustering based on miRNA expression.....	125
Figure 14. Correlation networks of positively correlated miRNAs.....	126
Figure 15: Distribution of motility related sperm quality parameters across breeds.....	127
Figure 13. NGS data analysis workflow at a glance.....	140

3) Tables et figures du second article :

Dynamics of cattle sperm sncRNAs during maturation, from testis to ejaculated sperm.

Table 1. Mapping statistics across the 17 libraries.....	162
Figure 1. sncRNA signature discriminates between sperm sampled along the male tract.....	163
Figure 2. Changes in sncRNA expression along the male reproductive tract.....	164
Table 2. Expression ratio for downregulated piRNAs between parenchyma and ejaculated sperm.....	165
Table 3. Abundance and diversity of sperm piRNA along the male tract.....	165
Figure 3. Nucleotide frequency at each position for all detected piRNA in each sampling region.....	166
Table 4. Number of sperm sncRNAs differentially expressed between the six sampling regions.....	167
Table 5. Fold changes SPZ versus PAR and pvalues of five differentially expressed miRNAs.....	168
Figure 4. Proportion of sncRNAs families among differentially expressed sncRNA between testis parenchyma and ejaculated sperm.....	169
Figure 5. Dynamics of tsRNA across regions .....	170
Figure 6. Dynamics of rRNA fragments across regions .....	170
Figure 7. K-means clustering illustrates the diversity of expression profiles among sncRNAs.....	171
Figure 8 . Expression fold changes observed between PAR and CAP, as well as CAP and COR for sncRNA belonging to downregulated K-means clusters .....	172
Figure 9. General characteristics of K-means clusters.....	174
Figure 10. sncRNA expression across K-means clusters .....	177
Figure 11. GO biological processes associated with K-means clusters.....	178
Figure 12. Differentially expressed miRNAs are shared between bovine epididymosomes and epididymal sperm .....	183



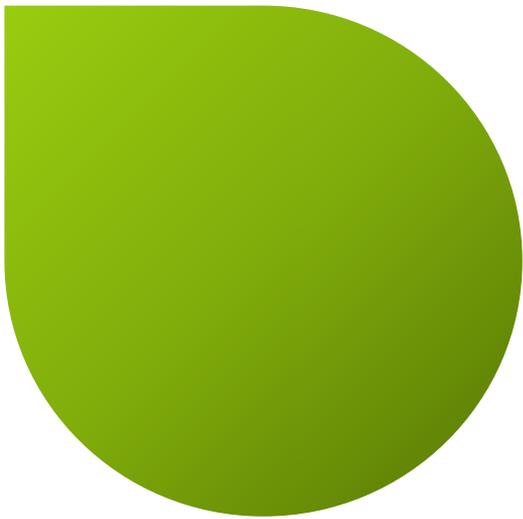
# Prologue

La coopération agricole, les fortes valeurs qui régissent le monde coopératif, l'insémination..., autant de notions lointaines, voire totalement inconnues pour moi il y a 20 ans, lorsque j'ai décidé de rejoindre l'UNCEIA (ALLICE aujourd'hui). Je ne me doutais absolument pas que le monde de l'insémination animale pouvait receler une multitude de problématiques concrètes qui appelaient à développer des programmes de recherche ambitieux à court, moyen et long termes. Aujourd'hui, 20 ans après, je reste très motivé par ce monde, au service des adhérents et des éleveurs et faisant miennes les valeurs de la coopération agricole.

Mon intérêt pour les différentes problématiques rencontrées, comme l'amélioration de la production de semence, la semence sexée ou la prédiction de la fertilité mâle et ma curiosité scientifique m'ont incité à vouloir participer activement à la création ou l'orientation des programmes de recherche mis en place. C'est ainsi que j'ai souhaité progressivement acquérir de nouvelles connaissances et compétences, via l'obtention d'un diplôme d'ingénieur de recherche au Conservatoire National des Arts et Métiers, puis ce doctorat commencé il y a trois ans.

Cette aventure a commencé il y a 20 ans par l'acquisition de connaissances sur la structure et la fonctionnalité des différents compartiments du spermatozoïde bovin. Tous ces travaux avaient pour but d'identifier des défauts de « qualité de semence » menant à une réduction du pouvoir fécondant de l'éjaculat et *in fine* à une réduction globale de la fertilité du taureau. Plusieurs phénotypes d'intérêt ont été mis en évidence, dont la combinaison permettait d'élaborer des équations prédisant environ 40% de la variation de la fertilité mâle. Ces résultats intéressants, confirmés depuis par plusieurs équipes, se sont toutefois avérés insuffisants pour une prédiction de routine suffisamment fiable et nous laissaient à penser que nous avions atteint un « plafond de verre » avec ce type de phénotypes. De ce constat et d'une rencontre avec l'équipe d'Hélène Jammes (INRAE – BREED – MEC<sup>2</sup>) est né en 2014 un Laboratoire Commun (Labcom ANR SeQuaMol), focalisé sur l'épigénome du spermatozoïde bovin. L'objectif était d'explorer de nouveaux phénotypes afin d'améliorer notre prédicteur de fertilité. C'est dans ce cadre que la thèse a été effectuée.

Au long du manuscrit, vous trouverez les éléments de contexte qui ont poussé les acteurs de l'insémination à développer des programmes de recherche autour de la fertilité mâle, ainsi que l'ensemble des travaux de recherche menés en collaboration étroite avec INRAE au cours de cette thèse.



Introduction : contexte socio-  
économique et scientifique

# Chapitre I : Le contexte de l'IA en France

« IA », acronyme d'Insémination Animale, cache une histoire passionnante, faite de savoirs empiriques, de valorisations biotechnologiques, de relations humaines et de rapports entre l'homme et l'animal (quelques dates clés sont illustrées dans la figure 1). Cette histoire débute par la domestication animale au cours du néolithique, il y a environ 13 000 ans, qui en fournissant nourriture, force de travail, vêtements et protections a provoqué des changements comportementaux et sociaux drastiques, l'Homme nomade devenant peu à peu sédentaire.

*C'était la naissance de l'élevage !*

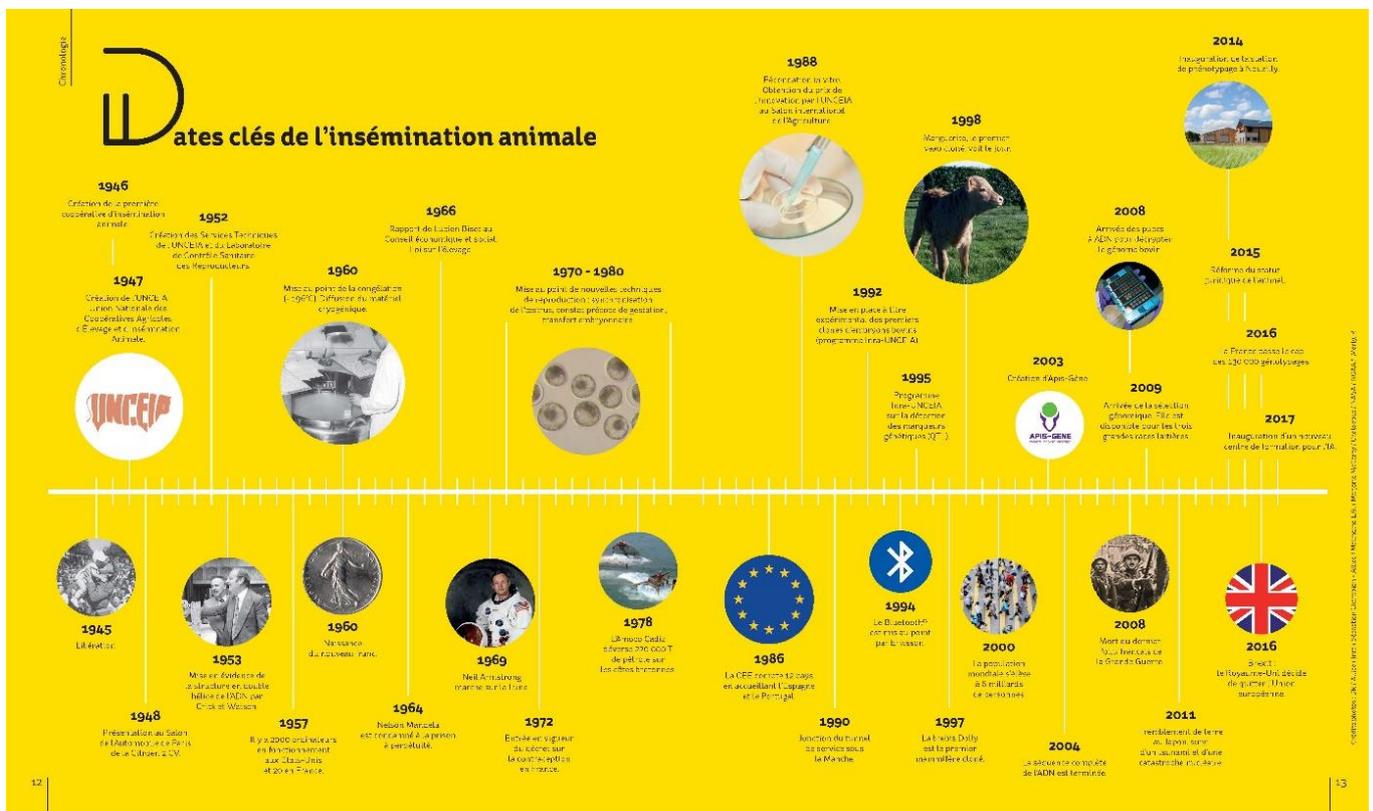


Figure 1 : Dates clés de l'insémination animale. ©Alice

Une certaine variabilité des aptitudes existant au sein des populations domestiquées, l'Homme a cherché très tôt à exploiter cette diversité pour conserver et reproduire les animaux les plus adaptés à ses besoins (lait, de viande, force de travail...), façonnant progressivement des rameaux ancestraux.

*C'était la naissance de la sélection animale « phénotypique » !*

Au sortir de la Seconde Guerre mondiale, face à la pénurie de denrées alimentaires, la puissance publique engage une action volontariste en faveur de l'augmentation de la productivité des productions animales et végétales pour assurer la sécurité alimentaire du pays. Face à un cheptel bovin décimé (de 8 700 000 têtes avant-guerre à 6 800 000 fin 1945) et de piètre productivité (les vaches laitières françaises produisaient 1 500 kg de lait par an quand celles d'outre-Atlantique affichaient près de 4 000 kg), la science et les progrès techniques ont été érigés en nécessité nationale. C'est dans ce contexte que l'IA, qui consiste à déposer le sperme au moyen d'un instrument dans le tractus génital femelle afin d'assurer la reproduction, a connu d'importantes améliorations.

Figure 2 : Lazzaro Spallanzani.



Les premières réussites d'insémination animale documentées remontent au 18<sup>e</sup> siècle, dans les années 1780, avec les travaux du moine savant Lazzaro Spallanzani<sup>1</sup> (Figure 2), qui a réussi à féconder une chienne après avoir récolté la semence de son chien. Mais il faudra néanmoins attendre le 20<sup>e</sup> siècle pour voir la pratique de l'IA se rationaliser par de nombreuses innovations techniques et technologiques, comme l'invention du vagin artificiel pour la collecte du sperme, les milieux de conservation de la semence et après-guerre, la paillette « Cassou » en 1949 (Figure 3) puis les techniques de congélation permettant son développement à grande échelle. Ce fut un succès ! Le nombre d'IA est passé de 5000 en 1946 à 2 500 000 en 1956.

*C'était le début d'une **utilisation massive de l'IA** comme mode de reproduction des bovins puis des petits ruminants !*

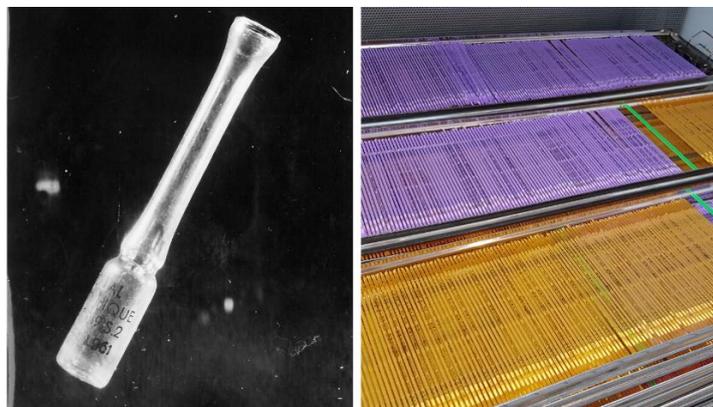


Figure 3 : Doses d'inséminations animale.

De la dose en fiole, utilisée pour de la semence fraîche à la paillette Cassou utilisée aujourd'hui pour de la semence congelée.

Au-delà du changement radical de pratique dans les élevages et des avantages sur le plan sanitaire, l'insémination permettait d'augmenter le nombre de descendants d'un même taureau et d'assurer une meilleure traçabilité des généalogies. Elle a ainsi contribué au développement de la sélection génétique au début des années 60, appuyée sur la science naissante de la génétique quantitative : l'intuition des sélectionneurs et la sélection phénotypique a laissé place à une évaluation scientifique basée sur les performances des descendants, appelée testage sur descendance. L'innovation technologique a ainsi permis d'améliorer considérablement les caractéristiques des animaux.

*C'était le début d'une utilisation massive de l'IA comme **mode de diffusion du progrès génétique** !*

La combinaison de l'IA et de la sélection génétique a en quelque sorte, pour reprendre un terme à la mode, « uberisé » l'activité de création génétique et de vente de reproducteurs jusque-là contrôlée par une certaine « aristocratie » de commerçants en bestiaux et autres sélectionneurs qui tenaient avec fermeté le marché des taureaux reproducteurs. En effet, l'innovation de rupture a conduit à un changement de paradigme où des collectifs d'éleveurs, organisés en coopératives, ont décidé de mieux maîtriser l'évolution de leur troupeau via la génétique des taureaux d'IA. Dès 1946, la première coopérative d'insémination animale (CIA) est créée en Eure-et-Loir. Elle regroupe des services et des outils coopératifs gérés par des éleveurs et pour des éleveurs. Le mode coopératif, basé sur des valeurs fortes d'honnêteté, de transparence, de responsabilité sociétale, de solidarité, et d'adhésion a permis aux éleveurs français de ruminants de garder la maîtrise des orientations de sélection et la propriété de la génétique (contrairement au monde végétal ou aux filières de monogastrique par exemple).

L'IA, comme toute innovation de rupture, a fait face à une certaine opposition, liée au frein majeur du changement de mentalité. L'union faisant la force, les acteurs de l'insémination se sont regroupés pour créer le 7 mars 1947 l'Union Nationale des Coopératives Agricoles d'Élevage et d'Insémination Animale (UNCEIA ; figure 4). Cette coopérative, qui deviendra ALLICE près de 70 ans plus tard, a eu pour mission dès sa création de fédérer et de défendre les intérêts collectifs du monde de l'insémination animale en France et à l'international, via par exemple un partenariat avec INRAE pour apporter des arguments scientifiques permettant de contrer efficacement les idées reçues.

# Compte-rendu de l'Assemblée Générale Constitutive du 7 Mars 1947

Les représentants des différentes Coopératives d'élevage et d'insémination artificielle réunis en Assemblée Générale Constitutive le 7 Mars 1947, Salle Olivier de Serres, 14 Bis, rue Scribe à Paris, décident la création d'une

Union Nationale des Coopératives d'Élevage et d'Insémination Artificielle de l'Espèce Bovine.

Après vérification des Statuts, un bureau provisoire de séance fut installé sous la présidence de M. Foucet Jean, Président de l'Union des Coopératives d'élevage de l'espèce bovine du département de l'Yonne, qui ouvre la séance à 10 heures.

L'Assemblée décida que l'Union serait constituée en application de l'article 3 du titre Ier du statut juridique de la Coopération agricole.

Il fut tout d'abord discuté du montant des parts qui fut fixé à 10.000 frs par coopérative adhérente.

Lecture des statuts fut ensuite faite et article par article ils furent adaptés. Le Conseil de l'Union des Coopératives d'élevage et d'insémination artificielle de l'espèce bovine fut ensuite nommé en se référant à l'article 18 des statuts, furent élus :

Président : M. Foucet Jean, Président de l'Union des Coopératives d'élevage de l'espèce bovine du département de l'Yonne.

Treasorier : H. Cheneveau Terul, Président de la Coopérative pour l'amélioration de l'élevage de la région de Rochefort (Ch'te Maritime).

Figure 4 : Première page du compte-rendu de l'Assemblée Générale Constitutive de l'UNCEIA.

ALLICE fédère plus de 95% des acteurs français de l'IA, réalisant plus de 7 millions d'IA en bovin, 818 000 en ovin et 74 000 en caprin. ALLICE, est une union de coopérative assez atypique (figure 5). Car en plus des missions de représentation, de lobbying et de formation, la fédération accompagne également l'innovation, avec dès 1952 un service technique chargé de l'optimisation des méthodes puis en 1959, le Centre Technique de Contrôle de la Descendance, à l'origine plus tard du service génétique du tout nouvel Institut technique de l'élevage bovin. Les thématiques de recherche abordées tournent autour du développement et de l'amélioration de la sélection génomique, de l'apport de connaissances en reproduction avec un intérêt plus particulier pour le développement de l'embryon, la fertilité des femelles et des mâles. Fortes de ces valeurs coopératives, les équipes de R&D d'ALLICE ont toujours privilégié le partage et le travail en commun. C'est pourquoi depuis le début un partenariat fort s'est noué avec l'INRA, l'institut national de recherche agronomique (devenu INRAE depuis 2020), qui perdure aujourd'hui avec la création du laboratoire commun SeQuaMol dans le cadre duquel s'est déroulée ma thèse.

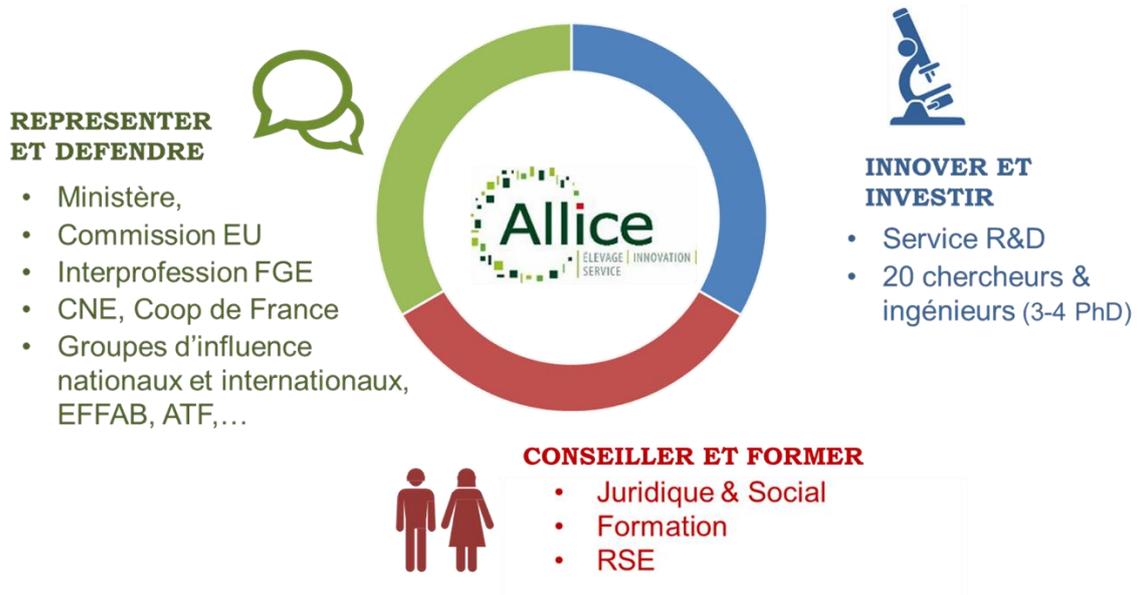


Figure 5 : Allice, une fédération atypique.

## 1) Importance de la maîtrise de la fonction sexuelle du taureau

La diffusion du progrès génétique s'appuie principalement sur la « voie mâle », via l'insémination. D'où l'importance non seulement de garantir le meilleur niveau génétique via une sélection rigoureuse, mais également d'assurer une production optimale de la semence et de garantir une bonne fertilité du taureau ou de ses éjaculats. Ces deux derniers points nécessitent une bonne connaissance de l'appareil reproducteur, de la spermatogenèse et de la physiologie du spermatozoïde, afin d'élaborer des conditions optimales d'élevage des animaux ainsi que des protocoles optimisés de collecte et de conditionnement de la semence, tout en ayant identifié les paramètres pertinents permettant de juger de la qualité de la production.

Le spermatozoïde est le résultat d'un long processus de différenciation aboutissant à une cellule hautement spécialisée, compartimentée et très particulière au regard des autres types cellulaires de l'organisme. Et pour cause, c'est la seule cellule qui a pour *destinée de vivre hors de son organisme*. Cellule haploïde, elle concourt par l'apport de son patrimoine génétique à travers la fécondation d'un ovocyte à la génération d'un nouvel individu. C'est le trait d'union entre les générations. Mais avant de parler de son rôle, reprenons son histoire : le long parcours de la différenciation gamétique.

### 1.1) La success-story du spermatozoïde : de sa conception à la fécondation

#### 1.1.1) La spermatogenèse

Le spermatozoïde est formé dans le testicule, via un processus appelé « spermatogenèse ». Ce processus biologique est défini comme l'ensemble des étapes (divisions et différenciations de cellules souches et intermédiaires) permettant la formation d'un spermatozoïde mature. En réalité, ce processus ne permet pas d'aboutir à une cellule fonctionnelle, car d'importantes modifications structurelles auront lieu en dehors du testicule permettant au spermatozoïde d'être pleinement apte à assurer la fécondation et le développement d'un embryon. Ces étapes seront traitées par la suite (voir la *maturation épидidymaire*). La spermatogenèse ne débute pas lors de la puberté, mais prend place bien avant la naissance par la mise en place structurelle et fonctionnelle du testicule (première étape de la différenciation sexuelle).

#### a) Durant la vie fœtale

Les premières ébauches des organes sexuels prennent place avec la formation des crêtes génitales, structurées par la prolifération des cellules épithéliales du mésonéphros<sup>2</sup>. L'expression de gènes clefs, tels que *GATA4*, *SF1* ou *LHX9*, est primordiale à leur mise en place puis à la survie des cellules composant ces crêtes génitales. Des travaux publiés chez la souris dont l'un ou l'autre de ces gènes a été invalidé (souris « KO »), permettent de confirmer leur importance dans la formation des gonades<sup>3,4</sup>. Pendant leur croissance, les crêtes génitales sont colonisées par les cellules germinales primordiales (PGC), qui proviennent du sac vitellin, structure transitoire dérivée de l'endoderme primitif. Chez le bovin cette colonisation intervient entre le 27<sup>e</sup> et 31<sup>e</sup> jour<sup>5</sup>. Les PGC vont alors exprimer d'importants facteurs favorisant leur prolifération et la différenciation germinale<sup>6,7</sup>, tels que les gènes *DDX4* et *DAZL*. Un défaut d'expression de ces gènes conduit à des perturbations d'établissement de la lignée germinale et *in fine* à la stérilité, comme démontré par les travaux de Gill et al., publiés en 2011<sup>8</sup> chez la souris.

Chez le mâle, sous l'action de l'expression des gènes *SRY* puis *Sox9*, les gonades embryonnaires s'orientent vers une différenciation testiculaire. Plusieurs types cellulaires interagissent alors pour former le testicule. Les cellules de Sertoli et les cellules myoïdes périvitubulaires s'organisent en « clusters » autour des cellules germinales afin de former les cordons séminifères. Les cellules de Leydig fœtales, supportant la prolifération des cellules de Sertoli, favorisent également l'expansion de ces cordons, via la production d'activin A<sup>9</sup>. À ce stade, plusieurs hormones sont nécessaires pour la poursuite du développement de l'appareil reproducteur mâle. Le testicule fœtal sécrète alors via les cellules de Sertoli, l'AMH (Anti-Müllerian hormone) qui est nécessaire à la régression des canaux de Müller. Cette hormone se lie à un récepteur sérine/thréonine kinase (AMHR2) exprimé dans les cellules du mésenchyme entourant les canaux de Müller, déclenchant la cascade d'évènements conduisant à leur régression. En parallèle de l'AMH, d'autres hormones androgènes telles que la testostérone nécessaire à la différenciation des canaux de Wolff en tractus interne (canaux efférents, déférents et épидидyme) ou la dihydrotestostérone nécessaire à la formation du tractus externe (scrotum et pénis), sont sécrétées.

Contrairement à l'action des différentes cellules somatiques, la lignée germinale ne participe aucunement à la formation du testicule fœtal. Néanmoins les cellules germinales ne restent pas inactives et entament chez le bovin, une phase de prolifération importante entre le 50<sup>e</sup> et le 80<sup>e</sup> jour de vie foetale<sup>10</sup>. Sous l'action des cellules de Sertoli les PGC se différencient en pro-spermatogonies. Elles présentent alors un profil épigénétique particulier (re-méthylation progressive) qui participe au blocage de l'expression des gènes de la pluripotence. Dans les gonades fœtales de souris, les pro-spermatogonies deviennent quiescentes et restent bloquées en mitose (phase G0/G1) jusqu'à un âge périnatal<sup>10</sup>. L'inactivation du gène d'initiation de la méiose *STRA8*, permet de conserver les pro-spermatogonies à l'état quiescent. Différents systèmes sont employés, comme la surexpression des gènes *FGF9* ou *NANOS2* qui répriment l'entrée en méiose<sup>11</sup>, ou encore l'action de la DNMT3L qui induit une reméthylation spécifique du génome, inhibant l'expression de gènes initiateurs de méiose. *DNMT3L* code un cofacteur de DNMT3A et DNMT3B, spécifique de la lignée germinale et impliqué dans la reméthylation de novo de l'ADN. Cette action sur le génome conduit aussi à une perte d'expression des gènes de pluripotence (*SOX2*, *OCT4* et *NANOG*) en faveur de ceux impliqués dans la différenciation germinale de la voie mâle.

#### b) De la vie périnatale à péri-pubertaire

Chez le bovin, du 80<sup>e</sup> jour de vie embryonnaire jusqu'à la 4<sup>e</sup> semaine de vie postnatale, le diamètre des microtubules n'excède pas 50 $\mu$ m<sup>10</sup>. La lignée germinale, positionnée au centre, reste quiescente durant toute cette période. Mais sous l'impulsion de la GnRH (Gonadotropin Releasing Hormone) produite par l'hypothalamus, l'hypophyse sécrète deux autres gonadotropes, la LH (Luteinizing Hormone) et la FSH (Follicule Stimulating Hormone). Ces hormones ont un rôle primordial dans le développement de la gonade : la LH stimule la production de testostérone par les cellules de Leydig, qui contribue à la stimulation de la spermatogenèse. La FSH, par son action sur les cellules de Sertoli, agit aussi sur la spermatogenèse. Le facteur de croissance fœtal, IGF1, favorise l'action de la LH en augmentant le nombre de récepteurs membranaires LH au niveau des cellules de Leydig. Tous ces éléments sont donc les clefs régulant la croissance du testicule, influençant le développement des tubes séminifères<sup>12</sup> et donc la production de gamètes matures.

De la 5<sup>e</sup> à la 12<sup>e</sup> semaine postnatale, le diamètre des microtubules augmente légèrement, atteignant chez le taureau Holstein 60 à 80 $\mu\text{m}$ <sup>10</sup>. L'augmentation s'accroît ensuite jusqu'à la semaine 32, où les microtubules présentent un diamètre d'environ 208 $\mu\text{m}$ <sup>13</sup>. Dès le début de l'augmentation du diamètre des microtubules, les cellules germinales migrent d'un positionnement central vers un positionnement basal. En parallèle, le nombre de cellules de Sertoli augmente tout au long de ces 32 premières semaines de vie. Leur nombre (par testicule) a été quantifié partant de quelques millions dans les semaines juste après la naissance à quelques milliards à 32 semaines de vie postnatale<sup>13</sup>. Leur multiplication chez le taureau s'arrête après le 6<sup>e</sup> mois de vie postnatale. Positionnées autour de la lame basale, polarisées et présentant d'étroites jonctions entre elles, elles forment une véritable barrière testiculaire. Le nombre de cellules germinales et leur différenciation sont étroitement liés au nombre de cellules de Sertoli<sup>13,14</sup>. En effet, les cellules de Sertoli assurent de nombreuses fonctions indispensables, telles qu'un maintien de la compartimentation du testicule, un soutien nutritionnel nécessaire à la prolifération et à la différenciation des cellules germinales, une participation à l'orientation de leur transfert vers la lumière des tubes séminifères et un rôle de « nettoyage » par phagocytose des cellules dégénérées et des corps résiduels<sup>15</sup>.

L'augmentation du diamètre et de la longueur des microtubules est corrélée à l'augmentation du volume des testicules. De la 12<sup>e</sup> semaine postnatale à la 32<sup>e</sup> semaine, la circonférence scrotale des testicules double pour atteindre en moyenne 27cm chez les taureaux Holstein<sup>13</sup>. Cette augmentation est elle-même corrélée à la prise de poids du taureau (en moyenne pour la race Holstein, de 79kg à 12 semaines et jusqu'à 275kg à 32 semaines).

Sous l'influence des signaux hormonaux et de l'action des cellules de Sertoli, la première phase de la spermatogenèse, appelée spermatocytogenèse, reprend (figure 6). Cette phase correspond à la différenciation des spermatogonies en spermatocytes puis en spermatides rondes. Longue de 44 jours chez le bovin<sup>15</sup>, elle se décompose en deux phases : mitotique et méiotique. La phase mitotique permet le renouvellement des spermatogonies, constituant la réserve de cellules souches pour la vie reproductive de l'individu. Une partie des spermatogonies entame une différenciation vers les différents types de spermatogonies (A1, A2, A3, B1 et B2) puis finalement les spermatocytes de Type I. À ce stade les cellules sont prêtes à entamer la phase méiotique. Cette étape cruciale voit la réduction de la moitié du patrimoine génétique : les spermatocytes de type I, cellules diploïdes, évoluent en

spermatides rondes, cellules haploïdes. Chez le taureau, l'apparition de spermatogonies de type A a lieu à partir de la 16<sup>e</sup> semaine postnatale. Quatre semaines supplémentaires sont nécessaires pour l'apparition des spermatogonies de type B. Les spermatocytes de type I, aux stades leptotène, zygotène et pachytène apparaissent à partir de la 24<sup>e</sup> semaine. Les premières spermatides en cours d'élongation ou allongées apparaissent dès 28 semaines postnatale<sup>13</sup>. La dernière grande phase de la spermatogenèse, la spermiogenèse est de 17 jours chez le bovin<sup>15</sup>. Elle se compose de nombreuses étapes conduisant à des modifications de structures formant *in fine* les spermatozoïdes<sup>16,17</sup>. Ce n'est qu'au bout de 32 semaines postnatales chez le bovin, que les premiers spermatozoïdes apparaissent dans la lumière des tubes séminifères<sup>13</sup>. Ces étapes, décrites dès 1952 par Leblond et Clermont<sup>18</sup> chez le rat, sont au nombre de 14 chez le bovin<sup>19</sup>. Les 3 premières étapes voient principalement la formation de la « granule acrosomique ». Cet élément continue de se développer pour former l'acrosome durant les étapes suivantes. Les étapes 4 à 12 présentent les plus fortes modifications :

- l'acrosome poursuit son extension,
- le noyau se condense fortement dû à un remplacement massif d'une grande majorité des histones par des protamines<sup>20</sup> permettant un fort enroulement de l'ADN et la formation de toroïdes (structure spécifique du spermatozoïde),
- l'axonème se développe et la cellule « s'allonge ».

Les dernières étapes correspondent à la formation de la pièce intermédiaire et du flagelle, à la condensation terminale du noyau et à la concentration du reste du cytoplasme dans la « gouttelette cytoplasmique » (qui sera expulsée au cours du transit dans l'épididyme).

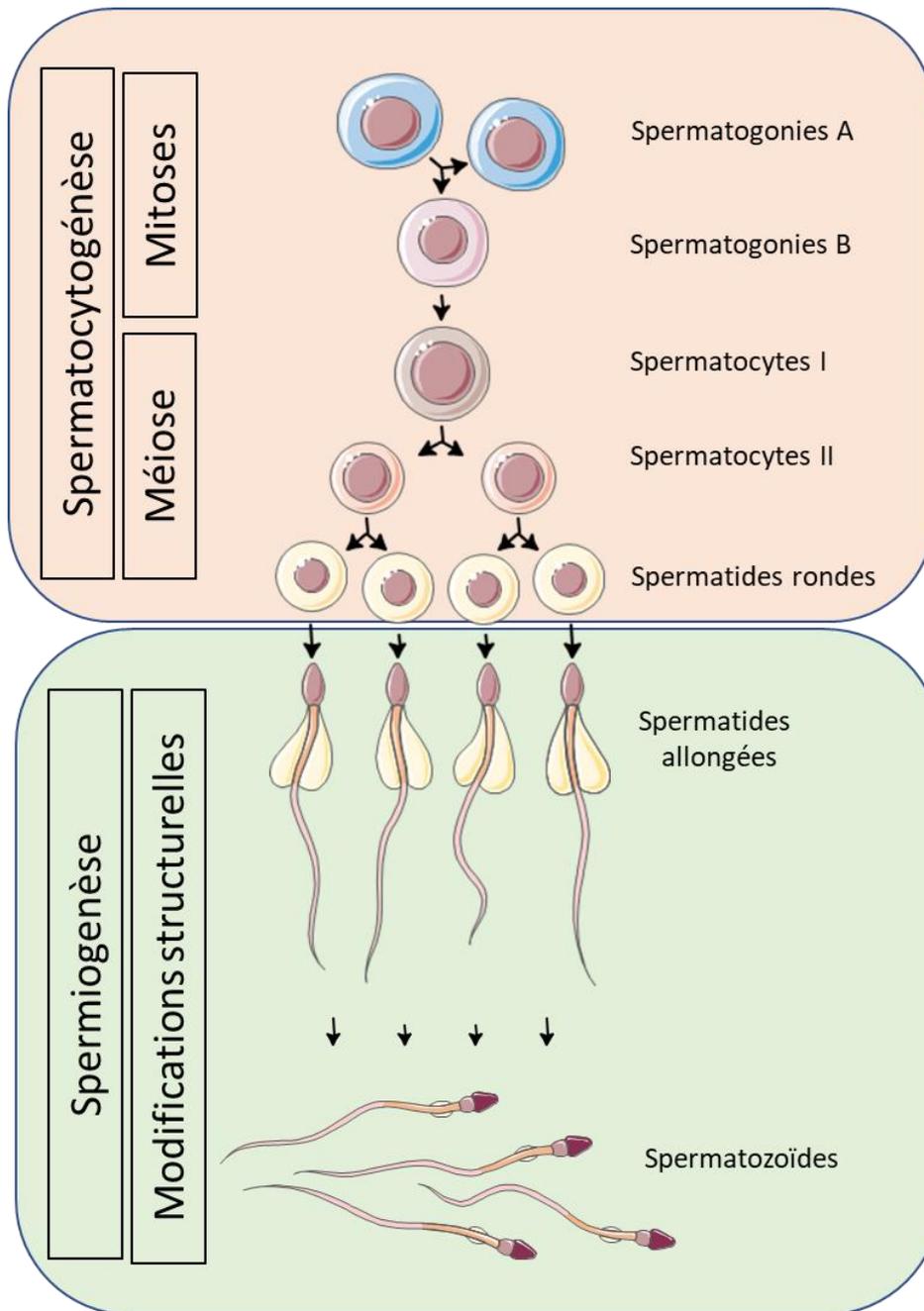


Figure 6 : La spermatogénèse.

La spermatogénèse comporte 2 grandes phases : la spermatocytogénèse et la spermiogénèse. La première est une succession d'étapes partant des cellules souches, les spermatogonies, en passant par les spermatocytes et en allant jusqu'aux spermatides rondes. Les mitoses permettent aux spermatogonies de se multiplier, renouvelant ainsi le stock de cellules souches. La méiose engagée à partir des spermatocytes de type I permet de passer d'une cellule diploïde à une cellule haploïde. La seconde phase est le siège d'un ensemble de modifications structurelles permettant aux spermatides rondes de s'allonger pour aboutir à la forme finale du spermatozoïde. @Alicia

Toutes ces phases de différenciation cellulaire s'accompagnent d'une migration progressive vers la lumière des tubes séminifères. Les spermatozoïdes nouvellement formés sont donc déversés dans la lumière des tubes séminifères, puis ils migrent dans le rete testis et quittent le testicule. Ils transitent par l'épididyme où ils termineront leur maturation<sup>21</sup>. Chez le bovin, les premiers cycles de spermatogenèse post pubertaires sont plus longs (environ 80 jours<sup>13</sup> à 32 semaines postnatales) qu'en pleine maturité (61 jours<sup>15</sup>). La quantité de spermatozoïdes produits est plus faible (quelques millions par ml dans les premiers mois contre 1 à 1.5 milliard par ml à partir de 12 mois postnatal). La qualité des gamètes produits est également moins bonne. Les spermatozoïdes sont peu mobiles<sup>22</sup> (moins de progressivité) et présentent des anomalies morphologiques de la tête, du flagelle et de l'acrosome<sup>23</sup>. Il faut attendre entre 4 et 5 mois supplémentaires post-puberté avant d'avoir une spermatogenèse efficace et stable. Chez les bovins, la puberté est considérée comme atteinte lorsque le mâle est capable de fournir un éjaculat contenant 50 millions de spermatozoïdes par ml ayant une mobilité (progressivité) supérieure ou égale à 10%. Ce seuil est variable sous l'influence de facteurs exogènes tels que l'alimentation du veau (forte corrélation entre poids de l'individu et survenue de la puberté), mais également en fonction des races et entre individus. Lunstra & Echtenkamp en 1982 ont illustré cette variabilité entre 6 races bovines, montrant un écart pouvant atteindre 62 jours entre les deux races extrêmes<sup>23</sup>.

### *1.1.2) La maturation épидидymaire*

L'épididyme est un organe adossé à chaque testicule, formé d'un très long tubule (plusieurs dizaines de mètres chez le taureau<sup>24</sup>), fortement contourné et pelotonné, qui met en communication les canaux efférents en amont et le canal déférent en aval. Plusieurs jours sont donc nécessaires à sa traversée (6 à 11 chez le bovin<sup>25</sup>). Il est habituellement décrit comme ayant 3 parties principales, la tête, le corps et la queue (figure 7), mais d'autres subdivisions ont été proposées selon les espèces. De nombreux échanges entre le spermatozoïde et les composants du fluide épидидymaire ont lieu au cours du transit des gamètes, participant aux importantes modifications de structures et/ou de fonctionnalités du spermatozoïde.

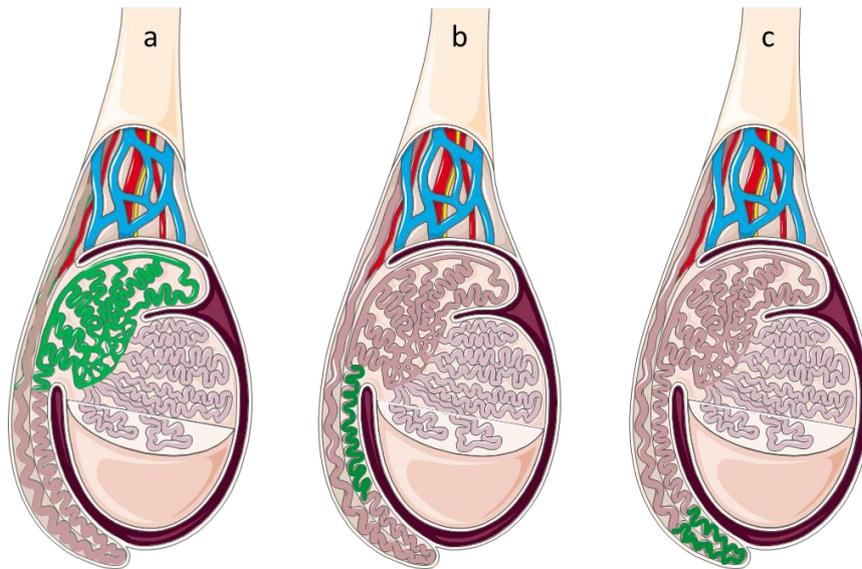


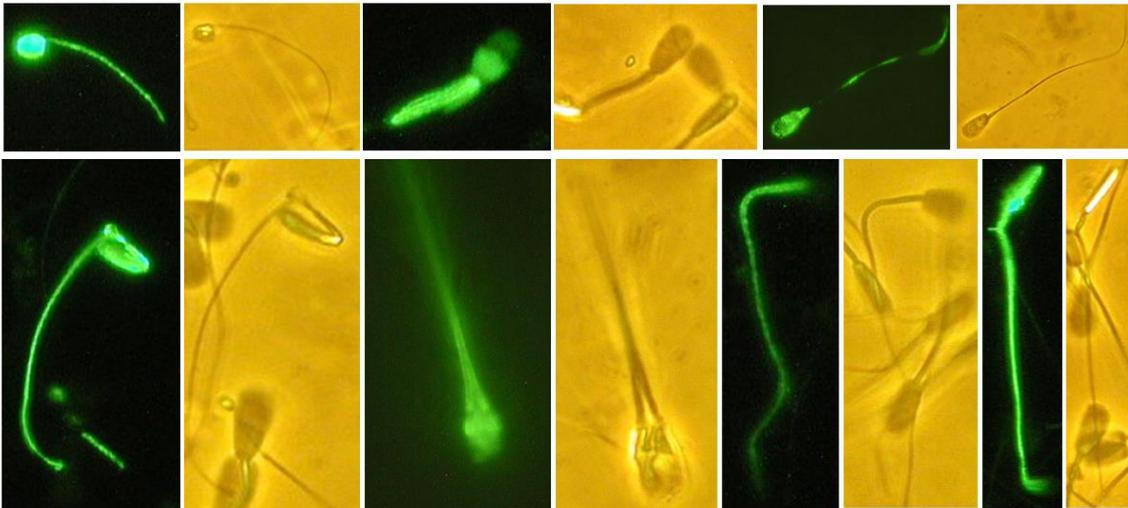
Figure 7 : Représentation des différentes parties de l'épididyme © Alice  
(a) la tête (caput) ; (b) Le corps (corpus) ; (c) la queue (cauda)

#### a) Les différents rôles de l'épididyme

Différents rôles sont attribués à l'épididyme. Il intervient tout d'abord dans le transport des spermatozoïdes<sup>24</sup>. Le gamète mâle, au sortir du testicule, n'a pas encore acquis sa capacité de mouvement. Il dépend donc de mécanismes extérieurs pour se mouvoir. Des contractions de la paroi des tubules épидидymaires et l'action des cellules ciliées permettent aux spermatozoïdes de transiter de la tête vers la queue de l'épididyme. Cet organe sert également de réservoir spermatique<sup>24</sup>. L'épididyme, et plus particulièrement la région caudale de l'organe est en mesure de conserver les spermatozoïdes pendant plusieurs jours. Le temps de stockage est variable d'une espèce à l'autre. Chez l'homme ou le taureau, deux tiers des gamètes extratesticulaires peuvent être conservés entre 2 et 5 jours dans la queue de l'épididyme, alors que chez le verrat la totalité des spermatozoïdes peut être conservée 3 semaines.

Un stockage de plusieurs jours des spermatozoïdes permet de mettre en place un système de « contrôle qualité » des gamètes<sup>24</sup>. Les spermatozoïdes trop vieux ou morphologiquement anormaux sont éliminés par l'épididyme. Ils seraient recouverts d'ubiquitine, un peptide de 8.5 KDa, les désignant comme une cible pour les protéines du protéasome<sup>26,27</sup>. Ces spermatozoïdes sont ensuite phagocytés par les cellules épithéliales épидидymaires, puis dégradés par les macrophages environnants. Sutovsky et son équipe confirment le rôle de

l'ubiquitine<sup>27,28</sup> et montrent une corrélation entre taux de spermatozoïdes anormaux et pourcentage de spermatozoïdes recouverts d'ubiquitine. Des travaux non publiés par ALLICE dans les années 2000 confirment ces résultats (figure 8). À noter que ces observations ont été faites sur de la semence éjaculée, laissant supposer que dans un contexte d'anomalies morphologiques chroniques, le système de « contrôle qualité » n'est pas suffisamment efficace pour éliminer tous les spermatozoïdes anormaux.



*Figure 8 : Différentes anomalies morphologiques du spermatozoïde bovin éjaculé. Visualisation des spermatozoïdes anormaux en microscopie à fluorescence (x400). Marquage de l'ubiquitine avec un anticorps couplé à FITC. Différentes structures anormales sont recouvertes d'ubiquitine : anomalies de la tête (problèmes de développement), de la pièce intermédiaire (courbée ou problèmes de développement) ou de l'ensemble du spermatozoïde (forme double) ©Allice*

L'épididyme assure également une protection des spermatozoïdes, qui se doit d'être d'autant plus efficace que le gamète mâle est une cellule fragile. En effet, son état quiescent ne lui permet pas de répondre de façon autonome à un quelconque stress, ni en déclenchant une réponse reposant sur l'expression de gènes de défense, ni en mobilisant ses faibles ressources cytoplasmiques en protéines de défense. Le spermatozoïde est en particulier vulnérable aux attaques radicalaires par les ROS (*Reactive Oxygen Species*), sensibilité accrue par la richesse en acides gras polyinsaturés de sa membrane plasmique. Le stress oxydant peut ainsi conduire à l'oxydation des membranes ou de la chromatine du spermatozoïde, le rendant plus fragile, moins fonctionnel et dans certains cas non fécondant<sup>29</sup>. Parmi l'ensemble des gènes exprimés par les cellules épithéliales de l'épididyme, le gène *GPX5*, appartenant à la famille des glutathion peroxydase, est fortement exprimé dans la tête de l'organe. La protéine GPX5 est alors sécrétée dans la lumière du tube et suit les spermatozoïdes dans leur transit vers la partie

caudale<sup>29</sup>. Impliquée dans la réduction des éléments oxydés, notamment le peroxyde d'hydrogène, tout au long du transit épидидymaire et surtout lors du stockage prolongé dans la partie caudale de l'organe, cette protéine participe à la protection des gamètes contre le stress oxydant<sup>29,30</sup>.

Le dernier rôle et non des moindres, attribué à l'épididyme est d'être le siège de l'acquisition du pouvoir fécondant du spermatozoïde<sup>24</sup>. Le spermatozoïde libéré dans la lumière des tubes séminifères n'est pas encore totalement mature. Certaines modifications de structure sont nécessaires et se mettent en place au cours du transit épидидymaire. Il s'agit par exemple de la migration de la gouttelette cytoplasmique proximale en position distale avant son évacuation complète du spermatozoïde. Cette migration est réalisée entre la tête et la queue de l'épididyme<sup>31</sup>. Cette gouttelette contient les derniers restes du cytoplasme non totalement éliminé lors de la spermiogenèse. Sa perte est concomitante avec l'acquisition de la motilité du spermatozoïde dans la partie caudale de l'épididyme. Il est possible de retrouver des spermatozoïdes éjaculés ayant gardé leur gouttelette. Il s'agit d'une anomalie morphologique majeure si elle est encore en position proximale ou mineure si elle est distale (figure 9). Des travaux de fécondation *in vitro* ont démontré l'impact négatif sur le taux de clivage quand plus de 30% des spermatozoïdes d'un éjaculat portent cette anomalie majeure<sup>32</sup>.



Figure 9 : Gouttelette cytoplasmique du spermatozoïde bovin éjaculé.

Un spermatozoïde éjaculé « normal » ne contient plus de gouttelette cytoplasmique (A). Si elle est retrouvée en position proximale (B) ou distale (C), il s'agit d'une anomalie morphologique. Le spermatozoïde de l'image (B) présente une seconde anomalie majeure : la décapitation. @LNCR

## b) L'épididyme, un lieu d'échange avec le spermatozoïde

Tous les changements majeurs que subit le spermatozoïde durant son transit épидидymaire proviennent d'interactions entre les composants du fluide épидидymaire et le gamète lui-même. L'épididyme étant un organe régionalisé, la composition du fluide présente des spécificités au niveau de la tête, du corps ou de la queue.

Le contenu protéique du fluide épидидymaire varie en quantité entre la tête et la queue de l'organe, de 2 à 4 mg/ml à 50 à 60 mg/ml<sup>21</sup>, respectivement. Chez le taureau et le verrat<sup>33</sup>, une dizaine de protéines représentent plus de 80% de l'expression totale des protéines de l'épididyme, dont la lipocaline, la clusterine, la transferrine, la glutathion peroxydase 5, l'albumine ou encore la prostaglandine D2 synthase. Leur concentration varie suivant leur localisation. En effet, la glutathion peroxydase 5, l'albumine, la prostaglandine D2 synthase ou la transferrine sont plus exprimées dans la tête de l'épididyme, alors que la clusterine, la lipocaline 5 ou la beta-N-acetyl-hexosaminidase le sont principalement dans la partie caudale<sup>21</sup>. Toutes ces protéines proviennent de l'activité exocrine des cellules épithéliales épидидymaires. Les changements séquentiels de composition protéique du fluide épидидymaire sont à l'origine des modifications protéiques de la membrane cytoplasmique du spermatozoïde<sup>21,33</sup>. Certaines protéines de surface du spermatozoïde « testiculaire » sont clivées lors de leur passage dans l'épididyme, modifiant leur conformation externe (fertiline ou angiotensin-converting enzyme)<sup>21</sup>. D'autres sont au contraire ajoutées à la membrane du gamète. Il s'agit entre autres de certaines Heat-shock protéine (Hsp70 ou 90), de la tubuline ou encore l'énolase 1 (ENO1). Des hypothèses ont été proposées concernant leur rôle dans le contrôle de la migration de la gouttelette cytoplasmique (due à la présence d'actine, de tubuline ou des Hsp) ou dans la participation à la défense antioxydante du gamète (ENO1, CCT2, PRDX5, HIP1...)<sup>21</sup> et l'invalidation chez la souris des gènes codants pour ces protéines de la lumière de l'épididyme, induit une baisse de fertilité, voire une stérilité (*SED1*, *ESR1*, *C-ROS*, *DICER1*, *APOER2*...figure 10)<sup>30</sup>.

Gene knockout	Fertility phenotype	Changes to epididymal environment	References (PMID)
<i>Sed1</i>	Infertile	Hypo-osmotic and alkaline epididymal fluid, disrupted fluid reabsorption, increased intracellular vesicles	20122713
<i>Esr1</i>	Infertile	Hypo-osmotic fluid	20130266
<i>C-ros</i>	Infertile	Defective initial segment development, increased luminal pH	10645273, 15095336
<i>Dicer1</i>	Subfertile	Imbalanced lipid homeostasis in proximal segments. Dedifferentiation of the epithelium and imbalance in sex steroid signaling	25366345, 22701646
<i>Rlx</i>	Subfertile	Delayed maturation and growth associated with increased collagen deposition	15956703
<i>He6</i>	Infertile	Reduced in size and dysregulation of fluid reabsorption	15367682
<i>Lur (testosterone treatment)</i>	Subfertile	Inflammation in epididymis	15514086
<i>Erα</i>	Infertile	Disruption in Na <sup>+</sup> reabsorption and passive water transport, abnormal epithelial ultrastructure	11698654
<i>Nhe3</i>	Infertile	Disruption in Na <sup>+</sup> reabsorption and passive water transport	11698654
<i>Lxr</i>	Infertile	Abnormal accumulation neutral lipids	15525595
<i>Apoer2</i>	Infertile	Dysfunction of clusterin and PHGPx protein impacting sperm maturation	12695510
<i>Fsh-r</i>	Subfertile	Smaller epithelial surface area in caput and corpus segments	15973687
<i>Gpx5</i>	Higher incidence of miscarriages and developmental defects	Excess of reactive oxygen species in the cauda segment leading to oxidative damage of spermatozoa	19546506
<i>Hoxa10</i>	Subfertile	Epididymis characterized by homeotic transformation	8787743
<i>Hoxa11</i>	Infertile	Epididymis characterized by homeotic transformation	7789268
<i>Hexa</i>	Infertile (age dependent)	Inability to degrade endocytosed substrates	12617783
<i>Tmf</i>	Infertile	Epithelial apoptosis and sperm stasis in the cauda segment	23000399
<i>Trpv6</i>	Subfertile	Defects in epididymal Ca <sup>2+</sup> absorption	22427671
<i>Slc9a3</i>	Infertile	Abnormally abundant secretions and calcification in the lumen	28384194

*Figure 10 : Impact de l'extinction de différents gènes sur l'environnement luminal de l'épididyme de souris<sup>30</sup>. La perturbation de l'expression de ces gènes impacte l'environnement épидидymaire, conduisant à d'importants problèmes de fertilité, incluant la stérilité.*

Des travaux récents ont par ailleurs montré l'existence d'interactions entre les spermatozoïdes et les vésicules extracellulaires du fluide, en particulier sur les épидидymosomes, qui pourraient jouer un rôle majeur dans la maturation via le transfert de matériel biologique depuis les cellules épithéliales de l'épididyme.

Ces vésicules extracellulaires, délimitées par une double couche lipidique, proviennent d'une sécrétion apocrine (bourgeonnement de la membrane plasmique) des cellules épithéliales de l'épididyme. Une nomenclature particulière a été établie pour les catégoriser<sup>34</sup>. Elle se base sur les caractéristiques physiques des vésicules, telles que leur taille (50 à 300nm), leur densité (gradient de sucrose) ou leur composition membranaire (CD63, CD9, CD81, Annexine V...). Le tissu excréteur d'origine permet également de les différencier. Suivant ces critères, ces vésicules appartiennent à des catégories particulières telles que les « oncosomes, ectosomes, microvésicules, petites moyennes ou grandes vésicules, exosomes, épидидymosomes, prostasomes...). Leur contenu (protéique, lipidique...) semble être différent suivant la catégorie de vésicule. Dans ce manuscrit, nous utiliserons les termes employés par les auteurs des publications choisies pour illustrer nos propos. Sans information spécifique de la catégorie, nous utiliserons le terme de vésicules extracellulaires.

Chez le bovin, plusieurs centaines de protéines ont été identifiées à partir d'épididymosomes purifiés par ultracentrifugation sur gradient de sucrose<sup>35</sup>. Bien que sécrétés principalement dans la tête de l'épididyme<sup>36</sup>, la composition des épididymosomes semble différer en fonction de la région dans laquelle ils sont prélevés<sup>35</sup>. Les épididymosomes seraient capables de fusionner avec la membrane cytoplasmique du spermatozoïde afin d'y déverser leur contenu protéique<sup>37,38</sup> ou d'induire des modifications biochimiques (lipidique<sup>36</sup> notamment). Chez le bovin il a été démontré que les protéines ELSPBB1 et GBB2 étaient transférées depuis les épididymosomes vers les spermatozoïdes<sup>35</sup> dans la partie caudale de l'épididyme. D'autres exemples ont été publiés comme la cenexine chez le chat<sup>38</sup>, transférée entre la tête et la queue de l'épididyme via des épididymosomes.

Le contenu lipidique varie également suivant la section de l'organe. Dans la tête de l'épididyme, 60 à 70% des lipides du spermatozoïde murin sont des phosphatidyléthanolamines et phosphatidylcholines. Cette proportion diminue et ne représente plus qu'environ 40% des lipides pour les spermatozoïdes prélevés dans la queue de l'épididyme. En parallèle, la proportion de sphingomyéline spermatique augmente d'un facteur 2 entre la tête et la région caudale de l'épididyme, où elle constitue le lipide le plus représenté<sup>36</sup>. À *contrario*, le cholestérol, la phosphatidylcholine et le phosphatidylinositol, abondants dans la tête de l'organe, sont nettement moins présents dans la queue de l'épididyme. Toutes ces différences quantitatives et qualitatives sont également observées dans le contenu des épididymosomes<sup>36</sup>, suggérant leur implication dans les variations observées au niveau du spermatozoïde.

### 1.1.3) *Le plasma séminal*

Le plasma séminal joue également un rôle important dans la protection et la maturation du spermatozoïde éjaculé. Ce liquide, qui représente jusqu'à 80% du volume de l'éjaculat, est un milieu nutritif et protecteur, composé d'ions ( $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Zn}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Cl}_2$ ), de substrats énergétiques (fructose, sorbitol...) et de diverses protéines<sup>39,40</sup>. Il assure une pression osmotique optimum, apporte des nutriments, participe à l'amélioration de la motilité et au blocage de la capacitation précoce du spermatozoïde<sup>41</sup>. Le liquide séminal favorise aussi la protection des gamètes vis-à-vis des réactions inflammatoires dans les voies génitales femelles. Ce contrôle immunitaire contribue aussi à la préparation de l'utérus à recevoir

l'embryon nouvellement conçu favorisant son implantation et son développement<sup>42</sup>. *In fine* le liquide séminal influence ainsi la fertilité<sup>39</sup>.

La composition en protéines du plasma séminal bovin a été étudiée. Plus de 1000 protéines ont été détectées par spectrométrie de masse. La famille de protéines la plus abondante, appelée BSP (Bovine Seminal Plasma) comprend plusieurs membres : BSP-A1, A2, A3 ou BSP-30, qui se lie aux phosphatidylcholines présentes dans la membrane plasmique<sup>41</sup>, évitant une capacitation précoce du spermatozoïde. Ces protéines interagissent par la suite avec les HDL (High Density Lipoprotein) présentes dans le fluide oviductal, conduisant à la libération du cholestérol membranaire du spermatozoïde et *in fine* à sa capacitation<sup>41</sup>. Parmi ces protéines, 80 ont été reliées à la fertilité des mâles<sup>39</sup>, certaines sont surexprimées (TIMP-2, BSP5...) et d'autres sous-exprimées chez les taureaux fertiles (clusterine, galectin-3-binding). Or, chez le bovin, les traitements biotechnologiques de la semence utilisée en IA conduisent à réduire le plasma séminal à l'état de trace. En effet la semence est hautement diluée pour la fabrication des doses d'insémination. Au regard de l'importance de ce milieu pour les spermatozoïdes, mais également potentiellement pour le tractus génital femelle, il serait intéressant de mieux le prendre en considération et d'orienter la composition des milieux de conservation de la semence en fonction de la composition moléculaire et des caractéristiques physico-chimique du plasma séminal. Ceci permettrait potentiellement d'améliorer le pouvoir fécondant des spermatozoïdes et plus globalement la fertilité des taureaux<sup>39,43</sup>.

Le plasma séminal est aussi chargé en vésicules extra cellulaires provenant de la prostate<sup>44</sup>. Ces prostasomes, tout comme les épидидyosomes transfèrent différents constituants protéiques ou lipidiques<sup>45</sup> vers les spermatozoïdes depuis les cellules épithéliales de la prostate. Ils contiennent une forte proportion de sphingomyéline et de cholestérol, qui permettraient de rigidifier un peu plus la membrane cytoplasmique afin de contribuer au blocage de la capacitation et/ou une réaction acrosomique précoce<sup>45</sup>, en complément des autres systèmes (BSP notamment). Outre l'albumine qui constitue leur constituant principal, les prostasomes contiennent diverses protéines en commun avec les épидидyosomes (prostaglandine-D2-synthase, lipocaline, GPX5 ou clusterine), ainsi que des protéines spécifiques (FOLH1, PSMA, TGM4...)<sup>45</sup>. Chez l'homme, certaines protéines des prostasomes ont été décrites comme biomarqueurs d'infertilité<sup>45</sup> et signeraient la présence d'anomalies de

la spermatogenèse, telles que l'azoospermie (clusterine), l'oligoasthenozoospermie (cathepsin, ATP5A1), le varicocèle (SERPINA1) ou encore la teratozoospermie (LTF).

#### 1.1.4) Dans les voies femelles et les premières étapes de la fécondation

« *Nous partîmes cinq cents, mais par un prompt renfort nous nous vîmes trois mille en arrivant au port* ». À l'inverse de cette célèbre citation de Corneille (Le Cid 1636), le périple des spermatozoïdes dans le tractus femelle débute par plusieurs millions de spermatozoïdes pour qu'*in fine* un seul puisse arriver à bon port et féconder l'ovocyte<sup>46</sup>.

Lors de l'insémination, les spermatozoïdes sont déposés dans le corps de l'utérus, après le col. De ce fait, ils évitent les interactions parfois négatives avec le milieu vaginal (notamment dues au pH). Il en résulte un besoin bien moins important en nombre de gamètes pour la réussite de l'IA (de 2 à 20 millions de spermatozoïdes, respectivement pour la semence sexée ou conventionnelle). Entre 6 et 8 heures sont nécessaires aux gamètes pour atteindre en nombre suffisant l'ampoule de l'oviducte, lieux de rencontre avec l'ovocyte<sup>47</sup>.

Dans l'utérus, les spermatozoïdes, aidés par les mouvements engendrés par les contractions des muscles lisses, mettent en jeu leur capacité de rhéotaxie pour rejoindre l'oviducte. Plusieurs mécanismes semblent concourir à réguler leur entrée dans l'oviducte au niveau de la jonction utero-tubaire<sup>46</sup> : la viscosité du mucus, l'anatomie de la zone en « entonnoir » et l'obligation pour le spermatozoïde d'avoir la bonne « clef » pour ouvrir le passage. En effet, la présence de certaines protéines de la famille des ADAM<sup>46,48</sup> semble être un prérequis obligatoire. Chez la souris, seuls les spermatozoïdes présentant les protéines ADAM1a, ADAM2 et ADAM3 à la surface membranaire, sont retrouvés dans l'oviducte<sup>48</sup>. De plus, même dans un mélange de spermatozoïdes déficients ou non en ADAM, seuls ceux avec les protéines de cette famille sont retrouvés dans l'oviducte. Les spermatozoïdes déficients en ces protéines sont incapables de franchir la jonction utero-tubaire<sup>48</sup>. La présence de spermatozoïdes normaux ne suffit donc pas à « ouvrir » le passage à un groupe de gamètes mâles. Il semble donc, tout du moins chez la souris, que cette famille de protéines représente un « *sauf-*

*conduit* » individuel pour le passage vers l'oviducte. En revanche, la présence ou l'absence de ces protéines n'impacte pas la fécondation des ovocytes réalisée *in vitro*<sup>48</sup>.

Dans l'oviducte, les spermatozoïdes progressent dans le fluide oviductal visqueux. Des échanges sont possibles, via vraisemblablement les oviductosomes<sup>46</sup>, les composants de ce fluide et le spermatozoïde qui acquière ainsi de nouvelles protéines, telles que l'oviduct-specific glycoprotein (OGP) ou l'oveosontin. Ces deux protéines ne sont pas strictement nécessaires pour la fécondation ou le développement embryonnaire (chez la souris), mais elles permettent d'améliorer (*in vitro*) la mobilité des spermatozoïdes, leur viabilité, et rendent la zone pellucide (une triple couche dense en protéines glycosylées) plus permissive aux gamètes mâles.

Les spermatozoïdes ayant un acrosome intact peuvent également se lier aux cellules épithéliales<sup>46,49</sup>, permettant une rétention locale et formant un « réservoir spermatique ». Cette liaison stoppe la mobilité des spermatozoïdes, évite potentiellement le déclenchement de la capacitation en bloquant l'influx de calcium intra cellulaire<sup>46</sup> et prolonge dans le temps les capacités fonctionnelles du spermatozoïde<sup>50</sup>. Si les mécanismes sous-jacents restent encore aujourd'hui méconnus, le rôle de l'interaction directe entre le gamète et les cellules épithéliales est toutefois établi, le fluide oviductal n'étant pas responsable de ce prolongement de fonctionnalité<sup>46</sup>. L'interaction semble par ailleurs induire une modification de l'expression des gènes des cellules épithéliales, notamment ceux impliqués dans la réponse inflammatoire<sup>46</sup>. Chez le bovin, cette interaction entraîne une réponse anti-inflammatoire par une surexpression de l'interleukine IL-10, du TGF $\beta$  et de la prostaglandine E2. Toutes ces actions permettent aux spermatozoïdes d'évoluer dans un environnement plus favorable jusqu'à la fécondation.

Les mécanismes responsables du relargage ultérieur des spermatozoïdes ne sont pas connus. Plusieurs hypothèses sont avancées, comme l'existence d'un « signal » inducteur provenant du complexe ovocyte-cumulus, ou un relargage progressif et continu des gamètes, afin d'assurer une plage spatio-temporelle optimale pour la rencontre du spermatozoïde et de l'ovocyte. Ce relargage contrôlé permettrait également de limiter les risques de polyspermie<sup>50</sup>.

Deux évènements surviennent alors pour achever la maturation finale du spermatozoïde : la capacitation et l'hyperactivation. La capacitation correspond à l'ensemble des modifications membranaires rendant apte le spermatozoïde à pénétrer la zone pellucide et l'ovocyte. L'une des premières modifications réside dans l'exclusion du cholestérol membranaire, conduisant à un remodelage de la composition de la membrane plasmique<sup>51</sup>. Ces changements membranaires rendent le spermatozoïde plus perméable à l'entrée d'ions, tels que  $\text{Ca}^{2+}$ ,  $\text{HCO}_3^-$  ou  $\text{K}^+$ . L'internalisation de ces ions à fortes concentrations est requise pour le déclenchement de la réaction acrosomique. De façon intéressante, il semble que le spermatozoïde ait besoin de la présence d'oxydants (ROS) pour initier sa capacitation<sup>52</sup>. Un équilibre doit donc être trouvé entre la concentration en ROS requise pour ce processus biologique et celle, trop forte, délétère pour les structures. Concernant l'initiation de la capacitation, les ROS agissent sur l'exclusion du cholestérol membranaire et provoquent la génération d'AMPC via la stimulation de la soluble adenylyl cyclase (sAC). Cette enzyme catalyse la formation d'AMPC à partir d'ATP en présence des ions  $\text{Ca}^{2+}$  et  $\text{HCO}_3^-$ , la formation d'AMPC à partir d'ATP, point de départ d'une cascade de phosphorylation amenant à l'hyperactivation du spermatozoïde<sup>52</sup> (processus biologique qui permet d'augmenter l'amplitude des mouvements du flagelle, conférant aux spermatozoïdes une motilité plus efficace ; figure 11).

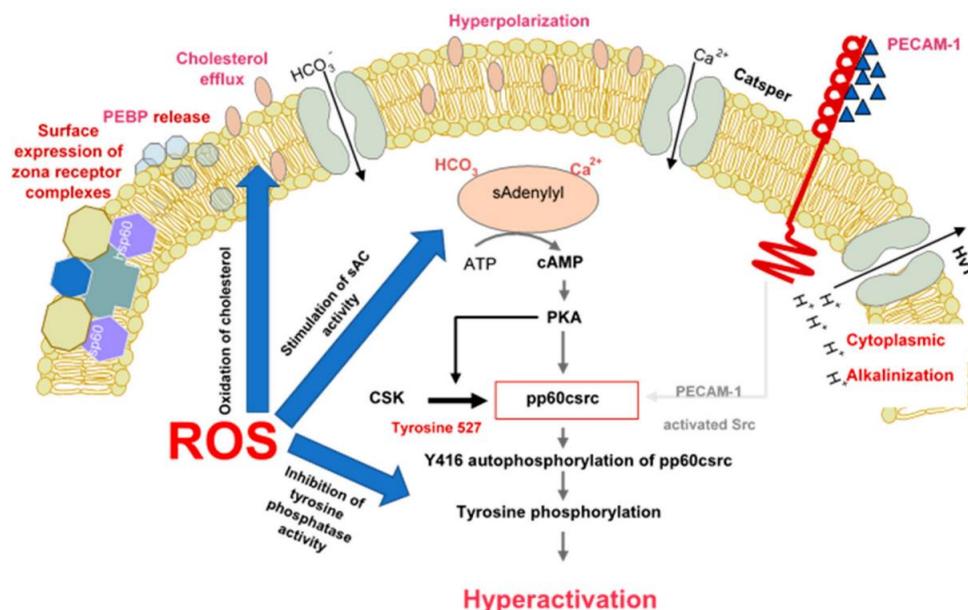


Figure 11 : Rôle des ROS dans l'induction de la capacitation et hyperactivation du spermatozoïde<sup>52</sup>. Les « espèces réactives à l'oxygène » ou éléments oxydants, participent à l'initiation de la cascade conduisant à la capacitation et à l'hyperactivation du spermatozoïde via l'oxydation du cholestérol (entraînant leur exclusion de la membrane plasmique) et la stimulation de la sAdenylyl (produisant l'AMPC nécessaire au reste de la cascade).

Au moment de la rencontre entre les gamètes mâle et femelle, la liaison du spermatozoïde par la zone apicale de l'acrosome à la zone pellucide de l'ovocyte initie la fusion des membranes interne et externe de l'acrosome, libérant ainsi le contenu enzymatique acrosomique<sup>53</sup>. Ces actions permettent « l'ouverture » d'un passage pour le spermatozoïde au travers de la zone pellucide et l'exposition de protéines particulières de la zone équatoriale de la tête du spermatozoïde, telles que l'équatorine, CD9, CD81 ou IZUMO1<sup>53</sup>. Isolément, ces protéines ne sont pas indispensables pour le succès de la fécondation, même si la fertilité est réduite chez la souris femelle KO pour les récepteurs CD9 ou CD81. En revanche, un double KO pour ces récepteurs conduit à une stérilité<sup>53</sup> suggérant un processus de compensation entre ces différents récepteurs. L'absence de la protéine IZUMO1 à la surface du spermatozoïde ou celle de son récepteur à la surface de l'ovocyte, JUNO, induit une stérilité chez la souris. Le rôle de ces protéines n'est pas totalement élucidé, mais les hypothèses avancées chez le bovin font état d'une implication dans l'adhésion des membranes plasmiques des deux gamètes et l'entrée de la tête du spermatozoïde dans l'ovocyte<sup>53</sup>. Dès pénétration d'un spermatozoïde, l'ovocyte développe des processus moléculaires le rendant hermétique à tout autre spermatozoïde évitant ainsi la polyspermie. Une fois libéré dans le cytoplasme ovocytaire, le pronucléus mâle (patrimoine génétique paternel) se décondense et participe pour moitié à la formation du patrimoine génétique du zygote puis de l'embryon. L'activation du génome embryonnaire s'effectue progressivement et plus ou moins rapidement en fonction des espèces pour piloter le développement embryonnaire (au stade 8-16 cellules chez le bovin).

Chez le bovin, la première étape du développement embryonnaire démarre 2 jours après la fécondation par l'initiation de la segmentation, suivie de divisions cellulaires successives. Au stade morula (32-64 cellules), 5 jours après la fécondation, la première différenciation cellulaire intervient : les cellules trophoblastiques se différencient à la périphérie de l'embryon et seront à l'origine du placenta, alors que les cellules dites de la masse interne seront à la base de toute l'organogenèse embryonnaire. La formation d'une cavité dite « blastocèle » par compaction des cellules, complète la formation du blastocyste (J6-J7 après fécondation)<sup>54</sup>. Après 9 jours, l'éclosion du blastocyste survient par la rupture de la zone pellucide. À partir du 11<sup>e</sup> jour, le blastocyste s'allonge et commence à occuper l'espace utérin.

L'implantation de l'embryon se déroule du 21<sup>e</sup> au 35<sup>e</sup> jour post-fécondation. La mise en place de toutes les structures placentaires permettant les communications (« échanges de nutriments et gazeux) entre la mère et le fœtus, s'achève au 60<sup>e</sup> jour de gestation. En parallèle de la « nidification ou implantation », l'organogenèse débute par la différenciation des trois feuilletts embryonnaires (endoderme, mésoderme et ectoderme) au sein de l'embryon et s'achève au 42<sup>e</sup> jour après fécondation. S'ensuit principalement une phase de croissance des organes jusqu'à la naissance. La durée de gestation chez les bovins varie en fonction de la race, entre 282 (Holstein) et 294 (Blonde d'Aquitaine) jours.

### 1.2) De la production à la mesure de la fertilité : comment sont utilisés les taureaux et leurs gamètes en France ?

L'amélioration de la génétique bovine française repose sur des schémas de sélection mis en place par les entreprises de sélection. Ces schémas comportent plusieurs étapes, qui ont évolué au gré des améliorations techniques et méthodologiques : la procréation d'un grand nombre de mâles candidats à partir des taureaux élites (les « pères à taureaux ») et des meilleures femelles (les « mères à taureaux »), la sélection des meilleurs candidats selon divers critères (morphologie, production, reproduction, santé...), le recrutement des candidats et leur élevage en station jusqu'à leur puberté, le contrôle de la fonction sexuelle afin de déceler et d'éliminer les taureaux présentant des problèmes inhérents à la production ou à la qualité de la semence, l'entrée en centre de production et l'entraînement à la collecte de semence, la collecte et le conditionnement de la semence pour une diffusion large par insémination et enfin la mesure des performances des descendants pour alimenter en données les modèles d'évaluation génétique.

#### 1.2.1) *Du testage sur descendance à la sélection génomique*

À la mise en place de l'évaluation génétique par INRAE avec la loi sur l'élevage de 1966, l'évaluation reposait sur un « testage sur descendance » : chaque taureau était collecté pour réaliser un nombre limité d'inséminations, afin de procréer une centaine de filles dont les performances étaient mesurées pour permettre en retour l'évaluation de leur père. Durant les 3 à 5 ans d'attente nécessaires à cette évaluation, les taureaux étaient entretenus en

station d'élevage (phase de « lay off »). À l'issue de cette phase, seuls les meilleurs taureaux étaient conservés et reprenaient la production de semence pour être diffusés à grande échelle. Cette approche était donc extrêmement coûteuse. Elle permettait par ailleurs une évaluation de la fertilité des taureaux avant diffusion.

Un changement de paradigme a eu lieu en 2009 avec l'arrivée de la sélection génomique. D'abord réservée aux 3 grandes races laitières (Holstein, Montbéliarde et Normande), cette nouvelle méthode a progressivement été étendue aux autres races (à l'exception des races à faibles effectifs) et aux ovins et caprins. Elle s'appuie sur une analyse du polymorphisme du génome à l'aide de puce de génotypage, permettant d'interroger de 10.000 à 50.000 positions réparties sur le génome (« *Single Nucleotide Polymorphisme* » ou SNP, polymorphisme d'une base). Ces SNP, rarement causaux, peuvent être en déséquilibre de liaison avec les gènes impliqués dans les caractères d'intérêt. Il est donc possible d'estimer l'effet de chacun des allèles de ces SNP sur les caractères d'intérêt en utilisant une large population de référence d'animaux génotypes et phénotypés pour ces caractères (plusieurs milliers à dizaines de milliers d'animaux). Et par extension, il est possible d'établir des équations de prédiction combinant les SNP d'intérêt pour estimer la valeur génétique de tout animal génotypé, en particulier les jeunes mâles candidats, au stade embryonnaire ou à la naissance. Il n'est ainsi plus nécessaire de procréer une descendance et d'attendre d'en mesurer les performances pour diffuser la semence des taureaux, ce qui permet de réduire par deux l'intervalle entre les générations<sup>55</sup> et donc de doubler le progrès génétique annuel, tout en réduisant le coût de fonctionnement des schémas de sélection. Mais, revers de la médaille, la production des jeunes taureaux est moins importante en volume et la semence est diffusée sans information préalable sur sa fécondance et la fertilité globale des taureaux.

### *1.2.2) La production de semence en centre de collecte*

La fréquence de collecte de la semence dépend de l'âge, de la race et des pratiques du centre (2 à 3 fois par semaine). Une fois collecté, l'éjaculat est soumis à un premier contrôle qualité : la mesure du volume de l'éjaculat, de sa concentration en spermatozoïdes et de la motilité

massale. Ce dernier examen consiste en l'estimation, sous un microscope, du niveau de vitalité et de vivacité de la semence. Des seuils ont été établis selon les paramètres, en dessous desquels les éjaculats ne sont pas conservés. Les éjaculats retenus sont dilués (de 10 à 30 fois) dans un milieu de cryoconservation et incubés à 4°C pendant 4 heures pour équilibrer les échanges entre milieu et spermatozoïdes, avant mise en paillette, à raison de 2 à 25 millions de spermatozoïdes par dose est réalisée (10 à 125 millions/ml). Suivant l'âge du taureau et ses caractéristiques de production, un éjaculat peut fournir entre 80 et 600 doses d'IA. Les paillettes sont plongées dans les vapeurs d'azote avant d'être stockées dans l'azote liquide. Un dernier contrôle qualité sur chaque lot de paillettes est réalisé après décongélation. Ce contrôle permet de vérifier la résistance de la semence aux processus de congélation/décongélation. En effet, ce processus inflige de tels dommages, qu'environ 50% des spermatozoïdes « meurent » à la décongélation et que les survivants sont moins performants<sup>56,57</sup>. En dessous d'un seuil de spermatozoïdes vivants et progressifs fixé par le centre, les éjaculats sont éliminés. Les autres vont être distribués aux inséminateurs et diffusés sur le terrain (figure 12).

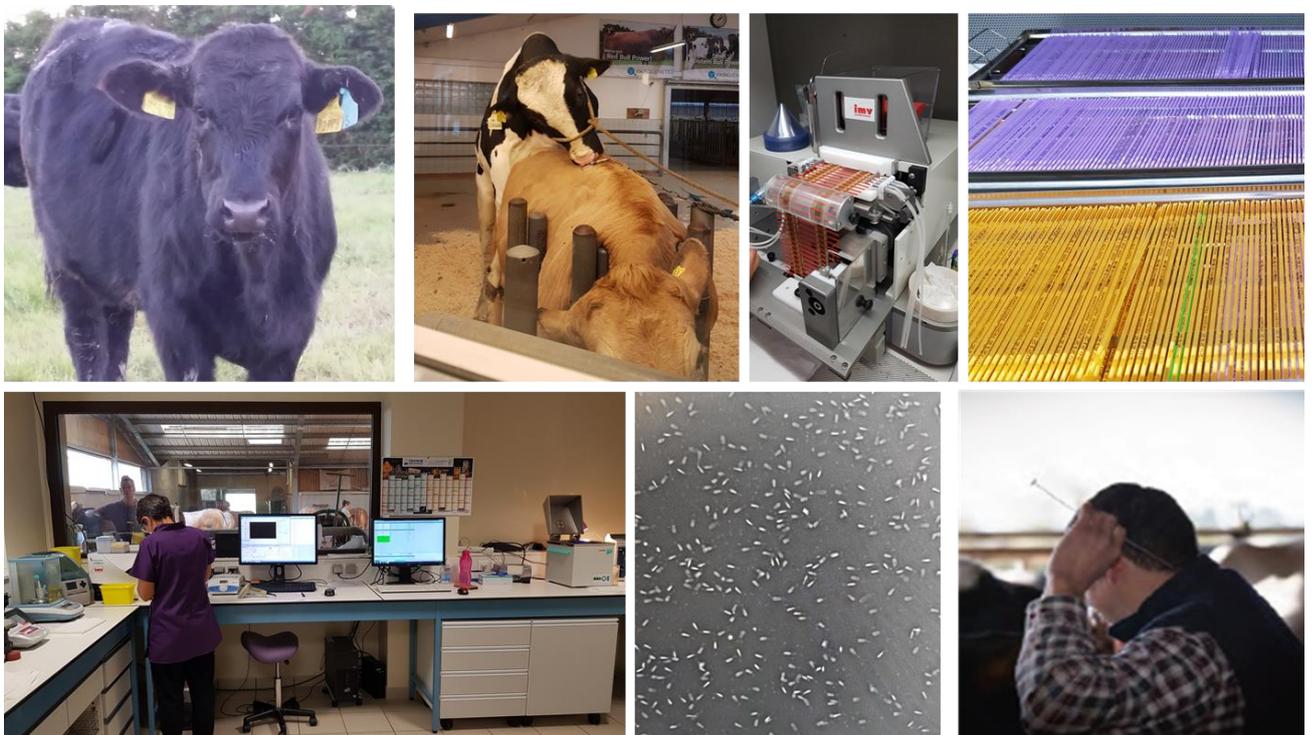


Figure 12 : De la récolte à l'insémination.

Les différentes étapes de la récolte à l'insémination sont représentées ici en images. Après la récolte de l'éjaculat au vagin artificiel, la mobilité massale ainsi que les données de production (volume et concentration) sont mesurées. En parallèle l'éjaculat est dilué avec le milieu de congélation et maintenu à 4°C pendant 4 heures. La semence est alors fractionnée en de nombreuses doses d'insémination avant d'être plongée dans l'azote liquide.

Un contrôle post-décongélation de motilité individuel est réalisé. Si tous les critères sont remplis, l'éjaculat peut être distribué aux inséminateurs et suivant la demande, inséminé. ©Alice

De nombreuses recherches sont menées pour améliorer la qualité de la semence post-décongélation et son contrôle<sup>58-63</sup> : plus de 5000 articles contiennent comme mots clefs « amélioration de la qualité de la semence ». Les différentes stratégies employées se résument à l'ajout de molécules protectrices (cystéine, glutathion...) ou stimulantes (caféine, penthoxyfiline...) dans le milieu de congélation. D'autres stratégies emploient des compléments alimentaires pour modifier la composition lipidique de la membrane des spermatozoïdes (huile de lin, DHEA...). Le principe est de fluidifier la membrane cytoplasmique par l'incorporation de phospholipides polyinsaturés, la rendant plus résistante à différents stress. De même, divers paramètres sont évalués par les centres de production de semence. Ces paramètres, tout d'abord simples, ont suivi l'évolution technologique pour devenir de plus en plus complexes et précis (figure 13).



Figure 13 : Évolution technologique du contrôle qualité de la semence.

Le contrôle qualité de la semence réalisé lors de la collecte, consiste en la mesure du volume, de la concentration et de la motilité massale de l'éjaculat. Celui réalisé après la décongélation peut être beaucoup plus complexe (motilité individuelle, analyses fonctionnelles...). L'évolution des pratiques a suivi l'évolution technologique : de l'estimation à l'œil des caractéristiques de la semence, le contrôle qualité en centre de production a peu à peu évolué en adoptant la microscopie classique, puis celle assistée par ordinateur (système CASA) et pour certains, la cytométrie en flux. Les technologies plus avancées, comme la spectrométrie de masse, sont encore réservées aux laboratoires de recherches.

Ces dernières décennies, la production de semence a également connu son lot d'innovations techniques et technologiques. La possibilité d'obtenir un taux de gestation acceptable après insémination avec de la semence congelée a permis de repenser totalement la logistique de

stockage de la semence (technique de congélation, composition des milieux de cryoconservation), le maillage du territoire national et la distribution des doses à l'international. Le développement de nouvelles technologies automatisées a permis d'aborder le contrôle qualité de la semence de façon plus précise et plus objective. Enfin, l'une des dernières innovations réside dans le tri des spermatozoïdes en fonction de la présence d'un chromosome X ou Y, afin d'assurer à plus de 90%, le sexe du descendant lors d'une IA utilisant la semence sexée.

### *1.2.3) Les indicateurs de la fertilité mâle*

L'efficacité reproductive conditionnant une part importante du revenu des élevages (mais variable selon les systèmes d'élevage), les coopératives d'insémination et leurs unions, entreprises de sélection, attachent une attention particulière à la fertilité de la semence qu'elles diffusent. Pour cela, divers indicateurs sont calculés tout au long de la vie du taureau, basés sur les résultats d'insémination.

Le Taux de Non-Retour (TNR) est un des indicateurs les plus utilisés. Il correspond au pourcentage d'inséminations fécondantes et peut être calculé à différentes périodes, suite à un constat de gestation : TNR 24, 56, 90 et 282 jours après l'insémination. Le TNR24 est basé sur l'observation du non-retour en chaleur des vaches, éventuellement complétée par un dosage de progestérone sérique (possible dès 21 jours post-IA). Les TNR 56 ou 90 sont également déterminés par l'observation du non-retour en chaleur, complétée éventuellement par un diagnostic de gestation échographique. Le TNR 282 correspond au taux de mises bas.

Sachant que la réussite à l'IA dépend de plusieurs facteurs (figure 14), intrinsèques (qualité des gamètes mâles et femelles) et extrinsèques (conditions d'élevage, état physiologique de la vache, détection adéquate des chaleurs, technicité de l'inséminateur...), le TNR doit être corrigé d'un certain nombre d'effets. Si le nombre d'inséminations prises en compte est suffisant (plusieurs centaines voire milliers), cet indicateur permet de s'affranchir des effets « femelle ». Toutefois, il reste soumis à de forts aléas techniques qu'il est nécessaire de

prendre en compte (effets inséminateur, saison, mois, jour de la semaine, région...) par diverses approches statistiques.

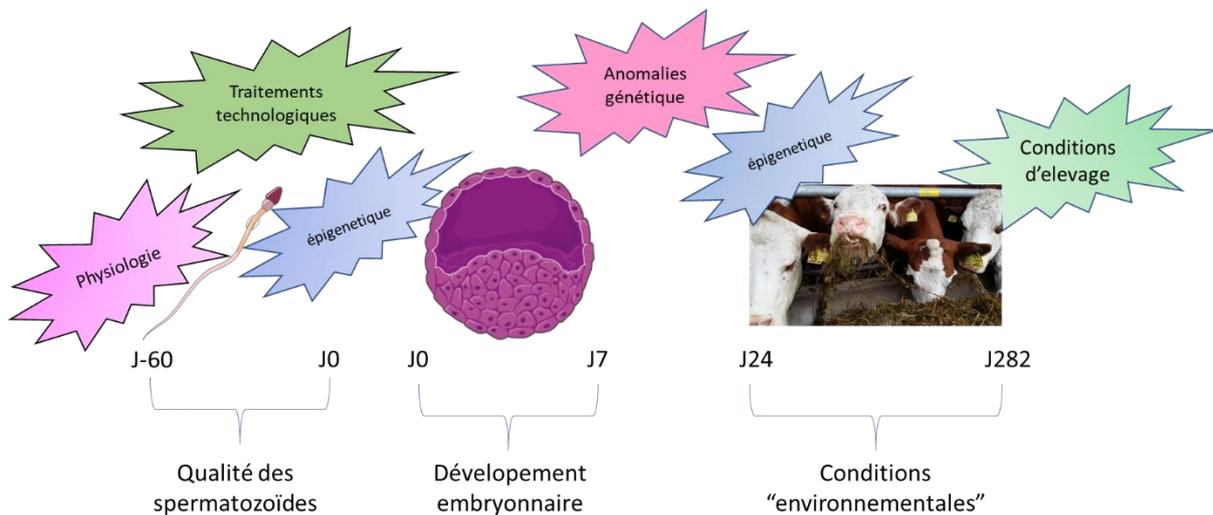


Figure 14 : Le résultat de l'IA est multifactoriel.

La réussite de l'insémination se mesure finalement par le taux de mise bas. De multiples facteurs sont impliqués dans la réussite de l'insémination. Une perturbation de l'un d'entre eux, que ce soit au niveau « cellulaire » (spermatozoïde, ovocyte ou embryon) ou au niveau plus général de l'organisme (environnement dans lequel évolue la vache gestante), peut conduire à un échec de l'IA ©Alice.

Il est à noter que ces 4 TNRs sont susceptibles de renseigner sur les causes de l'échec : le TRN24 traduirait un défaut de fécondance en lien avec la physiologie et les caractéristiques du spermatozoïde, ou une forte mortalité embryonnaire précoce liée à une anomalie génétique par exemple. Le TNR 56 signifierait plus un défaut d'implantation, d'origine embryonnaire ou maternel, le TRN90 un défaut d'organogenèse et le TNR282 un défaut de développement foetal éventuellement en lien avec les conditions d'élevage ou une anomalie génétique. Ainsi, le TNR à 24 jours serait le plus lié au pouvoir fécondant de la semence et serait le plus pertinent, mais il s'agit aussi du taux le moins précis du fait des difficultés de détection de la gestation à ce stade et des délais de remontées d'information dans les bases de données nationales.

La mise en place du code-barre sur les paillettes et sa lecture lors de l'IA, permet aujourd'hui de lier les résultats du TNR non seulement à un taureau, mais aussi à un éjaculat spécifique. Les avantages qu'apporte cette nouvelle technologie, notamment pour la traçabilité des lots (sanitaire), ont contribué à son utilisation importante sur le terrain. Dans certaines races,

comme la Holstein, plus de 90% des IA réalisées en France peuvent être retracées au niveau éjaculat. Mais le nombre de doses réalisées par éjaculat étant faible (de 80 et 600 selon les taureaux et les centres), la précision du TNR éjaculat est limitée. Le TNR est donc un indicateur pratique, mais à prendre en compte avec précaution compte tenu de sa précision, surtout s'il s'agit de prendre une décision de réforme anticipée pour cause de fertilité trop faible.

Par ailleurs, l'évaluation génétique femelle en France intègre un effet fixe Taureau x Campagne, qui traduit l'effet du taureau sur la fertilité femelle en année « n », correction faite des autres effets (inséminateur, jour de la semaine, région...). Il constitue ainsi un indicateur global de fertilité, exprimé relativement à la moyenne de la population (variable centrée-réduite). Les valeurs négatives correspondent donc à des taureaux ayant une fertilité plus basse que la moyenne et inversement pour les valeurs positives.

Comme indiqué précédemment, l'adoption de la génomique ne permet plus, même à petite échelle comme lors du testage, de bénéficier de données de TNR avant diffusion à large échelle de la semence. Dans un monde génomique, il est certes envisageable de calculer un index fertilité mâle pouvant servir à écarter les candidats ayant une fertilité jugée trop faible. Mais, outre la faible héritabilité ( $h^2 < 10\%$ ) qui limite la précision de l'évaluation, une telle approche ne prend en compte que les effets génétiques, oubliant tous les effets environnementaux susceptibles de dégrader la fertilité d'un animal ayant pourtant un bon potentiel génétique.

## 2) La prédiction de la fertilité mâle

Les indicateurs décrits précédemment permettent d'évaluer la fertilité a posteriori. Mais une gestion optimale de la production nécessiterait un outil permettant de prédire la fertilité d'un mâle, afin soit de l'éliminer pour recruter un autre candidat ayant un index génomique équivalent, mais une meilleure fertilité, soit d'adopter une gestion particulière de ce mâle pour pallier ses problèmes de fertilité. De même en ce qui concerne le contrôle qualité des éjaculats, une prédiction a priori de la fertilité permettrait de retirer les éjaculats présentant des défauts de fertilité avant toute diffusion. Par ailleurs, dans un contexte R&D, un tel outil

permettrait d'évaluer plus simplement l'impact de nouveaux dilueurs ou de nouvelles stratégies de conditionnement de la semence etc...

Si de nombreuses études portent sur l'amélioration des paramètres spermatiques post-décongélation, les études explorant réellement la prédiction de la fertilité sont moins courantes. Elles reposent sur le dogme, toujours en vigueur aujourd'hui, qu'une perte, même minime, de la qualité ou de la fonctionnalité d'un des compartiments cellulaires du spermatozoïde, peut impacter le pouvoir fécondant de la semence et entraîner une réduction de fertilité<sup>64</sup>. Dès lors, de nombreuses équipes ont travaillé au développement de tests permettant l'évaluation de chaque compartiment du spermatozoïde<sup>65-67</sup> : intégrité de l'acrosome, intégrité et fluidité de la membrane plasmique, niveau de fragmentation ou de décondensation de la chromatine, sensibilité à l'oxydation, développement énergétique, vitesse de déplacement). Certains de ces paramètres mesurés ont été significativement reliés à la fertilité, mais pris individuellement, ils présentaient un faible pouvoir prédictif de la fertilité. Le concept d'une approche multifactorielle de la prédiction de la fertilité a donc été proposé : la fertilité étant multifactorielle, la prédiction doit l'être également.

De nouvelles approches ont donc été développées pour combiner de façon optimale les différents tests et élaborer le meilleur prédicteur possible. Au-delà de l'aspect scientifique, la recherche d'un outil valorisable pour les professionnels a été le moteur de la plupart de ces études. Il faut donc trouver le meilleur compromis entre le pouvoir prédictif du modèle, lié au nombre de phénotypes à mesurer et la facilité d'utilisation sur le terrain. Le coût de fonctionnement est également un critère majeur.

Les résultats obtenus depuis les 5 dernières années laissent à penser que les aspects de fonctionnalité ou de qualité cellulaire ne suffiront pas à expliquer toute la variabilité de la fertilité des bovins. En effet, les meilleurs prédicteurs, obtenus par différentes équipes sur différentes cohortes de taureaux, n'expliquent que 40% environ de la variabilité de fertilité<sup>64,68</sup> (figure 15), ce qui est loin d'être suffisant pour une utilisation pratique. Il manque vraisemblablement des informations pour atteindre une bonne prédiction et d'autres paramètres biologiques doivent donc être pris en compte pour mieux expliquer les variations de fertilité. Les informations portées par le génome<sup>69</sup>, comme les mutations touchant un gène du développement embryonnaire<sup>70</sup>, ou tout ce qui se rapporte à l'épigénétique<sup>71</sup> pourraient être intéressants à intégrer dans le prédicteur.

Une liste non exhaustive des différents paramètres évalués est exposée dans les paragraphes suivants.

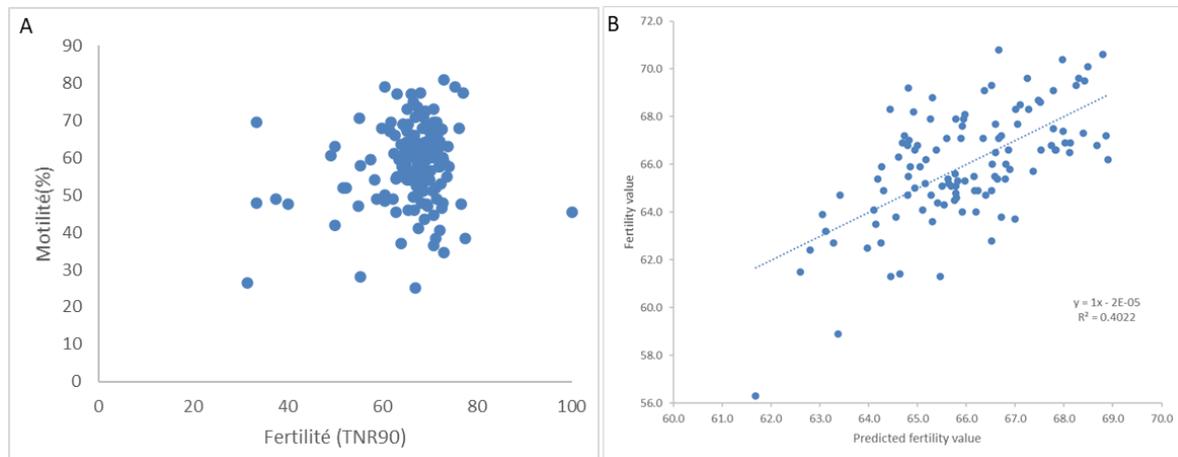


Figure 15 : Prédiction de la fertilité du taureau avec les données de fonctionnalité de la semence.

(A) La relation entre la motilité (%) et la fertilité est très faible. Il faut combiner plusieurs facteurs pour mieux prédire la fertilité. (B) Relation entre la fertilité prédite (combinatoire de plusieurs phénotypes) et le taux de non-retour.<sup>64</sup> ©Alice

## 2.1) Les paramètres historiques

Les premières technologies mises en place par les centres et encore utilisées aujourd'hui sont basées sur l'utilisation d'un microscope à platine chauffante afin d'évaluer la motilité des spermatozoïdes et d'un spectrophotomètre afin de déterminer la concentration cellulaire. La motilité des spermatozoïdes se mesure soit de façon globale, on parle alors de motilité massale, soit de façon individuelle (figure 16). La motilité massale est évaluée par une note de 0 à 5, représentant l'intensité des « vagues » formées par le mouvement de milliards de spermatozoïdes. La motilité individuelle, réalisée sur une dilution de la semence, permet d'estimer un pourcentage de cellules motiles et une intensité de mouvement (progressivité), notée également de 0 à 5. La motilité des spermatozoïdes bovins (pré ou post-congélation) a longtemps été considérée comme un paramètre crucial, expliquant la variabilité de la fertilité observée. En effet, de nombreuses études ont révélé une relation significative avec la fertilité<sup>64,72</sup>. En revanche même si statistiquement cette relation existe, elle reste très faible. Elle permet effectivement d'éliminer des éjaculats présentant une motilité extrêmement

basse et un pouvoir fécondant très faible. Mais au-delà d'un seuil de motilité relativement faible (10 à 20%), ce paramètre n'est plus pertinent pour prédire la fertilité.

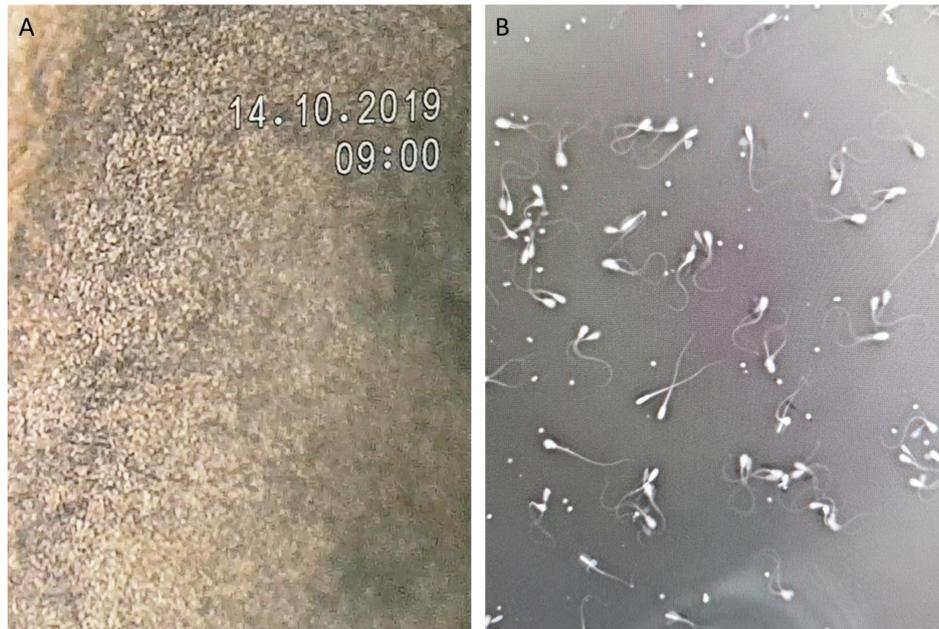


Figure 16 : Analyse de la motilité de la semence.

(A) Motilité massale : l'opérateur apprécie les vitesses de déplacement des « vagues » de spermatozoïdes. (B) Motilité individuelle : l'opérateur estime le pourcentage de spermatozoïdes mobiles et le type de mouvement. @ Alice

## 2.2) Les paramètres « cellulaires » via l'essor de nouvelles technologies

De nouvelles solutions techniques ont progressivement été proposées par les constructeurs, permettant d'analyser de nouveaux phénotypes cellulaires. Les systèmes « CASA » (*computer assisted semen analysis*), microscopes assistés par ordinateur, ont permis d'évaluer automatiquement le pourcentage de spermatozoïdes motiles, ainsi que leur vitesse de déplacement, le type de mouvement et pour les nouvelles générations, certaines anomalies morphologiques portées par le gamète (figure 17), permettant une évaluation de la qualité de la semence plus précise, rapide et objective.

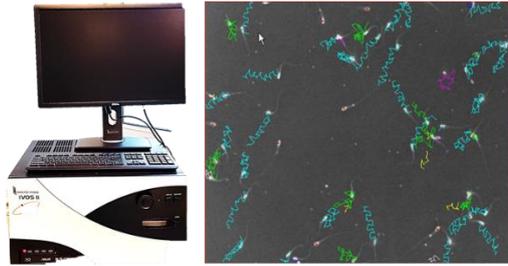


Figure 17 : Analyse de la motilité des spermatozoïdes avec la technologie « CASA ». L'IVOS II (Hamilton Thorne) remplace l'œil d'un technicien par une caméra. Le logiciel détecte les spermatozoïdes par contraste et les photographie à 60 images par seconde. Les trajectoires et vitesses de déplacement pour les gamètes mobiles peuvent donc être déterminées. @Alicia

La simplification de l'utilisation des cytomètres en flux et la réduction de taille des appareils ont permis l'implémentation de certains tests en centre de production de semence, permettant d'étudier avec précision et objectivité des caractéristiques fonctionnelles comme la viabilité, l'intégrité des membranes cytoplasmique et acrosomique, la fluidité membranaire, la réaction acrosomique, l'état de condensation et de fragmentation de la chromatine, le niveau de production énergétique (mitochondrie), la sensibilité à l'oxydation, la présence ou absence de certaines protéines membranaires, le flux calcique (figure 18)... Tout comme pour la motilité, les corrélations entre ces paramètres et la fertilité, bien que significatives, restent faibles. Ce n'est que par la combinaison de l'ensemble de ces paramètres que le prédicteur devient plus efficace ( $R^2 \sim 0.40$ )<sup>64</sup>. Le fait d'associer plusieurs phénotypes différents complexifie l'obtention du prédicteur et augmente son coût, limitant son intérêt en routine. C'est pour cela que certaines équipes cherchent à combiner les mesures cytométriques en un seul protocole<sup>73,74</sup>.

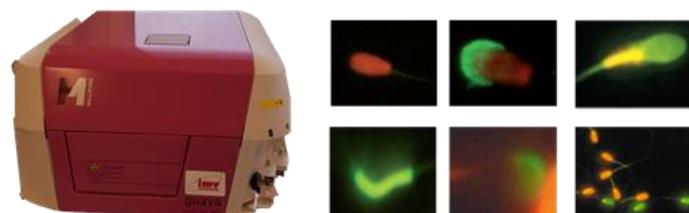


Figure 18 : Analyse de la fonctionnalité des spermatozoïdes avec le cytomètre de flux. Basés sur des marquages fluorescents, différents protocoles ont été établis pour l'analyse de la viabilité, de l'intégrité de l'acrosome, du développement énergétique, du niveau d'oxydation cellulaire, de la fragmentation ou de la décondensation de l'ADN, des anomalies morphologiques...@Alicia

Au-delà de la prédiction de fertilité, la mesure de ces phénotypes assure une qualification objective des spermatozoïdes et des éjaculats. Dans le milieu professionnel, de plus en plus d'opérateurs demandent lors d'import-export de semences, un contrôle qualité des éjaculats incluant ces nouveaux phénotypes. Cette démarche se généralise en France, mais aussi au niveau européen. Cette situation a ouvert la voie à une nouvelle collaboration européenne multicentrique via le groupe « QualiVets ». Il s'agit d'un regroupement de centres de production de semence et de centres de recherche de plusieurs pays européens (Irlande, Hollande, Belgique, Suisse, Danemark et France). La première conclusion émanant de ce groupe a porté sur la nécessité de standardiser les protocoles de mesure de ces phénotypes, avant de pouvoir les regrouper dans une base de données commune pour améliorer les équations de prédiction. Car même s'il s'agissait des mêmes phénotypes mesurés, différents protocoles de mesure étaient employés, introduisant une variabilité technique importante entre les centres. Des travaux visant l'établissement d'un guide des bonnes pratiques sont en cours.

### 2.3) Les nouveaux paramètres moléculaires

Ces nouveaux « phénotypes » sont associés à l'essor des nouvelles technologies à haut débit (séquenceur nouvelle génération, spectrométrie de masse, RMN). Ils sont pour la plupart réservés aux laboratoires de recherche. L'approche est encore plus fine que pour les phénotypes cellulaires, car elle aborde la prédiction de la fertilité via une analyse moléculaire du spermatozoïde et du plasma séminal : analyse du génome (génotype : SNP et épigénome : cytosine méthylée<sup>75</sup>), du transcriptome via l'analyse des ARN (codants et non codants<sup>76</sup>), du protéome<sup>77</sup>, du lipidome et du glycome (identification ou spectre)... Ces nouvelles technologies génèrent « facilement » de grandes quantités de données. La difficulté (qui est d'autant plus vraie dans la prédiction de fertilité) est l'intégration de toutes ces données hétérogènes dans un modèle le plus parcimonieux possible et le plus prédictif possible.

Après avoir consacré de nombreuses années à l'analyse et à la combinaison des phénotypes cellulaires, nous avons décidé d'explorer dans le cadre du laboratoire commun SeQuaMol (ALLICE et INRAE/BREED/MECP<sup>2</sup>), le rôle des mécanismes épigénétiques sur le développement du spermatozoïde et leur impact sur le pouvoir fécondant de la semence.

## Chapitre II Épigénétique et fertilité mâle

Toutes les cellules d'un même organisme possèdent exactement le même patrimoine génétique, mais tous les organismes vivants complexes possèdent différents organes nécessitant des structures cellulaires différentes aux fonctions spécifiques. Il est donc nécessaire de sélectionner au sein de chaque type cellulaire l'information génétique conduisant à la différenciation cellulaire et au métabolisme approprié. Comme le proposait Conrad Waddington<sup>78</sup> dès 1942, des mécanismes biologiques doivent orchestrer « la partition écrite » par les gènes et produire une interprétation aboutissant à un phénotype spécifique du type cellulaire. Parmi ces mécanismes moléculaires, les mécanismes dits « épigénétiques » jouent un rôle majeur. Étymologiquement le mot épigénétique signifie « au-dessus des gènes » et fait référence au contrôle exercé par ces mécanismes sur le fonctionnement du génome et en particulier sur l'expression des gènes. Ces éléments jouent un rôle majeur dans la destinée cellulaire : maintien de l'état indifférencié ou différenciation induite de manière spatio-temporelle. Ces marques épigénétiques ne modifient pas la séquence génomique, mais elles sont tout de même héréditaires d'une cellule à l'autre au cours de la mitose voire dans certains cas de la méiose. Les marques épigénétiques sont aussi reconnues comme un patrimoine non génétique transmis à la descendance qu'il s'agisse de l'empreinte parentale (concernant une petite centaine de gènes) ou d'une « mémoire » de l'historique de vie des parents. Les mécanismes épigénétiques participent à l'architecture de la chromatine en définissant de grands domaines au sein de l'ADN nucléaire : l'euchromatine et l'hétérochromatine, se différenciant par leur activité transcriptionnelle, respectivement active et silencieuse. Le contrôle de l'expression génique peut s'exercer selon deux grands types de régulation (figure 19) : la régulation de la transcription des gènes (via la méthylation de l'ADN et les modifications post-traductionnelles des histones) et la régulation de la traduction des ARN messagers (via les petits ARN non codants). Ces mécanismes (décrits plus en détail dans les paragraphes suivants) agissent de façon permissive ou répressive, de concert ou en compétition, basée sur un véritable dialogue modulant l'efficacité propre aux différents systèmes.

Les différents mécanismes épigénétiques impliqués dans le cas particulier du spermatozoïde sont décrits dans les chapitres suivants.

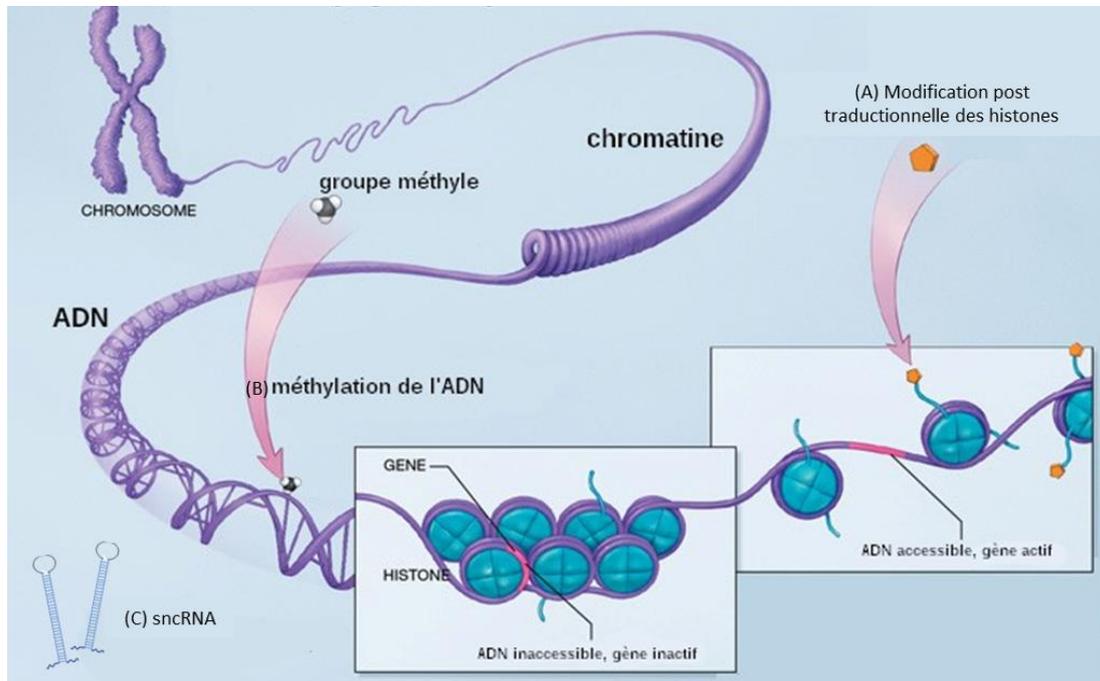


Figure 19 : Les principaux acteurs des mécanismes épigénétiques.

(A) Modifications post-traductionnelles des histones et (B) la méthylation de l'ADN, contrôlant l'accessibilité locale à l'ADN et la transcription des gènes. (C) les petits ARN non codants (sncRNA), régulant la traduction des ARN messagers.

## 1) Des modifications post-traductionnelles des histones...aux protamines

### 1.1) Rôles et fonctionnement biologique des histones

Dans le noyau cellulaire, l'ADN s'enroule autour de structures appelées « nucléosomes », constituées par un assemblage de protéines basiques, les histones. Les nucléosomes sont formés par 4 paires de variants d'histone (H2a, H2b, H3 et H4) et leur structure est maintenue par la liaison d'une molécule d'histone H1, permettant l'enroulement de 147 paires de bases d'ADN). Cette structure permet de condenser l'ADN et d'atteindre un niveau de compaction maximum des chromosomes. La structure que prend l'ADN compacté s'appelle la chromatine.

Selon l'état de la chromatine, plus ou moins condensé, les facteurs de transcription ont la capacité d'atteindre leur cible et de jouer leur rôle. Les histones participent donc à la régulation de la transcription via « l'ouverture ou la fermeture » locale des régions de l'ADN. Ces changements de compaction sont possibles via les modifications post-traductionnelles des histones, portées par certains acides aminés. Les modifications chimiques sont de différentes natures<sup>79</sup> : méthylation, acétylation, phosphorylation, glycosylation, carbonylation, ubiquitylation, biotinylation. Elles peuvent intervenir en combinaison et participent globalement à des modifications notables des capacités d'interactions entre les histones nucléosomales et l'ADN. Une telle complexité a été référencée sous le terme de « code des histones », et concourt donc à définir l'état de la chromatine. Par exemple, la méthylation des lysines 4, 36 et 79 de l'histone H3 est synonyme d'ouverture de la chromatine et donc d'activation de la transcription. Au contraire, la méthylation des lysines 9 et 27 de la même histone H3 est synonyme de fermeture de la chromatine et donc du silence transcriptionnel<sup>79</sup>.

Dans le cas des spermatozoïdes, certaines histones canoniques sont remplacées par de variants spécifiques dans les stades précoces de la spermatogenèse<sup>80</sup>, afin de participer à la régulation de l'expression génique essentielle à l'établissement de la lignée mâle. C'est le cas des histones TH2A et TH2B, variants spécifiques spermatiques, qui remplacent leurs histones canoniques H2A et H2B, dans les spermatogonies. À des stades plus avancés de la spermatogenèse, ces variants d'histones sont remplacés par d'autres protéines, les protamines.

## 1.2) Des histones aux protamines dans le spermatozoïde

Le noyau du spermatozoïde présente une organisation de la chromatine hautement spécifique, totalement différente de celle d'une cellule somatique. Durant les dernières étapes de la spermatogenèse, les variants d'histones sont massivement remplacés par des protamines. Ce sont des protéines basiques, plus petites que les histones et qui participent à la formation de structures extrêmement compactées, contenant de 50 à 100 kb, appelées toroïdes. Ces structures sont dix fois plus condensées que celles formant les nucléosomes<sup>81</sup>. Deux isoformes de protamines sont présentes dans les spermatozoïdes (Prm1 et Prm2). Selon

les espèces, le ratio entre les deux isoformes change de façon importante. La Prm2 peut être absente (chat domestique) ou représenter plus de 77% des protamines (macaque)<sup>82</sup>. Chez le bovin, la Prm2 a longtemps été considérée comme inexsistante<sup>82</sup>, mais les travaux publiés par Hamilton et son équipe<sup>83</sup> en 2019, ont permis de démontrer sa présence en faible quantité (100 fois moins que la Prm1). Tout comme les histones, les protamines peuvent subir des modifications post-traductionnelles. Il s'agit principalement de méthylation, phosphorylation et acétylation (données recueillies chez l'homme et la souris<sup>81</sup>). Le rôle de ces modifications sur les protamines n'est pas connu. Elles seraient impliquées dans la phase de substitution des histones par les protamines au cours de la spermatogenèse et dans l'ouverture locale de la chromatine pour leur élimination lors de la phase de reprogrammation épigénétique du noyau embryonnaire.

Le remplacement par les protamines n'est pas total et il reste une proportion d'histones résiduelles, variable selon les espèces : de 2% chez le spermatozoïde murin à 5 à 10% chez l'homme<sup>82</sup>. Chez le bovin, à notre connaissance, aucune donnée n'a été publiée sur le pourcentage d'histones résiduelles par rapport aux protamines. Chez la souris, ces histones seraient présentes sur les brins d'ADN reliant deux toroïdes adjacents<sup>82</sup>. Il est aussi montré un enrichissement en histones au niveau des régions répétées, telles que les centromères chez le bovin<sup>84</sup>. Une approche « Gene ontologie » réalisée sur la liste des gènes localisés dans les régions génomiques associées aux histones résiduelles, met en avant un seul processus KEGG statistiquement enrichi : cGMP-PKG. Il s'agit de gènes codants des sérine/thréonine protéines kinase qui phosphorylent leurs cibles sous une activation de cGMP<sup>84</sup>.

### 1.3) Impact sur la fertilité

Des défauts d'expression des différents variants d'histones induisent d'importants problèmes de fertilité. Les désordres interviennent soit au cours de la spermatogenèse, avec un défaut de condensation de la chromatine (souris KO H1t2, TH2A, TH2B, H2A.L1&2, H3.3)<sup>85</sup>, soit lors du développement embryonnaire (souris KO CENP-A ou H2A.Z)<sup>81</sup>.

Des défauts de modifications post-traductionnelles des histones ont été aussi décrits dans des cas d'infertilité masculine chez l'homme et la souris, comme par exemple lors d'une

diminution du niveau de l'acétylation de l'histone H4<sup>86</sup> ou l'inhibition de la méthyltransférase MII2 (H3K4) qui entraîne un blocage de la différenciation spermatique<sup>87</sup>.

De même, la transition histone-protamine est essentielle et l'invalidation de TNP1 et/ou TNP2 conduit à une infertilité<sup>88</sup>. Un défaut de protamination peut également avoir un impact délétère sur la fertilité : un ratio P1/P2 anormal entraîne une diminution de la qualité embryonnaire et des taux de grossesse<sup>89</sup>.

## 2) La méthylation de l'ADN

### 2.1) Rôle et fonctionnement biologique de la méthylation de l'ADN

La méthylation de l'ADN est l'apport d'un groupement méthyle (CH<sub>3</sub>) par liaison covalente sur le carbone 5' de la cytosine. Chez les animaux, la méthylation de la cytosine se fait quasi exclusivement dans un contexte « CpG », c'est-à-dire dans le cas d'une cytosine suivie d'une guanine<sup>75</sup>. La distribution des cytosines méthylables au sein du génome n'est pas homogène et conditionne le rôle joué par cette marque épigénétique. Certaines régions du génome sont riches en CG, dites « îlots CpG ». Une majorité de ces îlots est située dans des séquences répétées, principalement au sein de séquences retrotransposables<sup>90</sup>. Ces éléments, qui ont le potentiel de se répliquer et de transloquer dans le génome, peuvent engendrer de fortes perturbations d'expression génique (perte ou sur expression). L'hyperméthylation des îlots en amont de ces séquences permettrait de juguler leur activité transcriptionnelle. Ainsi, la méthylation de l'ADN contribue à la stabilité du génome. Le rôle de la méthylation de l'ADN est aussi de réguler l'expression génique, via l'activation ou la répression de la transcription<sup>75</sup>. En effet, l'impact de l'apposition des groupements méthyles sur ces CpG varie selon le contexte génomique de(s) l'îlot(s) concerné(s). La méthylation des îlots CpG situés dans le promoteur d'un gène réprime sa transcription<sup>91</sup>, alors que celle située dans le premier exon semble plutôt l'activer.

La méthylation de l'ADN est un processus dynamique, sous contrôle d'une famille d'enzymes, ayant des fonctions spécifiques : les DNMT (*DNA MethylTransferase*). La méthylation *de novo* est assurée par les DNMT3a et DNMT3b. Ces deux enzymes ne sont pas exprimées dans les mêmes tissus et/ou au même stade de développement. La méthylation dite de « maintien »,

assurée par la DNMT1, permet la méthylation des brins néosynthétisés au moment de la réplication de l'ADN. Cette fonction est capitale puisqu'elle assure l'héritabilité de l'information épigénétique au cours de la mitose. A contrario, il existe aussi des systèmes enzymatiques impliqués dans la déméthylation ; ces systèmes sont complexes et dépendent de la phase physiologique pendant laquelle ils sont activés. La famille d'enzymes TET (*Ten-Eleven Translocation*) est impliquée dans des phénomènes d'hydroxylation de la méthylation et la mise en place de 5-hydroxyméthyl cytosines est considérée aujourd'hui comme une marque épigénétique à part entière.

## 2.2) Impact sur la fertilité

Au cours de la différenciation gamétique, une vague de déméthylation massive (mais pas totale) intervient dans les PGCs lors de leur migration vers les crêtes génitales, écrasant leur profil épigénétique « pseudo somatique » pour un profil épigénétique de cellules souches. L'apposition de nouvelles marques de méthylation est ensuite dépendante du sexe de l'individu et de la présence de DNMT3A/DNMT3B, enzymes responsables de la méthylation de novo et du cofacteur DNMT3L, dépourvu d'activité de méthyl transfert, mais nécessaire à celles des DNMT3A et 3B. Chez la souris, une perte de fonction des DNMT3A, 3B ou 3L conduit à des désordres importants de la spermatogenèse, pouvant aller jusqu'à la stérilité (arrêt en méiose de la lignée germinale, apoptose des spermatocytes ou une absence de méthylation de novo). Chez l'homme des polymorphismes de séquences identifiées au niveau des séquences codantes des gènes *DNMT* sont associés à des spermogrammes anormaux et à une stérilité<sup>92</sup>.

De multiples études ont montré un lien entre défaut de méthylation de l'ADN et altérations spermatiques chez l'homme, avec un focus particulier sur les gènes soumis à empreinte. Par exemple, une hypo-méthylation d'*IGF2* et une hyper-méthylation de *MEST* ont été observées dans une cohorte de patients oligozoospermiques<sup>93,94</sup>. Une faible ou une absence de méthylation sur le locus *IGF2/H19* est synonyme d'une semence de mauvaise qualité (baisse de mobilité) et faiblement concentrée<sup>95</sup>. La méthylation des promoteurs de quelques gènes impliqués dans la spermatogenèse a également été analysée, montrant des altérations de

méthylation des promoteurs de DAZL, et CREM chez des patients infertiles<sup>96-98</sup>. Une étude a comparé le méthylome spermatique d'hommes normozoospermiques fertiles et infertiles, mettant en évidence 2752 CpG présentant un profil de méthylation altéré au niveau de CpGs hyperméthylés dans les spermatozoïdes<sup>99</sup>, suggérant une implication de la méthylation dans l'infertilité idiopathique.

### 3) Les petits ARN non codants

Le dogme central de la biologie moléculaire esquissé en 1957 par Francis Crick et revisité en 1965 par James Watson, postule que l'information génétique, codée dans l'ADN, est transmise à partir des acides nucléiques vers les protéines, ce qui est souvent résumé par « l'ADN produit l'ARN et l'ARN produit la protéine ». Si ce dogme est toujours valable concernant les ARN messagers transcrits par l'ARN polymérase à partir des gènes codants, on sait aujourd'hui qu'il existe plus de 600 familles d'ARN non-codants répertoriées dans des bases de données comme Rfam ou RNAdb. Ces familles diffèrent par leur structure (linéaire, tige-boucle ou circulaire) et leur longueur (de quelques bases à plusieurs centaines). On distingue les petits (<200 nucléotides) et les longs (>200 nucléotides) ARN non codants. Par ailleurs, ces ANR peuvent subir des modifications post-transcriptionnelles, comme le phénomène d'édition dû à diverses enzymes qui modifient l'ARN, notamment en transformant l'adénosine (A) en inosine (I) ou la cytosine (C) en uracile (U).

Les petits ARN non codants (*sncRNA*) se subdivisent en deux catégories : les *sncRNA* structurants et les *sncRNA* régulateurs<sup>79</sup>. Les *sncRNA* structurants sont généralement exprimés de manière constitutive et sont indispensables pour le fonctionnement normal de la cellule. Il s'agit par exemple, des ARN ribosomiques (*rRNA* ; constituant les ribosomes) ou des ARN de transfert (*tRNA* ; apportant les acides aminés lors de la traduction). Les *sncRNA* régulateurs sont exprimés de manière spécifique, répondant à des stimuli particuliers (étape du développement, différenciation ou réponse spécifique à l'environnement). Les microARN (*miRNA*), les piwi ARN (*piRNA*) ou encore les ARN circulaires (*circRNA*) font partie de ce second groupe d'ARN.

Certaines de ces familles sont étudiées depuis fort longtemps, comme les ARN de transfert ou ribosomiques, tandis que d'autres sont étudiées depuis une trentaine d'années (miRNA) ou

plus récemment encore comme les ARN circulaires (depuis 2012). Leur biogenèse, leurs rôles ou leurs mécanismes d'action ont été étudiés chez la souris, chez l'homme ou chez les organismes modèles (*C. elegans*, *drosophile*...). Bien que certaines données existent chez les bovins, l'état des connaissances dans cette espèce est nettement moins avancé. Les paragraphes suivants détailleront pour les principales familles de sncRNA les informations publiées concernant leur formation, leurs rôles et leurs mécanismes d'actions utilisés. Leur implication dans la spermatogenèse ou le développement embryonnaire sera abordée dans un paragraphe spécifique.

### 3.1) Les microARN (miRNA)

Parmi les sncRNA, les miRNA représentent la famille la plus étudiée. Au début des années 1990, le premier miRNA, lin-4 a été identifié chez le nématode *Caenorhabditis elegans*. Tout d'abord étiqueté comme ARNm codant une protéine régulant le développement, il a été ensuite établi que lin-4 n'existait que sous la forme d'une courte séquence d'ARN de 22 nucléotides dépourvue de phase codante<sup>100</sup>. Par contre, cette courte séquence pouvait se lier à un autre ARNm, Lin-14<sup>101</sup>. Cette liaison empêchait toute production de la protéine. Alors que lin-4 est spécifique de *C. elegans*, la découverte de let-7 chez de très nombreuses espèces, dont l'Homme<sup>101</sup>, a permis d'établir l'existence de petites séquences transcrites ayant un rôle régulateur, ouvrant un nouveau champ d'investigation concernant la régulation de l'expression génique. Dès lors de nombreuses recherches ont été menées conduisant à l'identification de plusieurs milliers de miRNA. Le graphique représentant le nombre d'articles publiés contenant le terme « miRNA » est éloquent (figure 20).

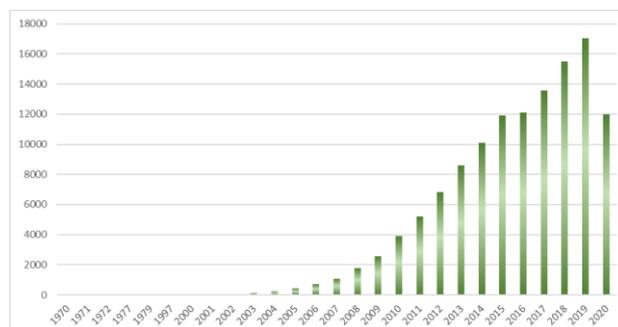


Figure 20 : Nombre de publications contenant l'occurrence « miRNA ». Les sncRNA et en particulier les miRNA, sont de plus en plus étudiés depuis le début des années 2000. Source PubMed (<https://pubmed.ncbi.nlm.nih.gov/>).

Andrew Fire et Craig Mello à l'origine des travaux d'interférence par l'ARN, ont été récompensés en 2006, par l'attribution du Nobel de médecine, reconnaissant ainsi l'impact scientifique très important qu'a eu cette découverte. Avec le nombre croissant d'études portant sur l'identification de ces sncRNA, la communauté scientifique s'est organisée pour regrouper, annoter et mettre à disposition une base d'information sur les miRNA. Aujourd'hui, *miRBase*<sup>102</sup> (<http://www.mirbase.org> ; release V22.1, octobre 2018), recense plusieurs milliers de miRNA connus dans 271 organismes différents, dont 2654 chez l'homme, 1978 chez la souris ou 1025 chez le bovin. Chez les ruminants, un consortium de chercheurs INRAE a constitué la base *RumimiR*, regroupant les miRNA publiés<sup>103</sup> (<http://rumimir.sigenae.org>) chez les bovins, ovins et caprins. Toutefois, peu ou pas d'informations sont disponibles concernant le contenu en miRNA (et plus largement en sncRNA) des spermatozoïdes bovins. Une seconde initiative, concernant cette fois-ci les sncRNA du spermatozoïde, a vu le jour en 2016<sup>104</sup>. *SpermBase* (<http://spermbase.org/index.php>) regroupe les listes de sncRNA publiés chez l'homme, la souris et le rat.

Les miRNA sont devenus en quelques années un domaine d'étude prometteur en thérapie et en particulier en oncologie, de par une meilleure compréhension des mécanismes impliqués dans l'établissement de la tumeur, la recherche de variation d'expression de ces miRNA comme marqueurs de la présence des cellules tumorales<sup>105</sup> ou encore en étudiant les possibilités des RNA interférences pour bloquer l'expression de gènes délétères. C'est le cas par exemple pour la tumeur de Wilms<sup>106</sup> ou pour l'ostéosarcome<sup>107</sup>, où l'analyse de l'expression de certains miRNA signe ces tumeurs. Au-delà de la cancérologie, d'autres domaines de la médecine utilisent les sncRNA comme biomarqueurs d'un traumatisme dont l'origine peut être diverse (mauvaise pratique du sport, stress thermique...) <sup>108</sup>, ou pour diagnostiquer rapidement une infection virale ou bactérienne<sup>109</sup>.

### 3.1.1) Biogenèse et régulation de l'expression des miRNA

Près de la moitié des séquences codant les miRNA sont situées en régions intergéniques, la seconde moitié étant localisée dans des régions codantes, principalement dans les introns ou dans les régions UTR<sup>100</sup>. La majorité des miRNA sont « isolés », mais certains sont organisés en « cluster » et transcrits de concert (par exemple le cluster de miRNA porté par le chromosome 21 chez le bovin figure 21).



Figure 21 : Exemple d'organisation génomique en « cluster » de certains miRNA positionnés sur le chromosome bovin 21 (<https://www.ensembl.org>).

Les miRNA sont transcrits par l'ARN Polymerase II en longs précurseurs ayant une structure caractéristique de « tige-boucle », appelés **pri-miRNA**<sup>100</sup>. Un même pri-miARN peut servir de précurseur à un miRNA (mono-cistronique) ou à plusieurs (poly-cistronique). La transcription du pri-miARN peut être déclenchée par la transcription puis la maturation de l'ARN messenger<sup>100,110</sup> de leurs gènes hôtes lorsque le pri-miARN est intronique, ou être déclenchée indépendamment par des promoteurs dédiés. Ces précurseurs sont ensuite pris en charge par DROSHA<sup>111</sup>. Cette enzyme, associée à DGCR8 (une enzyme de liaison à l'ARN), va entreprendre le clivage du précurseur en libérant les séquences flanquantes. Il en résulte une séquence en épingle à cheveux de 70 à 80 nucléotides de long appelée **pré-miRNA** (figure 22).

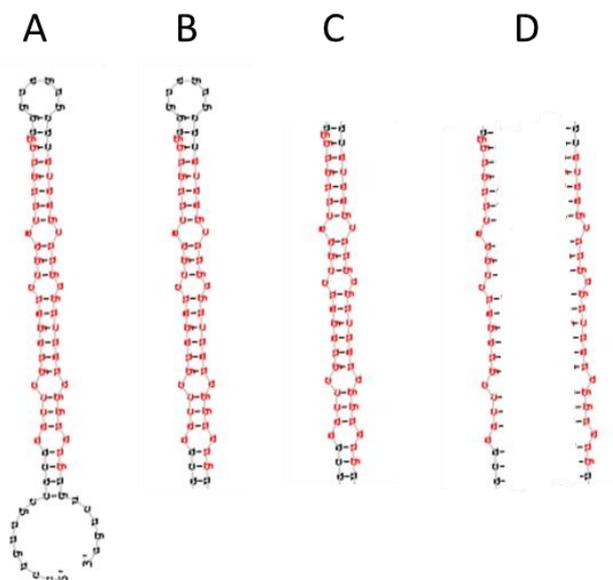


Figure 22 : Les miRNA de la forme native à la forme mature. Exemple du bta-miR-100. La forme native, nouvellement transcrite, est appelée pri-miRNA (A). Elle va subir différentes maturations, passant par la forme pré-miRNA (B) puis miRNA doubles brins (C), avant d'atteindre sa forme mature 5' ou 3' simple brin (D). @Alice

Cet intermédiaire est alors pris en charge par l'Exportine 5 pour être exporté hors du noyau jusque dans le cytoplasme, où la boucle de l'épingle sera clivée par la protéine DICER associée à TRBP, une autre protéine de liaison à l'ARN<sup>112</sup>. Il en résulte un complexe de deux brins d'ARN, imparfaitement complémentaires, d'une vingtaine de nucléotides (la taille est variable en fonction du miRNA), qui se séparent pour former les brins matures (figure 23).

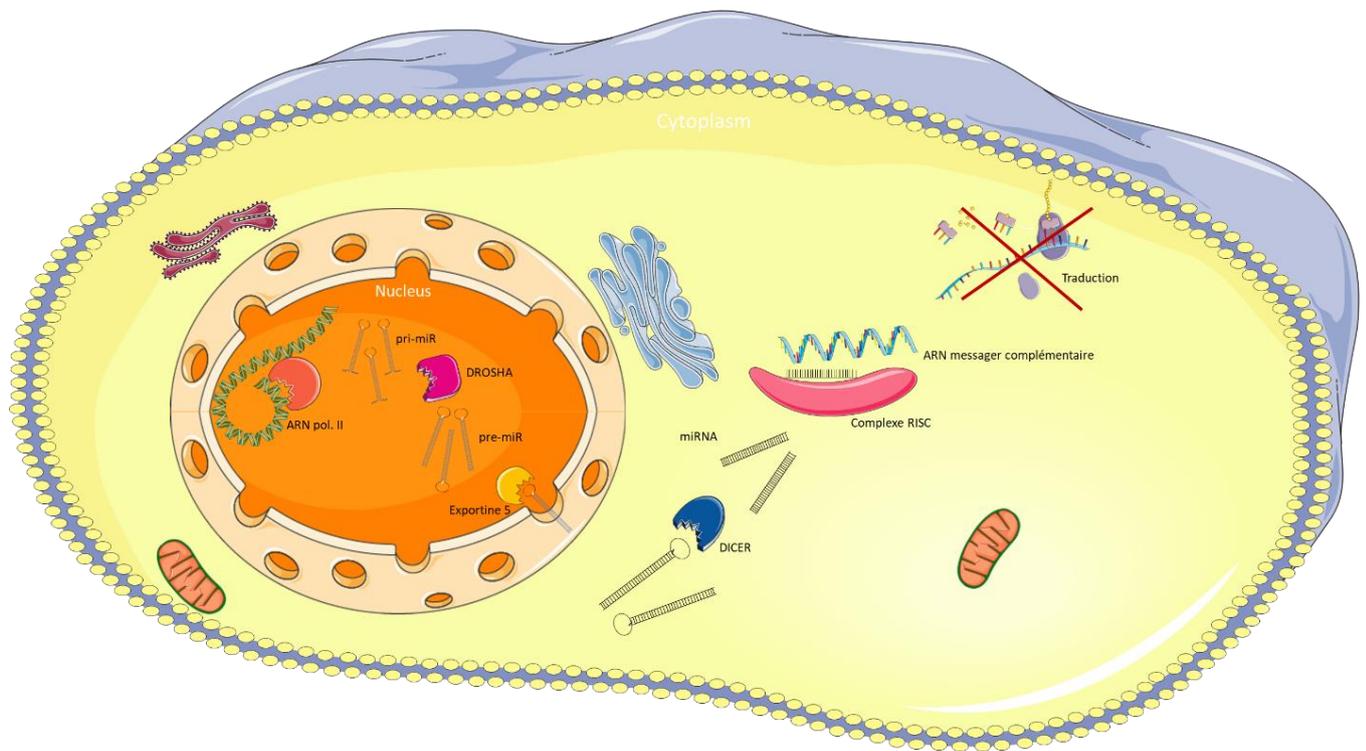


Figure 23 : Les miRNA, de leur biogenèse à leur fonction.

Les miRNA sont transcrits par l'ARN polymérase II. La forme native appelée pri-miR est prise en charge par DROSHA qui va tronquer les extrémités non complémentaires. Cette forme intermédiaire, appelée pré-miR est alors exportée hors du noyau. Leur tige boucle sera alors éliminée par DICER, pour former une séquence double brins d'environ 22 nucléotides. Le brin mature est ensuite chargé dans le complexe protéique RISC. Cet ensemble par complémentarité de séquence entre le miRNA et l'ARN messager régule sa traduction en le séquestrant et/ou en le clivant. @Alicia

L'existence d'isoformes, appelées isomiR, induit une variabilité supplémentaire des miARN (figure 24). En effet, les miARN sont sensibles à des transformations post-transcriptionnelles telles que des substitutions de nucléotides<sup>113</sup>, l'ajout d'adénine ou d'uracyle à l'extrémité 3' du transcrit<sup>114</sup>, le raccourcissement de l'extrémité 3' du transcrit<sup>115</sup> ou l'ajout de nucléotides en 5'<sup>116</sup>. Ces isomiR ont été décrits dans tous les tissus dans lesquels des miRNA canoniques ont été identifiés et une signature tissus-spécifique d'isomiR a été montrée<sup>117,118</sup>. En effet, les types d'édition et leur fréquence varient selon les tissus ou les stades de développement, certains miRNA y étant plus sensibles que d'autres (tout du moins chez l'homme, *C. elegans*

et la souris). C'est le cas des miR-100, miR-146a, miR-151-3p, miR-143, miR-331-3p, miR-23b, miR-24, miR-222, miR-199a-3p et miR-1308 qui présentent chez l'homme plus de variants que les autres miRNA canoniques<sup>119</sup>. Par ailleurs, l'ajout de nucléotide(s) en 3' n'est pas le fruit du hasard<sup>119</sup>, mais résulte de l'action d'au moins 7 nucléotidyl transferases, (telles que PAPD4 ou l'uridylyltransferase)<sup>120</sup>. Plusieurs éléments laissent à penser également que les isomiR ont un rôle fonctionnel<sup>117</sup>. D'une part, leur présence dans les complexes RISC a été montrée par des études de coprécipitation avec les protéines argonautes<sup>116</sup>. D'autre part, les travaux de Tan et al. publiés en 2014, ont démontré, en utilisant un système de gène rapporteur luciférase couplé au 3' UTR d'un gène cible du miRNA-9 (CDH1) ou de l'un de ses isomiR (DNMT3B), que le miRNA et son isomiR ne régulaient pas forcément l'expression des mêmes cibles<sup>116</sup>.

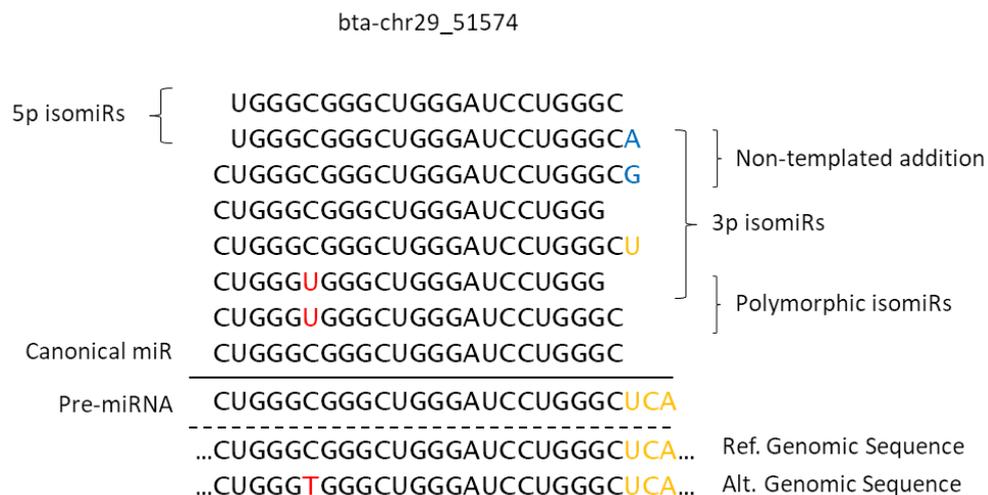


Figure 24 : Les différentes éditions d'un miRNA canonique : les isomiR. Exemple du miRNA putatif bta-chr29\_51574. Un miRNA dit canonique représente en réalité plusieurs séquences variant en générale d'une base. Ces éditions peuvent avoir lieu en 5', en 3' ou même à l'intérieur de la séquence. Une partie de ces éditions est liée à du polymorphisme de séquence. Pour le reste, ces éditions semblent être spécifiques. @Alice

### 3.1.2) Mécanismes d'action et rôle des miRNA

Le brin mature « sens » (5'-3') interagit avec Ago2, une protéine de la famille des argonautes, pour composer le complexe RISC (*RNA-induced silencing complex*). Le second brin, « antisens » (3'-5') peut également agir comme un miRNA mature, ou être dégradé. D'un point de vue nomenclature, ce miRNA est appelé miR-xx-3p alors que celui formé de la séquence

« sens » porte le nom de miR-xx-5p. Le complexe RISC reconnaît alors l'ARNm cible par complémentarité de séquence avec celle du miRNA mature, provoquant le clivage de l'ARNm cible ou le blocage de la traduction<sup>17,121-123</sup>.

Le miRNA peut être subdivisé en 2 parties : la « seed » (du nucléotide 2 au nucléotide 8) et le corps (du 9<sup>e</sup> au dernier). La région « seed » joue un rôle majeur pour la reconnaissance et l'appariement entre le miRNA et l'ARNm cible<sup>124</sup>, alors que le corps permettrait d'orienter le choix du type de régulation<sup>100</sup> : si la complémentarité est totale, l'ARNm serait détruit par l'activité catalytique des AGO<sup>125</sup> ; dans le cas contraire, la régulation résulterait plutôt d'un blocage de la traduction.

Les miRNA sont également soumis à des systèmes de régulation de leur activité. Au-delà de la régulation de leur transcription, leur maturation peut également être contrôlée empêchant la formation des brins matures. C'est le cas pour Let-7, pour lequel la liaison de la protéine LIN28, au niveau de la tige boucle, inhibe sa maturation<sup>126</sup>. Let-7 est ensuite dégradé par le recrutement d'une autre protéine, TUT4<sup>127</sup>. Un dernier système de régulation de l'expression des miRNA a été découvert il y a quelques années. Il s'agit d'une autre famille de petits ARN non codants, les ARN circulaires, qui agiraient comme des éponges à miRNA. Cette famille est décrite plus en détail dans les paragraphes suivants.

Les miRNA cibleraient environ 60% des gènes<sup>128</sup> et ont donc un rôle crucial dans la plupart des processus biologiques (prolifération, développement, différenciation, réponse au stress, adhésion cellulaire, mobilité, apoptose...). Il n'est donc pas étonnant que le KO constitutif de DICER chez la souris entraîne une létalité embryonnaire<sup>129</sup>. De même des KO tissus spécifiques démontrent un rôle des miRNA dans l'organogenèse et le maintien d'un grand nombre de tissus et d'organes (peau, follicules pileux (figure 25), cœur, poumon, muscle ...) <sup>130,131</sup> ainsi que le maintien des cellules souches<sup>132</sup>.



Figure 25 : Perturbation importante du système pileux chez la souris mutante du gène *Dicer*<sup>131</sup>  
Une KO spécifique du gène *DICER* dans les cellules de l'épiderme conduit à la génération de souris présentant des désordres importants du système pileux.

### 3.2) Les piRNA

Vers la fin des années 90, un ensemble de protéines, alors appelées PIWI (*P-element induced wimpy testis*), a été identifié dans les cellules souches germinales de *Drosophila melanogaster*, avant d'être également aussi décrit chez d'autres espèces<sup>133</sup>. Ces protéines nucléaires appartiennent aussi à la famille des Argonautes<sup>134</sup>. Chez la souris, il existe trois PIWI protéines : MILI, MIWI 1 et 2, exprimées différemment selon le sexe de l'animal. En effet, MILI existe dans les 2 sexes, alors que les MIWI sont exprimées uniquement dans la lignée germinale mâle. Les protéines PIWI comportent 4 domaines : N-terminal, PAZ, MID et PIWI.

Quelques années plus tard, plusieurs publications<sup>133,135</sup> font état de petits ARN qui se lient à ces protéines. C'est la découverte des piRNA, de petits ARN non codants de 26-33 nucléotides, présentant une méthylation de leur extrémité 3', ainsi qu'une base uridine à leur extrémité 5'. Les piRNA, étudiés principalement chez la Drosophile et chez la souris, demeurent mystérieux et des investigations supplémentaires restent nécessaires.

#### 3.2.1) Biogenèse des piRNA

La biogenèse des piRNA est totalement différente de celle des miRNA (figure 26). Elle implique deux voies distinctes : la biogenèse « primaire » et le cycle « ping-pong »<sup>136</sup>. Il semble que certains piRNA, comme ceux produits dans les cellules germinales au stade pachytène, soient exclusivement produits via la première voie<sup>137</sup>.

Les piRNA sont produits à partir de transcrits précurseurs codants pour des clusters de piRNA. Ces précurseurs sont exportés dans le cytoplasme et découpés en intermédiaires de piRNA par l'endonucléase ZUC qui, en association avec des cofacteurs tels que VRET, MINO ou GASZ, découpe le brin en 3' tous les 25 à 35 nucléotides, au niveau d'une uracile, générant la particularité de structure 5' 1U. Dans la voie primaire, ces intermédiaires sont pris en charge par PIWI, et divers mécanismes (coupure 3' par Trimmer, méthylation par Hen1) permettent de générer l'extrémité 3' méthylée des piRNA matures dits « primaires »<sup>136</sup>.

Le cycle « ping-pong » est initié par la protéine AUB associée à un piRNA primaire 1U. Par complémentarité de séquence, ce complexe dirige le clivage 5' du rétrotransposon ciblé, au niveau du nucléotide qui s'apparie avec le 10e nucléotide du piRNA. Le fragment 5' est éliminé,

alors que le fragment 3' constitue un nouveau piRNA sans biais de séquence 1U, mais avec une Adénosine en position 10. Ce nouveau piRNA, dit piRNA « secondaire », est pris en charge par Ago3 et son extrémité 3' est maturée par Trimmer et Hen1. Ce nouveau complexe cible et clive à nouveau le long transcrite primaire de piRNA, libérant un nouveau piRNA 5' 1U et initiant un nouveau cycle<sup>138</sup>.

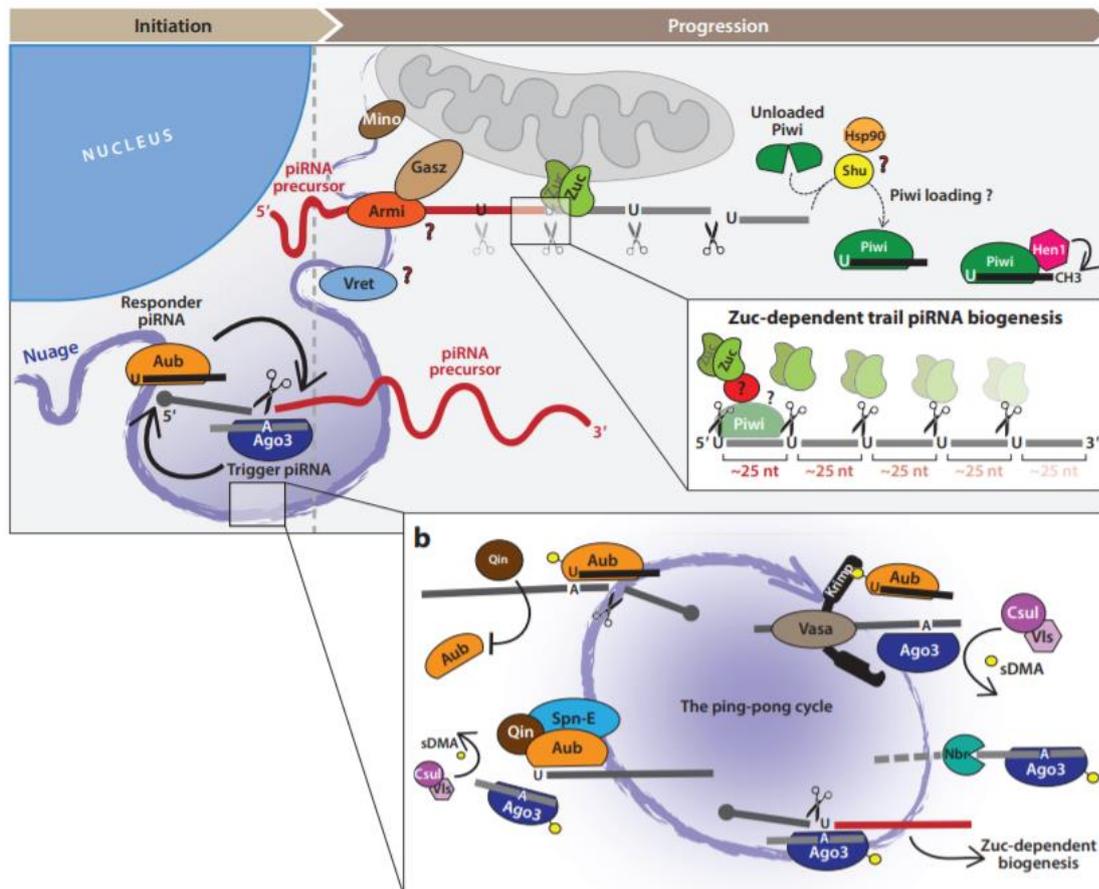


Figure 26 : Biogénèse des piRNA<sup>136</sup>.

La biogénèse des piRNA implique deux voies distinctes : La production de piRNA dit « primaire » et présentant un U en première position (5p). Cette voie fait intervenir principalement la protéine ZUC qui clive tous les 25 à 35 nucléotides le long transcrite primaire de piRNA. La production de piRNA dit « secondaire » au cours du cycle « ping-pong » et présentant une A en position 10. Cette seconde voie fait intervenir principalement AGO3 et AUB qui génèrent des séquences de piRNA primaire et secondaire en clivant soit des transposons soit des longs transcrits primaires de piRNA.

### 3.2.2) Rôles et mécanismes d'action

Les piRNA ont été initialement identifiés exclusivement dans la lignée germinale des animaux<sup>133</sup>, suggérant un rôle essentiel dans le développement des gamètes. Le premier rôle attribué à cette famille de petits ARN non codants est de lutter contre l'instabilité du génome, par la répression des transposons<sup>134</sup>. Les éléments transposables peuvent en effet se dupliquer et se transloquer tout au long du génome, provoquant des insertions dans des exons, conduisant à une protéine non fonctionnelle ou dans les 3' UTR, dérégulant l'expression du gène. Ils engendrent également des cassures de l'ADN<sup>139</sup>. Une fois captés par les PIWI protéines dans le cytoplasme et retournés dans le noyau, ces complexes sont capables à la fois de reconnaître les séquences transposables et de recruter les DNMT3L, 3A et 3B, afin d'induire localement une méthylation de l'ADN et bloquer ainsi toute activité de ces séquences (LINE1 ou LTR<sup>133,134,137,140</sup>). Il semblerait également que les complexes PIWI-piRNA répriment l'expression des transposons par des modifications locales de la condensation de la chromatine<sup>136</sup> due au recrutement de facteurs apposant des modifications post-traductionnelles des histones répressives (histone méthyltransferase Egg, WDE, MAEL). Ces actions conduisent à la fermeture locale de la chromatine, réprimant donc l'expression du transposon.

Mais depuis quelques années, l'existence d'une faible quantité de piRNA et/ou séquences dites « piRNA-like » présentant les mêmes caractéristiques que des piRNA a également été montrée dans les cellules somatiques<sup>139</sup>. Cette découverte s'accompagne également d'un nouveau rôle fonctionnel associé aux piRNA. En effet, certains piRNA identifiés chez plusieurs espèces, ne sont pas complémentaires d'éléments transposables<sup>133</sup> et seraient plutôt impliqués dans la régulation de l'expression génique, via l'activité catalytique médiée par l'un des domaines des PIWI protéines<sup>135</sup>. Chez la souris, cette activité catalytique est réalisée par une deadenylase en interaction avec MIWI, CAF1<sup>133,135</sup>.

### 3.3) Des ARN de transfert aux tRFs

Les ARN de transfert jouent un rôle central dans la synthèse protéique chez les procaryotes, et les eucaryotes. Leur structure et leur fonction font partie des systèmes les plus conservés du monde biologique. L'architecture des gènes à ARNt ne ressemble pas à celle des autres sncRNA<sup>141</sup>. Chez les eucaryotes, une structure particulière appelée « *Permuted genes* » est observée. Dans cette structure, les régions 3' et 5' de l'ARN proviennent du même locus, mais sont en position inverse. Suite à la transcription, le brin se replie et les deux parties complémentaires s'associent pour former l'ARNt<sup>141</sup>.

#### 3.3.1) Biogenèse des tRFs

Les tRFs sont des molécules qui dérivent du clivage des ARNt cytoplasmiques et mitochondriaux matures ou du transcrit précurseur. Cinq types structuraux de tRFs ont été décrits : les moitiés d'ARNt (5'-tRH, 3'-tRH) relativement longues (30-35nt) et les fragments courts (18-22nt) tRF5, tRF3 et i-tRF (Figure 27)<sup>142,143</sup>. Ils résultent de l'action de diverses enzymes telles la RNase Z, la RNase P, DICER ou l'Angiogenine<sup>144</sup>.

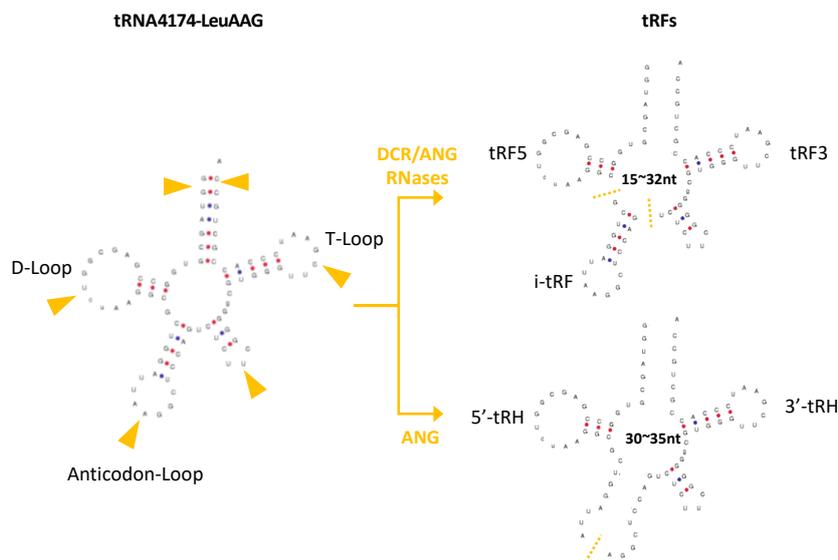


Figure 27 : Biogenèse des fragments d'ARNt.

Un ARNt est composé classiquement de plusieurs structures tige-boucles (appelées D-Loop, T-Loop et Anticodon-Loop). Ces structures peuvent être clivées par différentes endonucléases (DICER, RNase Z ou l'Angiogenine). Les fragments obtenus constituent les « tRFs ». Suivant le site de coupure, leur nomenclature est différente. Ces fragments peuvent être pris en charge par un complexe de type RISC pour réguler la traduction de certains ARN messagers.

### 3.3.2) Rôles et mécanismes d'action

De plus en plus d'études, portant sur différentes espèces, présentent l'implication spécifique des tRFs dans la réponse de l'organisme à un stress, la signalisation cellulaire via les vésicules extracellulaires, la réponse aux infections virales et les désordres neurologiques<sup>145,146</sup>. Leur expression augmente par exemple lors de l'infection par le phage T4 chez E.coli, sous l'effet d'une privation de nutriments, en cas de stress oxydatif important, ou tout simplement au cours du développement normal de l'organisme (bactérie et champignon)<sup>143</sup>. Par ailleurs, les tRFs semblent jouer un rôle dans l'interférence de l'ARN, en particulier dans la suppression des rétrovirus et des rétrotransposons qui utilisent l'ARNt comme amorce de réplication. Plus particulièrement, certains tRF3 permettraient de bloquer la rétrotranscription des rétrotransposons à LTR, constituant un deuxième niveau de contrôle lorsque le blocage de la transcription par la méthylation des histones (H3K9me3) est levé, par exemple lors de la reprogrammation épigénétique de l'embryon préimplantatoire de souris<sup>147</sup>.

De plus ces tRFs peuvent être captés par les argonautes AGO2 (mammifère), AGO3 ou AGO4 afin d'exercer une régulation de la transcription de type miRNA<sup>100,143</sup>. Globalement, les tRFs ont une action dans de nombreux processus biologiques dont les mécanismes restent à élucider.

## 3.4) Des ARN ribosomiques aux rsRNA

### 3.1.1) Biogenèse des rsRNA

L'ARNr est transcrit dans le nucléole, à partir de quelques longues régions de l'ADN (chez l'homme environ 43kb), par l'ARN polymérase I en un long précurseur, le pré-ARNr 47S. Ce pré-ARNr va subir une succession de clivages endonucléolytiques, de digestions exonucléolytiques et de modifications (pseudo-uridylation, méthylations) faisant appel aux « small nucleolar RNA » (snoRNA), pour donner naissance aux ARNr matures 18S, 5.8S et 28S<sup>148</sup>. L'ARNr 5S, troisième ARNr de la grande sous-unité ribosomique, est transcrit indépendamment par l'ARN polymérase III. Ces ARNr forment les deux sous unités du ribosome : SSU, *small subunit*, formée par l'ARNr 18S et LSU, *large subunit*, formée par les ARNr 5S et 28S<sup>149</sup>.

De nombreux fragments dérivés de l'ARNr ont été identifiés par séquençage NGS dans différents types cellulaires et pendant le développement, considérés initialement comme des produits de dégradation et écartés des analyses. Toutefois, Wei et al.<sup>150</sup> ont montré que 70% de ces rsRNA provenaient de l'ARNr 28S et Chen et al.<sup>151</sup> ont découvert que les rsRNA n'étaient pas distribués au hasard le long de l'ARNr 28S, mais au contraire étaient significativement enrichis aux extrémités 5' (rRF5) et 3' (rRF3). Cela suggère une production par un processus endonucléolytique spécifique plutôt qu'une digestion exonucléolytique aléatoire. La biogenèse de ces fragments fait vraisemblablement appel à des voies parallèles utilisant les mêmes endo- et exonucléases que celles utilisées pour les étapes de clivage et de maturation des ARNr. L'ARN 28S peut également donner naissance à des miRNA ou des piRNA via les voies « miRNA-like » et « piRNA pathways » correspondant à des modifications atypiques<sup>152</sup>. Les fragments d'ARNr 5.8S seraient quant à eux produits à partir soit de l'ARNr mature, soit de précurseurs d'ARNr, selon un processus similaire à celui donnant naissance aux tRF (clivage par la RNase Z), ou par un processus similaire à celui décrit pour les miARN et impliquant une ou plusieurs endoribonucléases de type Dicer<sup>152</sup>.

### 3.1.2) Rôles et mécanismes d'action

L'expression des rsRNA a été associée à des désordres métaboliques et aux processus inflammatoires chez la souris<sup>150,153,154</sup>. Leur expression varie selon le stade de développement ou l'état de stress<sup>155,156</sup>. Les travaux de Guan et al. montrent dans un contexte âge-dépendant que les rsRNA sont présents dans les complexes argonautes formés par AGO1 et AGO2 (figure 28) et des prédictions bio-informatiques indiquent qu'ils peuvent cibler des ARNm impliqués dans une grande variété de processus biologiques<sup>157</sup>. Mais leur importance fonctionnelle et biologique reste à étudier en détail.

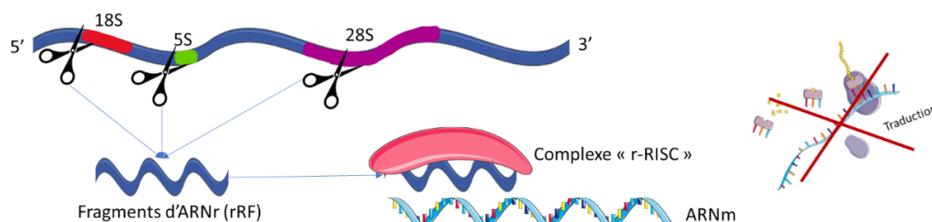


Figure 28 : Biogenèse des fragments d'ARNr<sup>157</sup>.

Les fragments d'ARNr générés par le clivage d'ARN ribosomiques, appelés « rRFs » peuvent être pris en charge par un complexe de type RISC pour réguler la traduction de certains ARN messagers.

### 3.5) Les autres familles de petits ARN non codants

D'autres familles de petits ARN non codants régulateurs ou structurants existent. Comme par exemple les « small nucleolar RNA », autrement appelés snoRNA, qui représentent une large sous-famille d'ARN non codants, présents dans toutes les cellules eucaryotes. Comme mentionné ci-dessus, ils jouent un rôle central dans la biogenèse des ribosomes.

La dernière famille à avoir été identifiée est celle des ARN circulaires (circRNA). Ces ARN sont caractérisés par une liaison covalente entre leurs extrémités 3' et 5'<sup>158</sup>. Il semble que le génome des animaux exprime une quantité importante de circRNA<sup>159</sup>. Peu de données sont publiées concernant leur rôle fonctionnel. Mais certaines études mettent en avant un rôle « d'éponge » à miRNA. En séquestrant les miRNA, leurs ARNm cibles ne se sont pas réprimés et leur traduction en protéine peut donc avoir lieu<sup>160</sup>. Chez le poulet l'ARN circulaire circSVIL, promeut le développement des muscles squelettiques embryonnaires par la séquestration du miR-203<sup>161</sup>. Rien qu'en 2018, plus de 160 travaux ont été publiés concernant l'implication de ces ARN circulaires dans l'initiation ou le développement de tumeurs, démontrant le niveau d'intérêt porté à ces circRNA. Il semblerait également que ces ARN soient transmis au sein de vésicules extracellulaires d'un tissu à l'autre<sup>162</sup>. Un autre point intéressant concerne le fait qu'il semble y avoir bien plus d'interactions (prédiction *in silico*) entre les isomiR et les circRNA qu'entre les miR canoniques et ces mêmes circRNA<sup>163</sup>.

Cette strate supplémentaire dans la régulation de l'expression génique n'est pas spécifique au règne animal. En effet, un système équivalent a été décrit chez les plantes, avec par exemple le rôle de l'ARN non codant IPS1 séquestrant le miR-399<sup>164</sup>.

### 4) Les petits ARN non codants au cours de la spermatogenèse

La spermatogenèse est un processus finement régulé qui s'accompagne de vagues de transcription, produisant un profil transcriptionnel spécifique à chaque stade<sup>165,166</sup>. Ainsi, un groupe spécifique de gènes est induit dans les spermatogonies en mitose, des gènes méiotiques sont exprimés dans les spermatocytes, et de nombreux gènes nécessaires à la

différenciation sont induits dans les spermatides. Les sncRNA ayant un rôle clé dans la régulation de l'expression génique, leur intervention dans le processus de spermatogenèse a naturellement été étudiée. Plusieurs classes de sncRNA ont pu être identifiées dans les spermatozoïdes, avec des différences de composition selon les stades de la spermatogenèse et de la maturation, comme illustré sur la figure 29<sup>167</sup>. Il s'agit principalement de piRNA, miRNA, tsRNA et rsRNA.

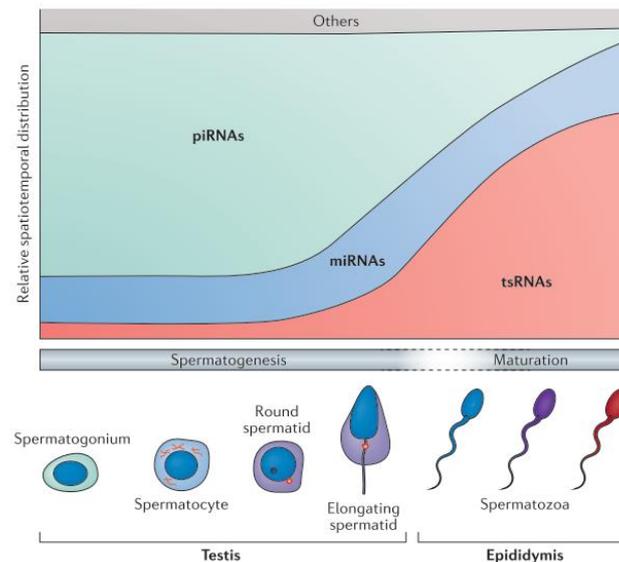


Figure 29 : Diversité des sncRNA dans les cellules germinales mâles murines. Distribution relative des principaux sncRNA dans les cellules germinales au cours de la spermatogenèses et post testiculaire<sup>167</sup>.

Le rôle majeur des piRNA et des miRNA est illustré, par exemple, par les défauts de spermatogénèse observé chez les souris après invalidation de l'un ou l'autre des gènes codant pour les protéines PIWI (souris KO MILI, MIWI2 ou MIWI)<sup>140</sup> ou l'effet délétère sur la spermatogenèse et la fertilité mâle de l'invalidation de DICER dans la lignée germinale et les cellules de Sertoli<sup>17,168-170</sup>.

Ce paragraphe décrit les fonctions remplies par les sncRNAs du spermatozoïde testiculaire au cours de la spermatogenèse, sachant qu'en dépit d'avancées majeures, d'importantes zones d'ombre persistent concernant les mécanismes moléculaires mis en jeu.

#### 4.1) Rôle des miRNA

La présence de DICER, d'Ago2, Ago3 et des miR-21, let-7a et miR-122a dans le corps chromatoïde des cellules germinales mâles<sup>171</sup>, la découverte de nombreux miRNA dans ses

cellules<sup>172</sup> et les variations d'expression observées entre stades (figure 30)<sup>17</sup> a très tôt suggéré un rôle des miRNA dans la régulation de la spermatogenèse.

		miRNAs (reported function)
 PGC	prenatal (dpc)	
	7	
 proSg		miR-17-92 cluster
	15	
 Sg	postnatal (dpp)	miR-291a-5p, -293, -294*, -290-5p
	3	<b>miR-21 (maintenance of undifferentiated Sg)</b>
 Sg	6	miR-34c, -182, -183, -146a, -465a-3p, -465b-3p, -465c-3p, and -465c-5p
		<b>miRNA-20,-106a (maintenance of undifferentiated Sg)</b>
		<b>miRNA-221/222 (maintenance of undifferentiated Sg)</b>
		<b>miR-17-92 cluster (KO mice have small testes and mild defects in spermatogenesis)</b>
		<b>miRNA-146 (repression of RA-mediated differentiation)</b>
		miR-136, -743a, -463*
		let-7f, -7a, -7b, -c7, -7g, -7i, miR-99a, -34a, -322, -201, -547, -204
		let-7a, -7b, -7c, -7d, -7e, 7g
 early Spc	PL	
	10	
 early Spc	Z	
	12	
 late Spc	P	
	14	miR-15b, -34b-5p, -34b-3p, -34c, -34c*, -449a, -425, -375, -3470a, -18a, -296-5p, -466i-5p, -3085-5p
 late Spc	D	
	18	miR-464, -113, -112, -120, -18, -114, -117, -469, -13
 late Spc	MI, MII	
	18-20	<b>miR-34c (induction of apoptosis)</b>
		<b>miR-34c (enhancing germ cell phenotype)</b>
		<b>miR-449 (cluster KO mice have unaffected spermatogenesis)</b>
 RS		<b>miR-18 (Regulation of Hsf2 mRNA)</b>
	20	<b>miR-469 (Regulation of Tp2 and Prm2 mRNAs)</b>
 RS		<b>miR-122a (Regulation of Tp2 mRNA)</b>
	28	

Figure 30 : Liste de miRNA impliqués dans la spermatogenèse chez la souris (figure adaptée<sup>17</sup>).

Différents exemples d'implication des miRNA dans la spermatogenèse sont illustrés dans cette figure. Que ce soit à des stades précoces ou au cours de la spermiogenèse, les miRNA semblent jouer un rôle primordial dans la bonne conduite de la spermatogenèse.

Mais la création de souris « floxées » permettant l'invalidation conditionnelle de *DICER* dans les cellules germinales constitue la première preuve de leur rôle. Ainsi, l'inactivation postnatale de *DICER* dans la lignée germinale n'a pas d'impact sur la viabilité des souris, mais les mâles adultes présentent des testicules deux fois plus petits et produisant moins de cellules haploïdes (les spermatozoïdes entrent en grande partie en apoptose), démontrant le rôle des miRNA à la fois dans les processus de prolifération des cellules germinales et dans la différenciation en spermatogonies. Par ailleurs, des problèmes importants de conformation du spermatozoïde ont été observés (élongation, organisation de la chromatine, compaction

et formation du flagelle défectueuse), aboutissant à d'importantes anomalies morphologiques (voir figure 31)<sup>168</sup>.

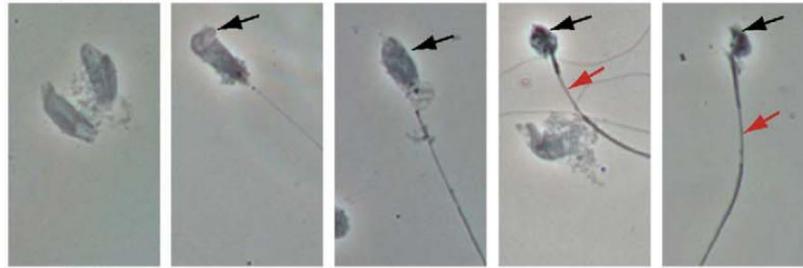


Figure 31 : Anomalies morphologiques du spermatozoïde murin délété du gène *Dicer*<sup>168</sup>. Une perturbation de l'expression de *DICER* conduit à de fortes perturbations de la spermatogenèse. Au-delà du faible nombre de spermatozoïdes retrouvé dans l'épididyme, des anomalies de conformation de la tête (flèches noires) et de la pièce intermédiaire (flèches rouges) sont également observées.

De même, l'inactivation de miR-34b/c et miR-449 (souris double KO miR-34bc<sup>-/-</sup> et miR-449<sup>-/-</sup>) induit une stérilité et l'analyse histologique des testicules montre une réduction drastique du nombre de spermatozoïdes produits, une perte de motilité et la présence d'anomalies morphologiques prononcées (décapitation ; figure 32)<sup>76</sup>. Il s'agit dans ce cas d'une oligoasthénotérazoospermie. L'analyse transcriptomique réalisée dans l'étude a mis en évidence quelques gènes dérégulés dans les spermatocytes sans pour autant établir un lien direct entre l'absence de ces miRNA, la dérégulation de ces gènes et la perturbation majeure de la spermatogenèse. De façon intéressante, la simple délétion d'un seul de ces deux miRNA n'a pas permis d'obtenir le même phénotype d'infertilité, mettant en lumière les possibles redondances entre certains miRNA. Chez l'Homme, 7 miRNA (miR-34c-5p, miR-122, miR-146b-5p, miR-181a, miR-374b, miR-509-5p, miR513a-5p) ont été associés à des cas d'infertilité masculine, notamment dans des cas d'azoospermie et d'asthénozoospermie<sup>17</sup>.

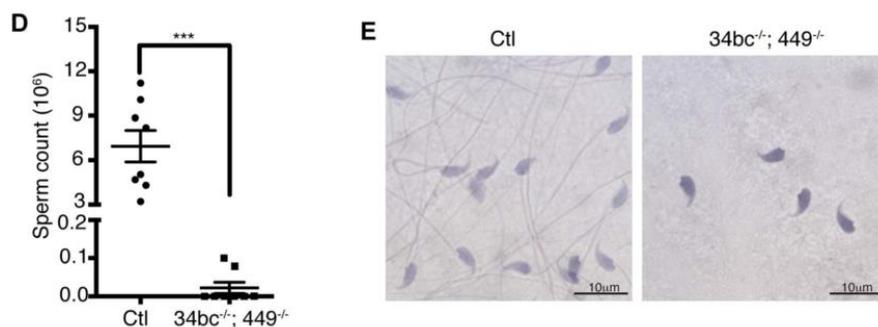


Figure 32 : Perturbation de la spermatogenèse chez les souris KO en miR-34bc et miR-449a<sup>76</sup>. Une délétion des miR-34bc et miR-449a entraîne une diminution très importante du nombre de spermatozoïdes produits (D) et une augmentation du nombre de spermatozoïdes décapités (E).

De façon générale, les miRNA interviennent dans la régulation de la différenciation des cellules souches spermatogoniales (CSS). Par exemple, miR-20, miR-21, miR-34c, miR-100, miR-106b, miR-146a, miR-182, miR-183, miR-222 et miR-383 sont préférentiellement exprimés dans les spermatogonies de type A. Chez la souris, miR-21 représente environ 11% de l'expression des miRNA dans les spermatogonies et son inhibition transitoire dans des spermatogonies en culture entraîne une augmentation de l'apoptose et donc une réduction du nombre de cellules<sup>173</sup>, démontrant son importance pour le maintien de la population des CSS. De même, miR-20, miR-100 et miR-106a se sont avérés essentiels pour le renouvellement des CSS via la répression de Stat3 et Ccnd1 qui régulent négativement la prolifération des CSS<sup>174,175</sup>. Un rôle de miR-146a et de miR-221 / miR-222 a également été montré dans le maintien du statut indifférencié des spermatogonies, via une modulation de l'activité de l'acide rétinoïque (inducteur de la différenciation des spermatogonies) pour miR-146a<sup>176</sup> ou la répression de cKIT pour miR-221 / miR-222<sup>177</sup>. Une expression de miR-34c a été détectée dans les CSS, où il participe à la régulation de la balance prolifération / différenciation des spermatogonies, via une action sur Nanos2<sup>178,179</sup> et favorisait indirectement le phénotype germinale des cellules déjà engagées dans cette lignée<sup>180,181</sup>.

De nombreux miRNA interviennent également dans la régulation des étapes méiotiques et la spermiogenèse<sup>17</sup>. L'augmentation de l'expression de miR-34c aux stades spermatocyte-pachytène et spermatide ronde peut favoriser le processus de méiose via la répression de Nanos2 et l'augmentation consécutive de l'expression de Stra8, un marqueur de l'entrée en méiose<sup>182</sup>. Mais miR-34c peut exercer son action via d'autres voies également, car Tgif2 (codant un inhibiteur de Tgfβ) et Notch2 sont des cibles directes du miR-34c<sup>181</sup> ayant un rôle important dans la spermatogenèse : Tgfβ inhibe la deuxième division méiotique de la spermatogenèse<sup>183</sup>, tandis que la signalisation Notch favorise la différenciation de la spermatogonie<sup>184</sup>. De façon intéressante, une baisse d'expression de miR-34c a été observée dans les tissus testiculaires de patients cryptorchides, suggérant l'implication des miRNA dans les mécanismes physiopathologiques associés<sup>178</sup>. Les miR-214 et miR-29 semblent également être impliqués dans la régulation de la méiose chez la souris<sup>185</sup>. Les miRNA interviennent par ailleurs lors de la transition histone-protamine, comme par exemple miR-469 et miR122a qui ont pour cibles des gènes codant pour des protéines de transition comme *TNP2*<sup>186,187</sup> (la protéine codée par ce gène permet le remplacement des histones par les protamines). Chez

la Drosophile, miR-8 contrôle l'expression testiculaire du gène codant la mitoferrine, un transporteur ayant un rôle dans le métabolisme mitochondrial du fer, dont l'absence entraîne une réduction forte du nombre de spermatozoïdes produits<sup>188</sup>.

#### 4.2) Rôle des piRNA

Les piRNA représentent la famille de sncRNA la plus exprimée dans la lignée germinale mâle. Plusieurs modèles de souris invalidées ont été créés (Miwi, Mili, Tdrd1, Tdrd9, Mvh, Mov10L1, Maelstrom, Gasz, Pld6, SPOCD1) pour étudier le rôle des protéines impliquées dans la biogenèse des piRNA et le fonctionnement des complexes PIWI, qui présentent tous des anomalies de la spermatogenèse<sup>16,134,137,140,189</sup>. Les phénotypes se manifestent principalement à deux stades distincts de différenciation spermatogénique, lors de la méiose des spermatocytes et au cours de la spermiogenèse ultérieure des spermatides.

Deux vagues d'expression des piRNA ont lieu dans les gamètes mâles. La première au stade foetal contribue au maintien de l'intégrité du génome des lignées germinales via l'expression de piRNA dérivés des rétrotransposons dans les *Primordial Germ Cells* (PGC) lors de la déméthylation massive du génome<sup>16,123,134</sup>. Afin de protéger ces lignées des effets délétères des éléments transposables, les piRNA associés aux protéines PIWI reconnaissent par complémentarité de séquence les éléments transposables pour les dégrader. De plus chez la souris, MIWI2 appose des marques de méthylation *de novo* sur les éléments transposables, via le recrutement de plusieurs facteurs tels que la DNMT3A, DNMT3L et le cofacteur SPOCD1<sup>190</sup> afin d'empêcher leur transposition.

Les souris KO montrent une élévation d'expression des transposons et une diminution de la méthylation *de novo* des transposons à l'issue de la déméthylation des PGC, au stade pro-spermatogonies. Mais le phénotype n'apparaît qu'au stade des spermatocytes, qui s'avèrent arrêtés au stade zygotène, avec des défauts d'appariement des chromosomes homologues et une fréquence élevée de cassures double brin de l'ADN. Plusieurs hypothèses ont été proposées pour expliquer ce phénotype : l'activité des transposons au stade précédent ; l'hypométhylation héritée des pro-spermatogonies peut engendrer une conformation de chromatine inhabituelle, qui pourrait conduire à des défauts de condensation chromosomique ou de séparation des chromosomes homologues<sup>189</sup>.

La seconde vague d'expression des piRNA a lieu durant la première division méiotique au stade pachytène. Au cours de cette phase, les rétrotransposons, en particulier LINE1, connaissent une reprise forte d'expression et d'activité s'ils ne sont pas sous contrôle des piRNA<sup>191</sup>. Chez les souris KO, l'hypométhylation héritée des pro-spermatogonies pourrait empêcher le contrôle de l'activité de ces rétrotransposons lors des premières phases de la méiose et contribuer ainsi un phénotype observé, d'autant que lors de cette seconde vague, les piRNA ne sont pas enrichis en séquences dérivés des rétrotransposons. Les souris KO *Tdrd5*<sup>-/-</sup> présentent des testicules plus petits et des tubules séminifères plus fins, contenant des spermatogonies et des spermatocytes, mais pas de spermatides allongées<sup>192</sup>. Cet arrêt de la spermiogenèse au stade spermatides rondes est également observé chez les mutants *Tdrd4*<sup>-/-</sup>, *Tdrd6*<sup>-/-</sup> et *Miwi*<sup>-/-</sup>. Chez la souris (MIWI) ou chez l'homme (HIWI), une mutation homozygote située dans le domaine D-Box et empêchant l'ubiquitination et la dégradation par le système ACP/C conduit à une stérilité mâle<sup>193</sup>. Les protéines PIWI mutées ne sont pas dégradées et s'accumulent dans le cytoplasme de la cellule, où elle séquestre RNF8, une protéine essentielle pour la transition histone-protamine, perturbant la spermatogenèse et conduisant à une stérilité.

Les mécanismes mis en jeu lors de cette seconde vague restent à clarifier puisque le seul rôle classique de blocage des rétrotransposons apparaît peu probable. Les travaux de Gou et., al<sup>135</sup> chez la souris, ont mis en lumière que les piRNA pouvaient être pris dans des complexes *pi-RISC* associés à la protéine CAF1. Ce complexe formé dans les spermatides en cours de spermiogenèse conduirait à la dégradation massive des ARNm contenus dans la cellule. Ces travaux apportent un nouvel éclairage sur les taux relativement faibles d'ARNm dans le spermatozoïde mature, sachant que l'hypothèse actuelle postule que la perte des ARNm est liée à l'éjection de la gouttelette cytoplasmique proximale au cours du transit épидидymaire.

## 5) Les petits ARN non codants et maturation de la semence

Le spermatozoïde post testiculaire étant considéré comme transcriptionnellement silencieux du fait de l'extrême compaction du noyau spermatique, les sncRNA présents dans le spermatozoïde épидидymaire ont été considérés initialement comme un héritage de la spermatogenèse. Mais cette hypothèse ne permet pas d'expliquer les évolutions de

composition du spermatozoïde observées durant les phases de maturation épидидymaires et post épидидymaires<sup>167</sup>. Le remplacement des piRNA par les tsRNA et, dans une certaine mesure, les miRNA, suggère des apports exogènes effectués par divers types cellulaires via des vésicules extracellulaires comme les épидидymosomes et les prostasomes.

### 5.1) Le contenu en sncRNA varie au cours de la maturation

Les spermatozoïdes épидидymaires contiennent, comme les spermatozoïdes testiculaires, une grande diversité de sncRNA, principalement les miRNA, piRNA, tsRNA et rsRNA. Les autres catégories de petits ARN dérivés d'ARN mitochondrial (mitosRNA) et snoRNA ont parfois été décrites, mais sont moins représentées<sup>194</sup>. Au-delà de cette diversité, le contenu en sncRNA présente une dynamique spatio-temporelle et évolue au cours du transit épидидymaire, présentant des caractéristiques spécifiques à chaque section de l'épididyme, que ce soit chez la souris ou l'homme. Les proportions relatives des différentes classes évoluent, de même que l'expression des sncRNA individuels au sein d'une même classe.

Globalement, le contenu en sncRNA, riche en piRNA à l'entrée dans l'épididyme, est profondément remanié et les tsRNA<sup>195</sup> et rsRNA<sup>153</sup> deviennent progressivement les sncRNA les plus abondants. Un enrichissement transitoire en snoRNA est observé dans les spermatozoïdes au niveau de la tête de l'épididyme<sup>194</sup>. Les miRNA ne représentent qu'une faible fraction des sncRNA du spermatozoïde épидидymaire, mais présentent d'importantes variations individuelles d'expression entre les différentes sections de l'épididyme<sup>196</sup>. Il est utile de rappeler que la comparaison entre les études et les sections est rendue compliquée, voire impossible, du fait des différences de pipelines bio-informatiques, excluant parfois certaines catégories de sncRNA et modifiant de ce fait les proportions entre classes. Ainsi, selon les études, le spermatozoïde murin épидидymaire peut contenir soit plus de 80% de tsRNA<sup>197</sup>, soit plus de 80% de rsRNA<sup>194</sup> ! Il a également été observé une tendance générale à l'accroissement de la taille des sncRNA au cours de la maturation, liée à l'émergence d'espèces de sncRNA plus grandes, telles les tsRNA et rsRNA-28S. Par ailleurs, une étude montre l'existence de piRNA au niveau de la queue de l'épididyme, produits vraisemblablement par la voie primaire, car les piRNA ne présentent pas d'enrichissement 10A<sup>198</sup>.

Les connaissances acquises dans les 10 dernières années indiquent que parallèlement aux modifications de leur membrane plasmique (lipides, protéines et glycans) et la poursuite de la condensation de la chromatine que subissent les spermatozoïdes pour acquérir leur pouvoir fécondant au cours du transit épидидymaire, les modifications substantielles de la teneur en sncRNA représentent également un événement biologique majeur de la maturation.

## 5.2) Origine des sncRNA du spermatozoïde épидидymaire

Compte tenu du caractère quiescent du noyau spermatique, les modifications de contenu en sncRNA sont probablement le résultat des interactions continues avec l'environnement luminal. Plusieurs études ont été menées afin de déterminer la composition de fluide luminal et de tester cette hypothèse. Ainsi, tout comme les protéines, les sucres ou les lipides, des sncRNA ont été identifiés dans la lumière de l'épididyme<sup>30,199,200</sup>. Une étude Canadienne<sup>199</sup> en 2013, s'est intéressée aux modifications du profil en miRNA du fluide de l'épididyme, à la suite d'une vasectomie chez l'Homme. Plus de 300 miRNA ont été identifiés chez les patients avant la chirurgie. Après la chirurgie, le profil des miRNA est modifié : environ une soixantaine de nouveaux miRNA sont observés (+17% de miRNA), suggérant une modification des profils d'expression sous l'influence de l'environnement systémique de l'épididyme. Cinq ans plus tard (2018), le nombre de miRNA identifiés dans la lumière de cet organe s'élève à 545 chez l'homme et 370 chez la souris<sup>30</sup>. Une portion importante (31%) de ces miRNA est conservée entre ces deux espèces, indiquant un rôle probablement important de ces éléments.

De nouvelles études ont révélé que les sncRNA n'étaient pas libres dans la lumière de l'organe, mais encapsulés dans des vésicules extracellulaires : les épидидymosomes<sup>201,202</sup>. Appartenant à la famille des exosomes, ces épидидymosomes sont des vésicules hétérogènes de 40 à 100 nm sécrétées par les cellules épithéliales bordant la lumière de l'organe. La distribution de taille des ARN transportés met en évidence deux pics majeurs chez l'Homme, indiquant qu'il s'agit de petits ARN : de 20 à 40 nucléotides et de 40 à 100 nucléotides<sup>44</sup>. Ces tailles d'ARN correspondent bien aux longueurs de plusieurs familles de sncRNA. Les analyses bio-informatiques, réalisées à partir de données obtenues principalement chez la souris (ou d'autre comme chez le ver), ont permis d'identifier des miRNA, des piRNA, des rsRNA et tsRNA<sup>30,203-206</sup>. De façon cohérente avec l'évolution des sncRNA du spermatozoïde

épididymaire (selon certaines études), les tsRNA représentent 87% du contenu des épididymosomes, majoritairement représentés par les 5' tRFs (tRF-Glu et tRF-Gly)<sup>207</sup>. Les piRNA semblent peu retrouvés dans ces vésicules. L'acquisition de ces séquences par le spermatozoïde dans la queue de l'épididyme pourrait donc relever d'un mécanisme différent ou d'une synthèse in situ<sup>198</sup>.

Le contenu des épididymosomes est le reflet du contenu de la cellule épithéliale d'origine, et il a bien été validé que les sncRNA transférés aux spermatozoïdes via les épididymosomes sont bien produits dans l'épithélium épididymaire<sup>208</sup>. Des variations d'expression des sncRNA contenus dans des épididymosomes ont été observées entre la tête et la queue de l'épididyme, conduisant à des profils de sncRNA « section-spécifique ». Concernant les miRNA chez l'homme, 16% d'entre eux ont été spécifiquement identifiés dans l'un des 3 compartiments de l'organe<sup>199,209</sup>. Au-delà de l'aspect « absence-présence », des variations d'expression de miRNA ont été observées suivant la localisation dans l'organe. Par exemple, les miR-204 et miR-196b sont différenciellement exprimés entre la tête et la queue de l'épididyme de souris avec un ratio proche de 40 fois<sup>209</sup>. Ces 2 miRNA font partie d'un plus grand ensemble de petits ARN non codants ayant une expression statistiquement différente entre les cellules épithéliales formant les zones de l'épididyme<sup>209</sup>. Les volcano plots présentés en figure 33, illustrent des miRNA différenciellement exprimés selon les deux parties de l'organe considérées.

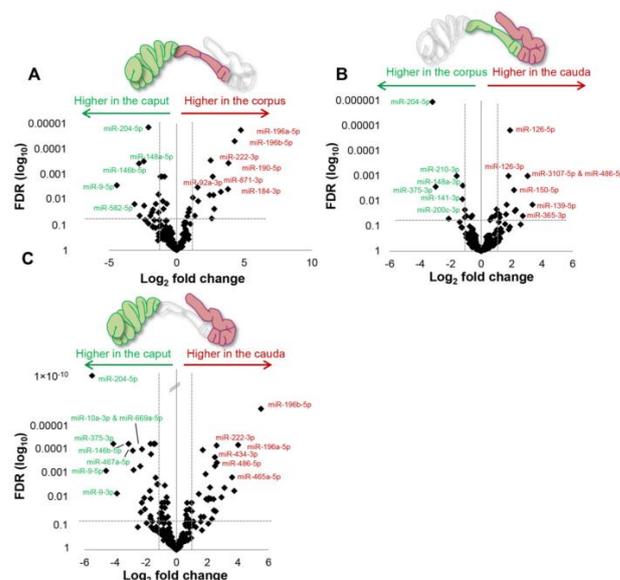


Figure 33 : Différence significative d'expression de certains miRNA en fonction de la section de l'épididyme<sup>209</sup>. Certains miRNA présentent une différence significative dans leur niveau d'expression en fonction de la section de l'épididyme (Tête vs Corps (A) ; Corps vs Queue (B) et Tête vs Queue (D)).

À la sortie de l'épididyme, le spermatozoïde rentre en contact avec le plasma séminal, qui contient également des vésicules extracellulaires appelées prostasomes. Quelques publications ont rapporté la présence de sncRNA au sein de ces prostasomes<sup>44,210,211</sup>. Chez l'Homme, plusieurs centaines de miRNA, piRNA, rsRNA et tsRNA ont été identifiés dans le contenu de ces vésicules extra cellulaires. Il est intéressant de noter que parmi les tRFs identifiés, l'acide aminé majoritairement représenté est encore la glycine<sup>44</sup> (23 à 45% de la diversité des tRFs selon les échantillons). Il se pourrait donc que ces vésicules, au même titre que les épидидyosomes, participent à la diversité du profil en sncRNA non codants du spermatozoïde éjaculé. Des variations de contenu en sncRNA ont également été observées dans les prostasomes de patients atteints d'azoospermie obstructive<sup>210</sup> ou d'oligoasthenozoospermie<sup>211</sup>. Par exemple, 40 miRNA dérégulés ont été identifiés dans le cas d'une oligoasthenozoospermie. La relation entre la perturbation du contenu en miRNA des prostasomes et les défauts du spermogramme n'est pas claire. Elle peut en être la cause ou juste un reflet de ce qui a conduit à une spermatogenèse anormale.

### 5.3) Rôles potentiels des sncRNA dans la maturation de la semence

L'acquisition du pouvoir fécondant des spermatozoïde est conditionné par les interactions avec l'environnement luminal, qui permettent la maturation de la semence et les modifications de structure associées via l'apport de nouveaux composants. Des exemples d'acquisition dans l'épididyme de protéines ou de lipides, contribuant à la maturation du spermatozoïde ont été publiés<sup>212</sup>. Par exemple, la cenexine, une protéine indispensable à la maturation des centrosomes, est produite par les cellules épithéliales de l'épididyme. L'augmentation de son expression dans le spermatozoïde au cours du transit épидидymaire résulte d'un transfert *via* les épидидyosomes<sup>38</sup>.

Il est tentant de spéculer sur un rôle similaire des sncRNA dans la maturation post testiculaire du spermatozoïde. Or, la plupart des sncRNA remplissent leurs fonctions via la modulation de l'expression de leurs gènes cibles, au niveau post-transcriptionnel ou traductionnel. Comme le spermatozoïde est transcriptionnellement inactif, une telle action ne pourrait s'envisager que sur le pool de transcrits déjà présent. Elle supposerait également que la traduction des protéines dans le spermatozoïde épидидymaire soit active. Le paradigme actuel établit

qu'aucune synthèse protéique n'a lieu dans le spermatozoïde, mais quelques études controversées montrent une incorporation d'acides aminés marqués dans les polypeptides pendant la phase de capacitation<sup>213-215</sup>. Cette activité de traduction dans le spermatozoïde éjaculé, situé dans les voies femelles, dépendrait exclusivement de la machinerie de traduction mitochondriale et concernerait des gènes nucléaires plutôt que les seuls gènes mitochondriaux. Si un tel mécanisme existe pour le spermatozoïde éjaculé, nous pourrions imaginer que cela soit également possible pour un spermatozoïde en transit dans l'épididyme. De fait, le rôle dans la maturation de la semence des sncRNA n'est pas encore élucidé et à notre connaissance, aucun article n'a encore été publié, qui démontre l'implication des sncRNA acquis via les épидидyosomes et les prostasomes dans le processus de maturation ou détaille les mécanismes en jeu.

#### 6) Les sncRNA du spermatozoïde support de la transmission entre générations

Le spermatozoïde a longtemps été considéré comme un « taxi », transportant et protégeant le patrimoine génétique du père. Cette idée venait du fait que le spermatozoïde bien plus petit que l'ovocyte après avoir éliminé une très large partie de son cytoplasme, ne pouvait léguer que son ADN hyper condensé. Mais les nombreux travaux concernant l'acquisition de sncRNA par le gamète mâle au cours de la maturation obligent à revoir ce dogme. En effet, comme indiqué précédemment, les épидидyosomes et prostasomes transfèrent leur cargaison de sncRNA aux spermatozoïdes au cours de la maturation. Ces sncRNA synthétisés par l'épididyme et la prostate sont donc délivrés au zygote et peuvent constituer une forme d'information épигénétique, potentiellement impliquée dans la régulation du développement embryonnaire, voir dans le développement à long terme des descendants. Les sncRNA pourraient ainsi être un des supports de l'hérédité non génétique, qui assure le transfert de la mémoire de l'exposition à divers environnements du père (altérations métaboliques, polluants, stress, certains comportements...), conditionnant la réactivité des descendants à divers environnements au cours de leur vie<sup>154,167,216-223</sup>.

Cette hypothèse est d'autant plus séduisante que l'épididyme constitue en quelque sorte une « fenêtre ouverte » sur le monde extérieur, capable de traduire les perturbations rencontrées

en signaux interprétables par le spermatozoïde ou le futur zygote. En effet, la barrière hémato-testiculaire formée par les cellules de Sertoli, joue un rôle majeur dans la protection des gamètes au cours de la spermatogenèse : le spermatozoïde est isolé du reste de l'organisme. Au niveau de l'épididyme, organe très vascularisé, les spermatozoïdes sont isolés de l'environnement par les cellules de l'épithélium, dont le fonctionnement est largement conditionné par l'environnement via la circulation générale. Par exemple, le contenu en sncRNA (plus particulièrement les tsRNA et les miARN) des épидидyosomes, reflet de la physiologie des cellules de l'épithélium, varie grandement en fonction du régime alimentaire ou du stress<sup>203,224</sup>. La sécrétion dans la lumière de molécules de diverses natures produites par les cellules épithéliales peut ainsi constituer un mécanisme de « mise à jour » des spermatozoïdes, pour prendre en compte les événements marquants rencontrés par l'organisme.

#### 6.1) Rôles potentiels des sncRNA dans le développement embryonnaire

L'importance des sncRNA de la semence pour le développement embryonnaire a été illustrée dès 2016 chez la souris en réalisant des expériences de fécondation in vitro (FIV) avec de la semence issue lignée de souris KO conditionnel pour DICER ou DROSHA dans la lignée germinale<sup>225</sup>. Les mâles de ces lignées présentent une oligoasthénotérazoospermie, nécessitant un recours à l'ICSI (intra cytoplasmique sperm injection). Le contenu en miRNA de leurs spermatozoïdes est profondément remanié, avec en particulier un nombre de miRNA exprimés très réduit (moins de 400). Dans les deux cas, le pourcentage de blastocyste obtenu après ICSI est nettement diminué et le nombre de nouveaux nés obtenus après réimplantation est très faible. D'autres mutations de DICER ont également été associées à un retard de développement embryonnaire et une mortalité fœtale (figure 34) liée à un défaut d'angiogenèse, confirmée par la dérégulation de plusieurs gènes clés de ce processus biologique (*vegf*, *flt1*, *kdr* et *tie1*)<sup>226</sup>.

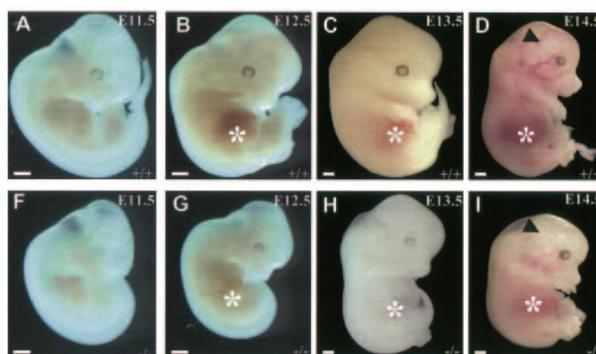


Figure 34 : Retard du développement embryonnaire murins KO pour le gène *Dicer*<sup>226</sup>  
 Les embryons normaux (A à D) et déficients en *Dicer* (F à I) présentent pour les mêmes stades des différences de développement : ils sont plus petits et sont moins vascularisés.

Plus récemment, Conine et al.<sup>227</sup> ont comparé les résultats de FIV réalisées chez la souris avec des spermatozoïdes de la tête et de la queue de l'épididyme des mêmes animaux. Aucune différence n'a été observée en termes de morpho-cinétique de développement jusqu'au stade blastocyste (J7). Cependant, une surexpression d'une cinquantaine de gènes a été observée dès le stade 4 cellules chez les embryons produits avec les spermatozoïdes de la tête de l'épididyme. Après réimplantation, aucun nouveau-né n'a pu être obtenu avec les embryons générés à partir de spermatozoïdes de la tête de l'épididyme. L'incubation des spermatozoïdes de la tête de l'épididyme avec des épидидyosomes récoltés dans la queue de l'organe permet de restaurer un phénotype normal, prouvant l'importance des sncRNA acquis en fin de transit épидидymaire pour le développement embryonnaire.

De même des expériences d'injection de miRNA et de leurs inhibiteurs dans le cytoplasme de zygotes de souris ont été réalisées pour analyser l'impact possible de ces sncRNA contenus dans les épидидyosomes. Ainsi, l'injection d'anti-miR-21 réduit de moitié le taux de développement jusqu'au stade 8 cellules (35% d'embryons à 8 cellules après injection vs 70% pour les contrôles)<sup>228</sup>. L'injection chez le porc d'un anti-miR-29b provoque un blocage du développement de l'embryon au stade blastocyste (18% de blastocystes vs 75% pour le témoin)<sup>229</sup>, associé à une augmentation du niveau de méthylation du génome liée à une surexpression des DNMT3a/3b et TET1.

Chez le bovin, une sous expression spermatique de miR-216b a été détectée chez des taureaux parmi les plus fertiles, ainsi que dans les embryons au stade 2 cellules produits avec leur semence, tandis qu'une surexpression de *K-RAS*, l'une des cibles de miR-216b, était

observée<sup>230</sup>. Par ailleurs, un taux de clivage plus élevé et une augmentation du nombre de cellules au stade blastocyste ont été observées chez les embryons produits avec cette semence. *K-RAS* étant impliquée dans la prolifération cellulaire, les auteurs font l'hypothèse que miR-216b transmis par les spermatozoïdes modulerait l'expression de *K-RAS* dans le zygote et régulerait le développement précoce en interférant avec le premier clivage et la qualité du blastocyste.

Les miRNA constituent actuellement la classe la plus étudiée, mais un rôle des autres sncRNA est vraisemblable, en particulier compte tenu des nouvelles connaissances sur les complexes RISC et leur association possible avec les tsRNA, rsRNA et piRNA. Un exemple récent d'implication de tsRNA spermatiques dans les premiers clivages de l'embryon porcin a été publié récemment. Des séquences anti-sens de tsRNA (tRF-Gln-TTG) ont été micro-injectées dans des zygotes, provoquant une réduction du taux de clivage et de développement des embryons dès le stade 2 cellules<sup>231</sup>, probablement via la régulation de gènes ou de rétrotransposons associés au cycle cellulaire.

Ces quelques exemples parmi d'autres (pour revue<sup>232</sup>) illustrent le rôle essentiel des sncRNA dans le développement normal de l'embryon, et confortent l'idée que les sncRNA de la semence puissent contribuer significativement au développement de l'embryon préimplantatoire.

## 6.2) Rôles potentiels des sncRNA dans la transmission intergénérationnelle

Au-delà du rôle sur le développement précoce de l'embryon se pose la question d'un impact à plus long terme sur les descendants. Un nombre grandissant d'études concerne cette question. Partant de l'observation qu'un régime paternel restrictif en protéines induisait une perturbation métabolique, en particulier une altération de la biosynthèse du cholestérol des cellules hépatiques dans la descendance chez la souris, Sharma et al. ont analysé la diversité du contenu en sncRNA codants des gamètes<sup>207,233</sup>. L'expression de tsRNA (tRF-Gly-CCC, Gly-TCC et Gly-GCC), apportés par les épидидymosomes, s'est avérée statistiquement augmentée par le régime, alors que celle des miRNA de la famille des let-7 était diminuée. Ces résultats suggèrent un lien entre ces sncRNA et le phénotype, qui reste toutefois à démontrer formellement.

De même, chez le rat<sup>234</sup> ou la souris<sup>235</sup>, les effets d'un régime paternel hyperlipidique sur le métabolisme des descendants ont été étudiés afin de relier le phénotype des générations suivantes avec les modifications engendrées dans les spermatozoïdes des pères. La génération F1 a présenté des symptômes d'ordre métabolique à la naissance et à l'âge adulte (problème de poids, réduction des cellules bêta pancréatiques, intolérance au glucose et résistance à la perte de poids). Les spermatozoïdes des rats et des souris ayant subi le régime hypergras, présentaient une modification d'expression de plusieurs centaines de sncRNA, dont des miRNA, des tRFs et des piRNA. Chez les rats de la génération F1 nés de pères F0 ayant subi le régime obésogène, le profil en sncRNA des spermatozoïdes présentent également une dérégulation des certains miRNA, ARNr et snoRNA. Seuls quelques piRNA et le miRNA let-7c ont montré une dérégulation significative chez les pères et leur fils. Parmi les cibles de let-7c, certaines sont impliquées dans le métabolisme du glucose ou dans les processus biologiques impliquant l'insuline, tels que AKT2, IGF2R ou UCP2<sup>234</sup>. Une dérégulation de l'expression de ces gènes pourrait être à l'origine du phénotype des générations F1 issues des pères ayant subi le régime obésogène. Chez la souris, après la réimplantation d'embryons micro-injectés avec le miR-19b (l'un des plus dérégulés dans la semence des pères), le poids à la naissance était altéré, tout comme pour les souris nées de l'accouplement entre un père « hypergras » et une mère contrôle<sup>235</sup>. Les épидидyosomes n'ont pas été étudiés ici, nous ne pouvons donc pas affirmer que ces modifications de profils en sncRNA proviennent des exosomes de l'épididyme. Néanmoins, les spermatozoïdes étudiés venant de la partie caudale de l'organe et en tenant compte des études précédentes, nous pouvons supposer un apport des épидидyosomes.

Il a également été rapporté chez la souris que les sncRNA pouvaient être impliqués dans la transmission de troubles du comportement. Par exemple, une exposition à un stress dans le jeune âge provoque une perturbation d'expression des miRNA spermatiques chez la souris et une altération des réponses comportementales et métaboliques chez les descendants<sup>216</sup>. L'injection d'ARN spermatiques de mâles stressés dans des zygotes contrôles permet de reproduire les altérations comportementales et métaboliques dans la descendance, suggérant un rôle des sncRNA dans cette transmission. De même, la transmission de dérégulations hypothalamo-hypophysaires induites par un stress a été associée à un jeu de 9 miRNA spermatiques<sup>216,236</sup>.

## Chapitre III : Objectifs de la thèse

Le rôle des sncRNA de la semence revêt un intérêt scientifique indéniable, comme en témoignent les nombreux travaux déjà réalisés chez l'homme ou les organismes modèles (souris, rat, drosophile, *C. elegans*). La connaissance détaillée des mécanismes chez les bovins pourrait trouver des applications dans le domaine du contrôle qualité de la semence pour garantir une bonne fertilité des taureaux et de leurs éjaculats, mais également dans le domaine de la gestion et de l'alimentation des taureaux ou de leurs mères au cours de la gestation, afin de garantir l'expression du potentiel génétique de la descendance.

Les connaissances chez les bovins restent parcellaires et dans l'idée de mener une recherche finalisée, il était essentiel d'acquérir les connaissances académiques et d'identifier des biomarqueurs potentiels dans cette espèce.

Aussi, trois objectifs ont été fixés pour cette thèse, menée en collaboration ALLICE-INRAE dans le cadre du LabCom SeQuaMol :

- (i) La description la plus exhaustive possible du contenu en sncRNA du spermatozoïde bovin éjaculé. À l'initiation de ce travail, seules quelques publications rapportaient l'identification de miRNA dans le spermatozoïde bovin.
- (ii) L'étude de l'origine de ces sncRNA (origine épидidymaire ou testiculaire).
- (iii) L'identification de sncRNA biomarqueurs de la fertilité mâle,

### 1) Objectif 1 : catalogue exhaustif du contenu en sncRNA du spermatozoïde bovin

Cet objectif se décompose en plusieurs tâches. La première a consisté à mettre au point un protocole robuste d'extraction des sncRNA du spermatozoïde bovin, à partir d'un grand nombre d'individus de différentes races, suivi d'une analyse NGS afin d'obtenir un catalogue le plus exhaustif possible. La seconde a porté sur l'annotation des séquences obtenues. En effet, les pipelines bio-informatiques habituellement utilisés se concentrent uniquement sur l'identification des miRNA. Or, nous souhaitons analyser l'ensemble des classes de sncRNA, qui pourraient toutes présenter un intérêt comme source de biomarqueurs. Par ailleurs, il

sous semblait intéressant d'analyser la variabilité des séquences de miRNA, en étudiant les « isomiR », plutôt que de se focaliser uniquement sur la séquence canonique. Une collaboration avec Sylvain Marthey (UMR GABI, INRAE) a permis de développer le pipeline bio-informatique adéquat.

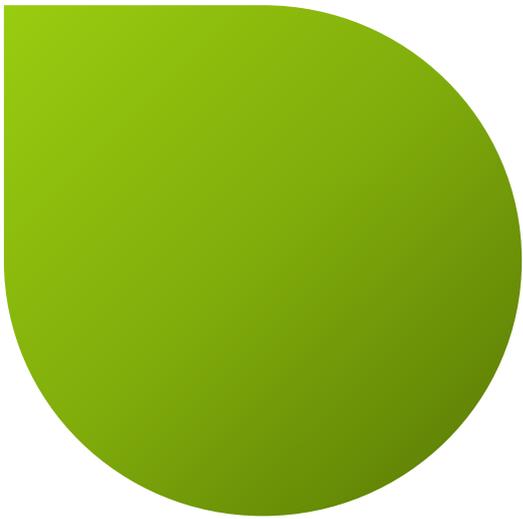
## 2) Objectif 2 : L'origine des sncRNA

Cet objectif a consisté à déterminer l'origine des sncRNA spermatiques en établissant en parallèle, à partir des mêmes animaux, les répertoires de sncRNA des spermatozoïdes testiculaires, épидидymaires (tête, corps et queue) et éjaculés. L'identification de l'origine des sncRNA permettra d'émettre des hypothèses quant à leur rôle dans la spermatogenèse, la maturation finale du spermatozoïde ou le développement embryonnaire.

## 3) Objectif 3 : biomarqueurs de la fertilité mâle

Il consiste à établir et analyser une cohorte importante de taureaux, incluant des animaux des races Holstein et Montbéliarde présentant une fertilité extrême, afin d'analyser le contenu en sncRNA de leur semence et d'identifier des biomarqueurs liés significativement à la fertilité des taureaux.

Ce travail de thèse s'inscrivant dans le contexte plus général du LabCom, une analyse d'autres marques épigénétiques (modifications histones/protamines, méthylation) et d'autres omics (protéomique, lipidomique, glycomique) est menée en parallèle sur les mêmes échantillons. Le but sera *in fine* d'intégrer l'ensemble de ces données, en collaboration avec l'équipe GIBBS (UMR GABI, INRAE), afin de mettre au point un prédicteur parcimonieux de la fertilité des taureaux, permettant une utilisation en routine au meilleur rapport coût/bénéfice.



# Résultats

# Chapitre I : le catalogue en sncRNA du spermatozoïde bovin éjaculé

## 1) Éléments de contexte

Le spermatozoïde a longtemps été considéré comme un taxi, ayant comme unique rôle le transport du patrimoine génétique mâle. Mais depuis quelques années, des travaux démontrent qu'il influence également le développement de l'embryon. En effet, l'ADN n'est pas le seul acide nucléique déversé dans l'ovocyte lors de la fécondation. Des petits ARN non codants, tels que les miRNA, les piRNA ou des fragments d'ARNt ou d'ARNr sont également transmis. Ces éléments ont la capacité d'influencer l'expression des gènes, en régulant par exemple la traduction des ARNm en protéine. Leur présence dans le spermatozoïde mature, transcriptionnellement éteint, pose donc question. Sont-ils des vestiges de la spermatogenèse ou constituent-ils des éléments de régulation du développement précoce de l'embryon, comme suggéré par de nombreux travaux chez la souris ? Qu'elles témoignent d'une altération de la spermatogenèse ou qu'elles impactent le développement de l'embryon, dans quelle mesure leurs variations sont-elles associées à la fertilité ?

Afin de pouvoir à terme répondre à ces questions chez le bovin, où peu d'informations sont disponibles, nous avons entrepris d'établir le catalogue exhaustif du contenu en sncRNA du spermatozoïde bovin éjaculé et préparé pour l'IA.

## 2) Résumé de l'étude

L'analyse de la diversité et de l'expression des sncRNA a tout d'abord nécessité l'optimisation du protocole d'extraction d'ARN total à partir de 40 millions de spermatozoïdes congelés (2 doses d'IA) de 40 taureaux de fertilité normale représentant 6 différentes races. L'objectif étant l'établissement d'un catalogue le plus exhaustif possible, le séquençage a été réalisé à forte profondeur, avec une cible à 40M de reads par échantillon.

L'analyse du contenu du spermatozoïde bovin mature éjaculé confirme certaines observations faites dans d'autres espèces, comme la présence de différentes familles de sncRNA (26% de piRNA, 25% de rsRNA, 20% de miRNA et 14% de tsRNA), ou la diversité de séquences de miRNA trouvée dans le spermatozoïde bovin (recouvrement important du nombre de miRNA avec l'étude de Capra<sup>237</sup> ou de Stowe<sup>238</sup> quand on s'intéresse aux séquences les plus exprimées). En revanche de nouvelles informations ont été apportées par cette étude, comme (i) l'extension du répertoire de miRNA connus et l'identification de plus de 2000 nouveaux miRNA prédits dans le spermatozoïde bovin, (ii) un répertoire de séquences éditées (isomiR), provenant principalement d'addition ou de délétion d'un nucléotide (A ou U) en position 3' ou encore (iii) les corrélations d'expressions entre miRNA formant des réseaux de régulation. De façon intéressante, l'analyse de la fréquence d'édition en fonction de la position des nucléotides des isomiR a permis de mettre en lumière que la région 3p (à partir du 19<sup>e</sup> nucléotide) est nettement plus éditable que le reste de la séquence et que le premier nucléotide de la « Seed » (en position 2) est plus variable que le reste des nucléotides de la « Seed » (du 3<sup>e</sup> au 8<sup>e</sup>).

Les autres familles de petits ARN non codants présentent également des particularités intéressantes. C'est le cas pour les tRNA, qui sont majoritairement présents sous forme de fragments (tsRNA). Que ce soit en termes d'isoaccepteur ou de type de fragments générés, le spermatozoïde bovin présente des particularités d'expression : La glycine et la glutamine représentent à elles seules plus de 70% des isoaccepteurs et le 5'-half-tsRNA plus de 57% des tsRNA présents. La moitié de nos séquences de piRNA (54%) est totalement complémentaire des séquences de référence. L'autre moitié présente des modifications de séquence, majoritairement une délétion d'une base en 5p. Ces séquences semblables aux piRNA peuvent être assimilées à des isopiR.

Un nombre important de séquences, quelle que soit la famille de sncRNA considérée, est différenciellement exprimé selon la race. L'établissement de ces catalogues fournit une base solide pour identifier ultérieurement des biomarqueurs du pouvoir fécondant de la semence.

### 3) Article publié dans *Epigenetics & Chromatin* en 2020

#### **A comprehensive overview of bull sperm-borne small non-coding RNAs and their diversity across breeds**

Eli Sellem<sup>1\*</sup>, Sylvain Marthey<sup>2</sup>, Andrea Rau<sup>2</sup>, Luc Jouneau<sup>3</sup>, Aurelie Bonnet<sup>1</sup>, Jean-Philippe Perrier<sup>3</sup>, Sébastien Fritz<sup>1,2</sup>, Chrystelle Le Danvic<sup>1</sup>, Mekki Boussaha<sup>2</sup>, Hélène Kiefer<sup>3</sup>, Hélène Jammes<sup>3</sup>, Laurent Schibler<sup>1</sup>

<sup>1</sup> R&D Department, ALLICE, 149 rue de Bercy, 75012 Paris, France.

<sup>2</sup> UMR1313 GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.

<sup>3</sup> UMR1313 BDR, INRA, ENVA, Université Paris Saclay, 78350, Jouy en Josas, France,

\*Corresponding author

E-mail: eli.sellem@alice.fr

#### **Abstract**

Mature sperm carry thousands of RNAs, including mRNAs, lncRNAs, tRNAs, rRNAs and sncRNAs, though their functional significance is still a matter of debate. Growing evidence suggests that sperm RNAs, especially sncRNAs, are selectively retained during spermiogenesis or specifically transferred during epididymis maturation, and are thus delivered to the oocyte at fertilization, providing resources for embryo development. However, a deep characterization of the sncRNA content of bull sperm and its expression profile across breeds is currently lacking. To fill this gap, we optimized a Guanidium-Trizol total RNA extraction protocol to prepare high quality RNA from frozen bull sperm collected from 40 representative bulls from six breeds. Deep sequencing was performed (40M single-50bp reads per sample) to establish a comprehensive repertoire of cattle sperm sncRNA.

Our study showed that it comprises mostly piRNAs (26%), rRNA fragments (25%), miRNAs (20%) and tRNA fragments (tsRNA, 14%). We identified 5p-halves as the predominant tsRNA

subgroup in bull sperm, originating mostly from Gly and Glu isoacceptors. Our study also increased by ~50% the sperm repertoire of known miRNAs and identified 2022 predicted miRNAs. About 20% of sperm miRNAs were located within genomic clusters, expanding the list of known polycistronic pri-miRNA clusters and defining several networks of co-expressed miRNAs. Strikingly, our study highlighted the great diversity of isomiRs, resulting mainly from deletions and non-templated additions (A and U) at the 3p end. Substitutions within miRNA sequence accounted for 40% of isomiRs, with G>A, U>C and C>U substitutions being the most frequent variations. In addition, many sncRNAs were found to be differentially expressed across breeds.

Our study provides a comprehensive overview of cattle sperm sncRNA, and these findings will pave the way for future work on the role of sncRNAs in embryo development and their relevance as biomarkers of semen fertility.

## **Keywords**

sncRNA, miRNA, isomiR, piRNA, tRNA, sperm, bull

## **Background**

Sperm is classically seen as a transcriptionally inactive vehicle (1) that delivers the paternal haploid genome to the oocyte. Yet, an increasing number of studies in several species have shown that mature ejaculate sperm carry thousands of RNAs, including messenger RNAs (mRNAs), long non-coding RNAs, transfer RNAs (tRNAs), ribosomal RNAs (rRNAs) and small non-coding RNAs (sncRNAs), especially microRNAs (miRNA), antisense RNAs, and Piwi-interacting RNAs (piRNA). Since transcription is mostly silenced in sperm, these RNAs are often considered to be remnant transcripts, produced during previous spermatogenesis stages (2-4). However, a comparison of sperm and testes in horses instead suggested that the repertoire of sperm RNAs is not a random spermatogenesis remnant, but rather a selectively retained and functionally coherent collection of RNAs (5). In addition, some studies in mouse have provided evidence for sncRNA regulation as sperm mature along the epididymis (6, 7), with sncRNA being transferred to sperm during post-testicular maturation, possibly through epididymosomes (8). The functional significance of sperm sncRNAs, if any, is still a matter of

debate, and future studies are required to shed light on their potential biological role. However, there is now some evidence of a role in fertilization and embryo development. Indeed, mouse sperm-borne RNAs are delivered to the oocyte at fertilization (9) and may provide essential resources for the first steps of embryo development (10). In agreement with this hypothesis, sperm-borne miRNA-34c has been shown to be delivered into the mouse zygote during fertilization and play a role in the first cell division (11). More generally, paternal miRNAs and/or endo-siRNAs have been identified as crucial for fertilization and preimplantation embryonic development (12) and post-natal outcome (13). In addition, involvement of sperm sncRNA has been demonstrated in epigenetic inheritance, including miRNAs (14, 15) as well as piRNAs and tRNAs (16, 17).

Unfortunately, the study of sperm sncRNAs is challenging due to technical issues related to sperm RNA isolation, including low RNA content and resistance to lysis by detergents (9, 18). Additionally, differences in sperm morphology and chromatin condensation may impair the interspecies use of sperm RNA isolation protocols. Consequently, highly variable sncRNA content has been obtained using different RNA extraction procedures, and until very recently only a small number of comprehensive studies were conducted. For instance, SpermBase was established in 2016 to publish RNA-seq data on mRNAs and sncRNAs in mouse, rat, rabbit, and human total sperm as well as sperm heads (19). Among sncRNAs, 18S and 28S rRNA fragments were reported to be abundant in sperm (20), as were tRNA-Derived Fragments (tsRNAs) (21). Sperm tsRNAs are mainly fragments of the 5' end of tRNA, ranging in length from 20 to 35nt, and are produced by specific cleavage of tRNA at either D-Loop, Anticodon-Loop or T-Loop by RNase Z, Dicer or Angiogenin (21, 22). They were recently proposed to mediate, at least in part, transgenerational effects associated with parental diet (16, 23). In addition, sperm tsRNAs are subjected to numerous RNA modifications and edits that contribute to their stability (for review see 24). Detailed studies of the sperm miRnome have also been undertaken in several species, including humans (25), mice (26, 27), horses (28), pigs (29) and cattle (30, 31). Extensive variation in length and sequence composition of miRNAs were detected by deep sequencing. These so-called isomiRs are thought to be produced via imprecise and alternative cleavage by Drosha or Dicer along the miRNA biogenesis pathway (32) and do not represent transcription or sequencing errors (33). Shortening of miRNAs may also arise by exonuclease trimming of the end (34). Conversely, the post-transcriptional

addition of one or more nucleotides by template-independent nucleotidyl transferases may increase miRNA length (35). In particular, seven nucleotidyl transferases have been implicated in human isomiR biogenesis, especially uridylyltransferases and adenylyltransferases (36), including PAPD4 which is considered to be the primary miRNA adenylyating enzyme (37). Polymorphisms within the internal canonical sequences have also been described in isomiRs expressed at low levels (38), resulting either from genetic polymorphism or from A-I editing, i.e. the hydrolytic deamination of adenosine to inosine in double-stranded RNA (36). A growing body of evidence indicates that these changes may affect isomiR stability *or* influence *target selection* (36, 37). IsomiRs have been shown to be non-randomly distributed and expressed in a developmental- and tissue-specific manner in several human tissues (39). Their broader biological significance is yet to be fully resolved, but some studies suggest that alteration in isomiR profiles, rather than in the overall miRNA abundance, is of biological significance, correlating for instance with cancer progression (40). In cattle, several miRNAs have been shown to be differentially expressed between high- and low-fertility bulls (41, 42), as well as between high- and low-motility sperm (30).

This study was thus designed to unravel the sncRNA content from frozen bull sperm, focusing on tsRNAs, miRNAs, and their structural variations (called "isomiRs" by Morin et al (43)), as well as the diversity of their expression profiles according to breeds. To do so, we optimized a Guanidium-Trizol total RNA extraction protocol for frozen bull sperm to ensure good sample quality and reproducibility.

## Results

### *RNA extraction and NGS sequencing*

Selected samples had normal sperm quality profiles, consistent with routine results obtained in semen production centers (data not shown). Contamination with somatic cells was negligible, as confirmed by microscopy (less than 1 somatic cell per 1000 sperm cells) and side-scatter profiles obtained by flow cytometry (data not shown). In addition, no RNAs could be extracted from the extender used by semen production centers to dilute the sperm cells.

Altogether, these results indicate that the identification of sperm sncRNA content is not jeopardized by contaminants.

On average,  $57.6 \pm 12.9$  ng of total RNA could be prepared from  $31.2 \pm 7$  million thawed sperm cells. Technical variation was evaluated based on multiple analyses of a standard ejaculate, showing good reliability with a coefficient of variation less than 10% in the amount of total RNA prepared using the protocol. Typical electrophoretic profiles were obtained, with no evidence of 18S and 28S rRNAs. RT-qPCR was performed to validate RNA quality and concentrations on a subset of 3 samples for each extraction batch, leading to consistent amplification results (Ct in the range of 20-21 starting from 5 ng of total RNA) and single peak melt curves, indicating that a single, specific product had been produced (Additional Fig S1).

Sequencing resulted in 1 444 722 663 total raw sequence reads. About 36.1 million reads were obtained on average ( $\pm 4.5$  million, min=28.3 and max=49.5) for the 40 libraries. The overall data quality was good, with >97% of the data having a Q-score over 30 (i.e., a base-calling accuracy of 99.9%). Typical examples of FastQC quality control plots are provided in Additional Fig S2a. After trimming sequencing adapters, read length distribution revealed two main peaks representing mostly microRNAs or siRNAs (~18-26nt) and intermediate sequences (28-32nt), including piRNAs and tsRNAs (Additional Fig S2b). About 72% of sequences could be mapped unambiguously to the cattle reference genome on average, while 15% were outmapped (i.e., mapped to abundant sequences such as polyA and polyC homopolymers or other repetitive sequences) and 13% were unmapped.

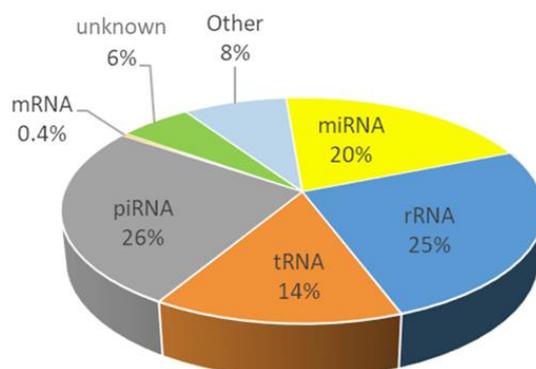
**Table 1. Mapping statistics across the 40 libraries.**

	Mapped (72%)	Outmapped (15%)	Unmapped (13%)	Reads
Reads no.	1 038 383 839	223 342 595	181 996 229	1 443 722 663
Mean	25 959 596	5 583 565	4 549 906	36 093 067
Std. deviation	3 542 949	1 932 607	4 370 783	4 483 073

*The overall small RNA content of bull sperm*

Reads were analyzed and categorized as described in the Material & Methods. First, 701 known and 2022 putative miRNAs were identified using miRDeep2 (Additional Table S1). The

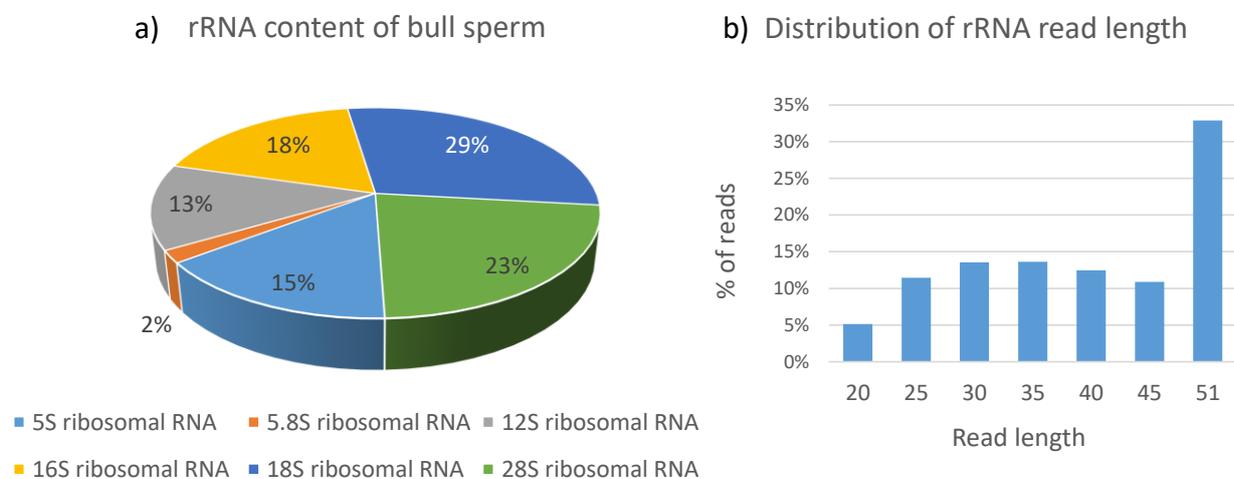
remaining reads were then annotated using several mRNA and sncRNA databases, highlighting the diversity of sncRNA in cattle sperm (Additional Tables S2-S6). As summarized in Fig 11, sperm sRNA contains mainly piRNAs (26%), rRNAs (25%), miRNAs (20%), tRNAs (14%), other sncRNAs (8%, including mainly Signal Recognition Particle (SRP) RNA, fragments of lncRNAs, Y\_RNA and snRNA) and 6% of unknown sequences. In addition, about 27655 reads (with a majority being 50nt in length) were identified as mRNA fragments originating from 3510 genes, of which 95% were already described in human or mouse sperm (<http://spermbase.org>). The gene coverage was usually low ( $7.7 \pm 3.1$  different unique reads per gene on average, preferentially clustered at the 5p end of the gene), but a few genes were found to be covered by well-distributed unique reads (37 genes with more than 100 different reads) and have high read counts (see for instance the *AKAP1* and *PRM1* IGV profiles in Additional Fig S3). A gene ontology analysis of these genes revealed several relevant enriched biological terms, including male gamete generation (GO:0048232), spermatogenesis (GO:0007283), and gamete generation (GO:0007276).



**Fig 14. sRNA content of bull sperm.** miRDeep2 was used to identify miRNAs, and the remaining reads were annotated (e.g. as rRNA, tRNA, piRNA) using several sncRNA databases. The percentages of total number of reads for each category are represented as a pie chart.

### *Bull sperm contains abundant rRNA-derived small RNAs*

About 25% of reads were annotated as rRNAs, mainly the 16S (18%), 18S (29%) and 28S (23%) rRNAs, as illustrated in Fig 2a. Though highly abundant, these rRNAs appeared to be fragmented, with two-thirds being less than 45nt in length (Fig 2b). In addition, the distribution of reads along rRNA transcripts revealed a series of peaks and read-poor sub-regions, especially for 18S and 28S rRNAs (Additional Fig S4).

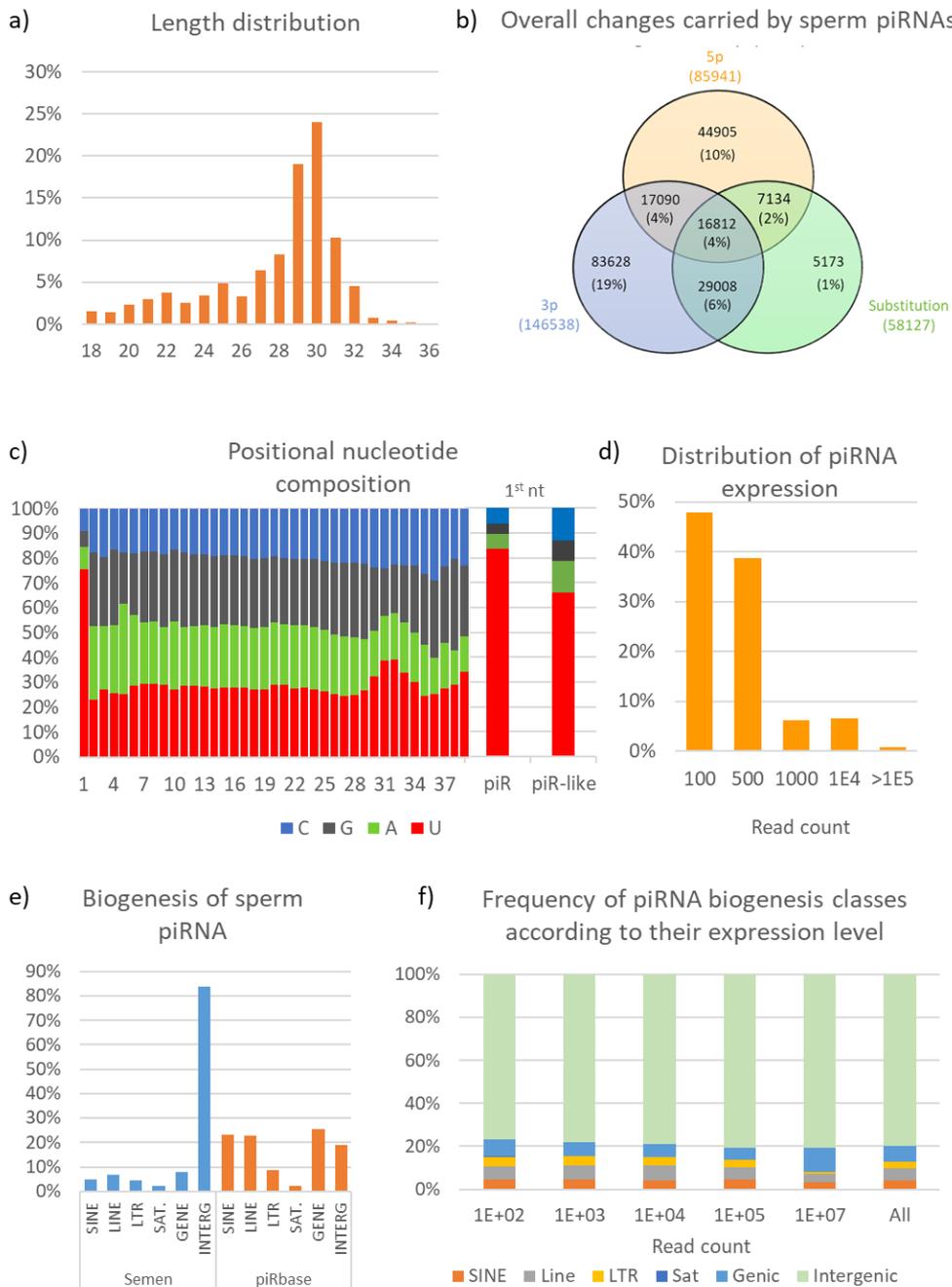


**Fig 215. Bull sperm rRNA content.** The proportion of reads annotated for different rRNAs is shown (a) as well as the distribution of read length (b).

### *Bull sperm contains abundant piRNAs*

About 26% of reads were annotated by piRNAs, with a size distribution peaking at 30 nt and two-thirds of reads being in the range of 27-32 nt (Fig. 4a). Among these reads, 54% matched perfectly with a published piRNA sequence. The remaining reads were considered to be piRNA-like sequences, bearing mostly 3p and 5p changes (19% and 10% of reads, respectively) or a combination (17%) of 5p, 3p and/or substitutions (Fig. 4b). Modifications of a single nucleotide at either the 5p or 3p end were the most frequent changes (81% and 67% of changes, respectively), with deletions accounting for about 85% of changes at both the 5p and 3p ends (see Additional Table S4). As expected, piRNA reads showed a typical nucleotide

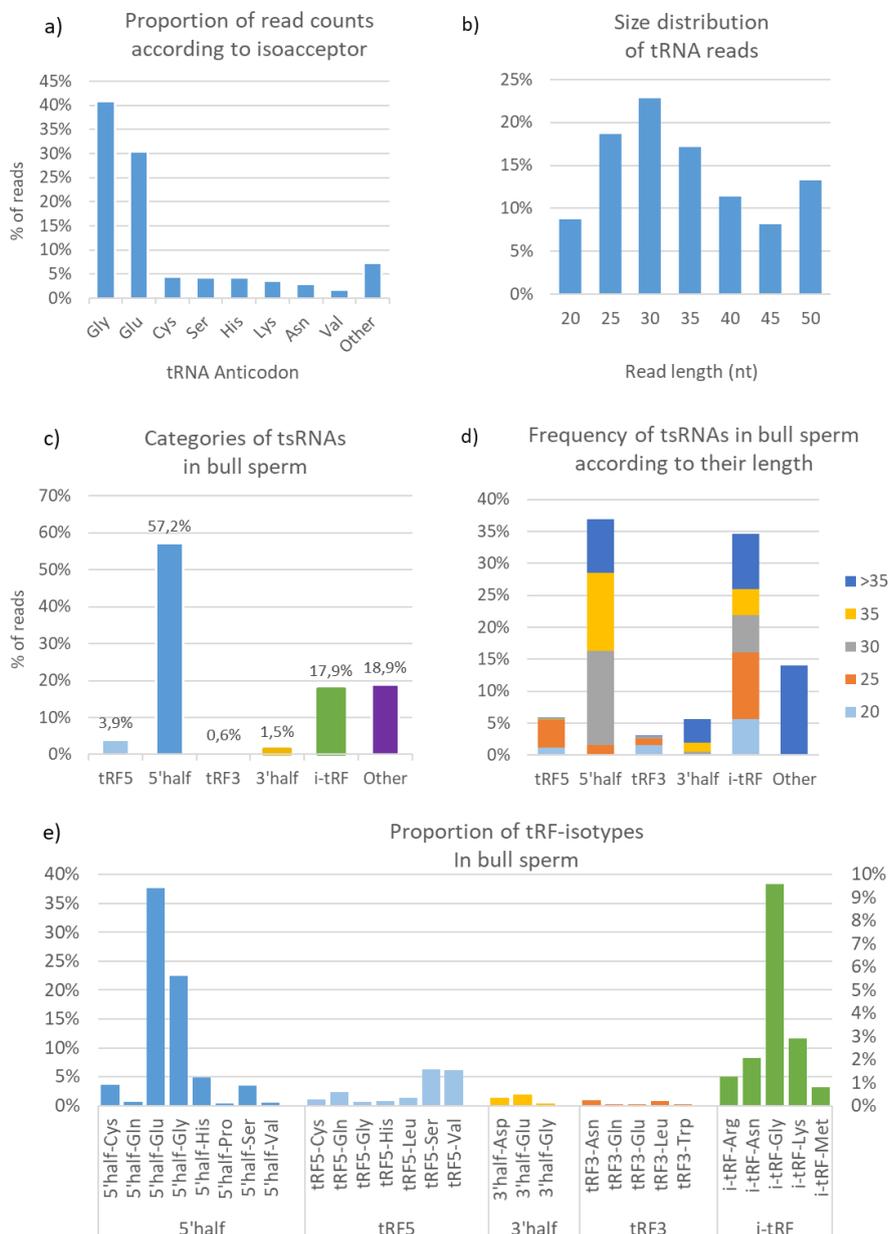
composition (Fig 3c), with strong enrichment for Uridine at the first position. This “1U bias” was also observed for piRNA-like sequences, although to a lesser extent (66% compared to 84% for piRNAs). No adenine enrichment at piRNA nt 10 could be observed for both piRNA and piRNA-like sequences (44). As shown by the distribution of expression (Fig 3d and Additional Table S4), about 50% of piRNAs were expressed at low levels (less than 100 mapped reads), and only 7% were very highly expressed (greater than 10,000 mapped reads). According to the piRBase biogenesis classification, sperm piRNAs were found to primarily derive from intergenic (84%), genic (8%) and LINE-rich (7%) regions (Fig 3e). Interestingly, a highly significant 4-fold enrichment was observed for sperm piRNAs derived from intergenic regions compared to the whole piRBase database (see Additional Table S4). Conversely, all other classes were found to be highly under-represented in sperm. When computing the frequency of each biogenesis class according to the piRNA expression level, no particular enrichment could be observed with the exception of gene-derived piRNAs, which account for a larger proportion of highly expressed reads (Fig 3f).



**Fig 3. Bull sperm piRNA content.** (a) Reads annotated as piRNAs showed a typical size distribution, with a majority of reads being 27-32 nt in length. (b) Among these reads, 54% matched perfectly to published piRNA sequences. The remaining reads were considered to be piRNA-like sequences, bearing mostly 3p and 5p changes (19% and 10% of reads, respectively) or a combination (17%) of 5p, 3p and/or substitutions. (c) Both piRNAs and piRNA-like sequences showed a specific nucleotide composition, with more than 84% and 66%, respectively, showing a preference for uracil at their 5p terminus. (d) The majority of piRNA are expressed at low levels (87% below 500 mapped reads), with only a few (1%) expressed above 100,000 mapped reads. (e) All piRNA biogenesis classes were identified in sperm, but sperm piRNAs were found to primarily derive from intergenic regions. (f) Gene-derived piRNAs showed a specific expression profile, with higher expression levels than the other classes.

*Bull sperm contains abundant tRNA-derived small RNAs (tsRNAs)*

A total of 61991 sequences were annotated as tRNAs, covering the full range of isoforms associated with the 22 amino acids. However, two isoacceptors contributed to 71% of all identified tsRNA: glycine (41% of read counts) and glutamine (30% of reads counts). As illustrated in Fig 4a, six tRNA isoforms each accounted for 2-5%, with the others together accounting for about 7% of read counts. The size of these reads was in the 18-50 nt range, with a bimodal distribution peaking at 30nt and 50nt (Fig 4b).



**Fig 4. Bull sperm tsRNA content.** (a) Strong expression of the Glycine and Glutamine isoacceptors was observed, as well as a high proportion of reads in the 25-35nt range (b). The proportion of each category and the proportion of read length in each category are shown in (c) and (d), respectively. To evaluate the proportion of tsRNAs isotypes, the percentage of read counts was computed for each tsRNA category and each associated anticodon (e). Only the most expressed tsRNAs from each category are reported on the histogram. The left axis (0-40%) refers only to 5p tRHs, which are the most expressed tsRNAs. The right axis (0-10%) refers to the other tsRNAs.

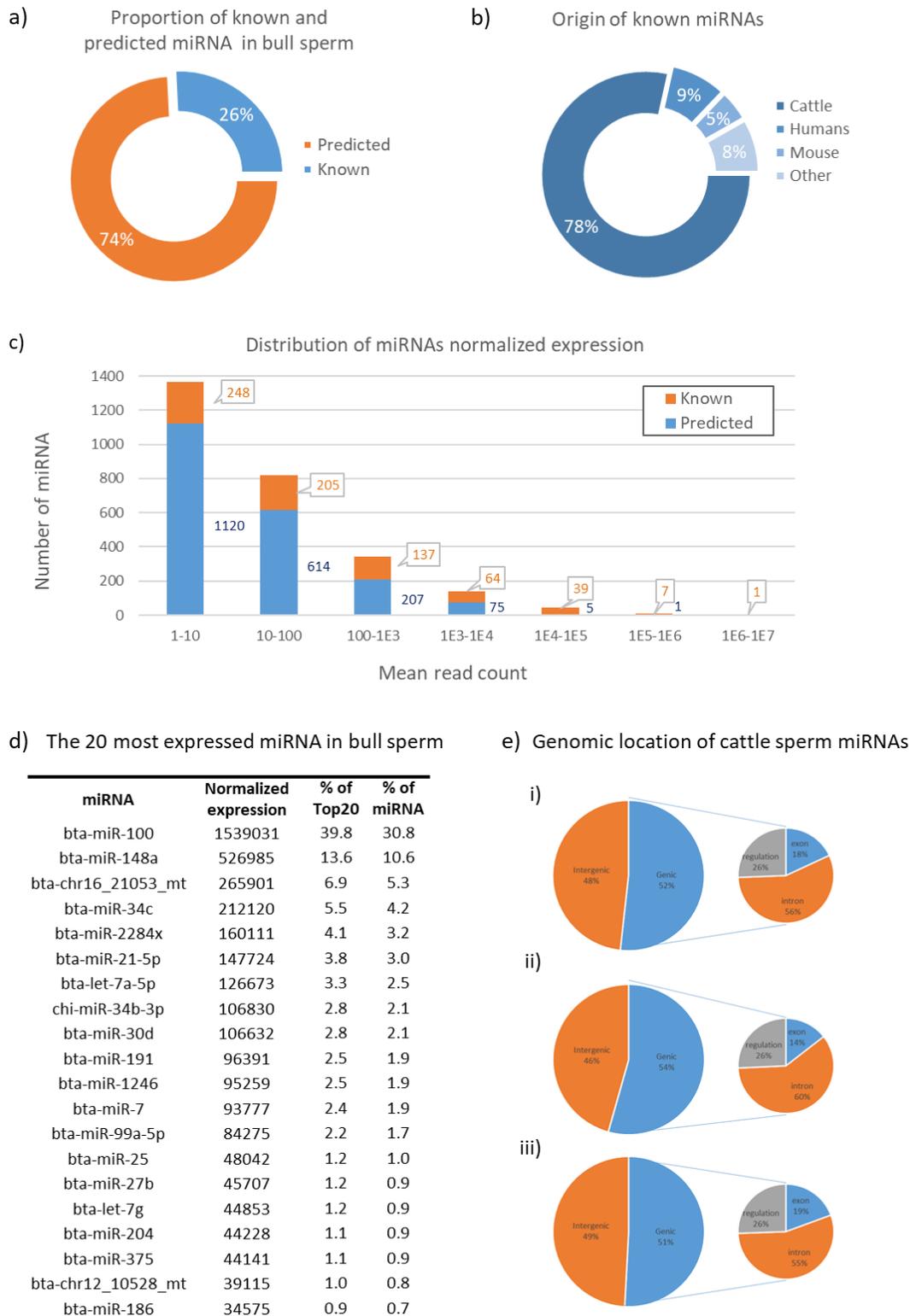
Previous studies have shown that 20-35 nt fragments do not represent random degradation products, but instead are generated by specific mechanisms producing four main categories of tsRNAs, as illustrated in Additional Fig S5. As shown in Fig 4c, 5p-tRHs were the most expressed tRFs, accounting for 57% of read counts, while i-tRFs and tRF5 accounted for 18% and 4%, respectively. Interestingly, nearly 65% of all i-tRFs (84% of expression) were found to be produced by Mitochondria (MT) tRNAs. About 19% of reads were classified as “other” and were shown to be more than 35 nt in length (Fig 4d). These reads were found to be mainly clustered at the 5p side of tRNAs, 75% of them starting at the 1st tRNA nucleotide. In contrast, the i-tRF group was found to be diverse in terms of size, while tRF5 and tRF3 were mainly in the 20-25nt range. When focusing on tsRNAs in the range 20-40nt, the most abundant tRFs were 5p-tRHs, especially Glu, Gly, His and Cys isoacceptors (Figs 5e and S6).

Among i-tRFs, the most abundant were i-tRF-Gly, i-tRF-Lys and i-tRF-Asn, with distinct read length distributions: e.g. reads 20-25 nt and 25-30 nt in length contributed equally (~42%) to i-tRF-Gly, while 25-30 nt reads accounted for 60% of i-tRF-Lys (Additional Fig S6a). The most abundant tRF5s were tRF-Ser and tRF5-Val, the majority of reads being 25nt in length. Most tRF3s are short fragments (20nt), with tRF3-Asn and tRF3-Leu being the most abundant. Interestingly, cleavage sites within the anticodon-Loop were found to be biased towards the 5p-side of the anticodon for both tRHs (Additional Fig S7). The 4th nucleotide appeared to be a preferential cleavage site for 3p-tRHs, while 5p-tRHs are produced by cleavage between the 4th and 7th nucleotides.

*Overall description*

Based on miRDeep2 software, 2723 unique miRNAs were detected, belonging to 2579 miRNA precursors (pre-miRNAs). Among them, 701 (26%) were already known (550 in cattle, 60 in humans, 33 in mouse, and 58 in other species), corresponding to 635 known cattle pre-miRNAs (Fig 5a and 6b). The remaining 2022 predicted miRNAs were associated with 1944 predicted pre-miRNAs (Additional Table S1). A total of 275 pre-miRNAs, including 184 known pre-miRNAs, were covered by both 5p and 3p miRNAs, while 26 predicted miRNAs may be produced by 97 pre-miRNA belonging to miRNA families (ranging from two to seven members). As shown by the distribution of normalized expression (Fig 5c and Additional Table S7), the vast majority (80%) of miRNAs (6% of the known and 85% of the predicted miRNAs) were expressed at low levels (2187 miRNAs expressed with fewer than 100 mapped reads on average) and only 53 miRNAs were highly expressed (more than 10,000 mapped reads). The most highly expressed miRNA, bta-mir-100, exhibited about 1.5M mapped reads.

The twenty most expressed miRNAs (>35,000 mapped reads) accounted for 77% of all miRNA-associated reads (Fig 5d). Bta-mir-100 accounted for 40% of the mapped reads for these top twenty, and for one third of all miRNA reads. The second most expressed miRNA (bta-mir-148) was three times less expressed, accounting for 13.6% of the mapped reads from the top twenty. Interestingly, these top twenty miRNAs included 18 known and 2 putative miRNAs.



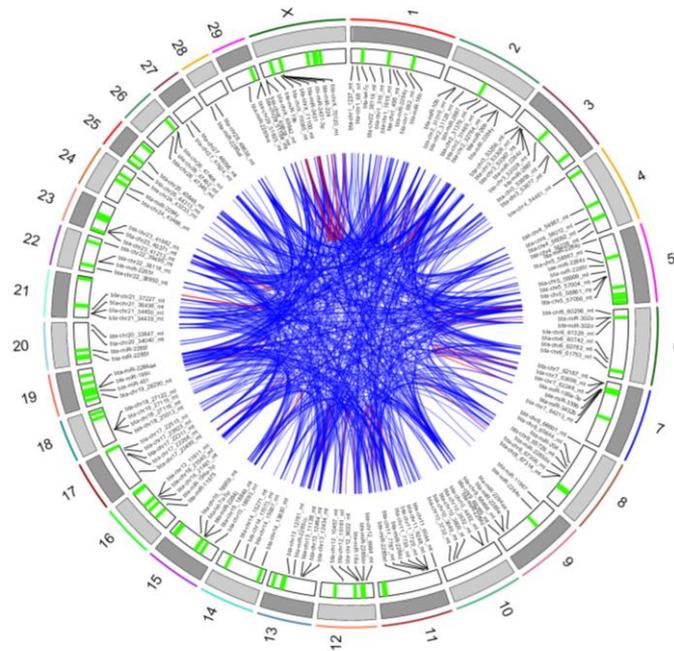
**Fig 5. Bull sperm miRNA content.** (a) Using miRDeep2 software, 2723 miRNAs were identified in bull sperm, including 701 (26%) known and 2022 (74%) predicted miRNAs. (b) Among known miRNAs, 550 were already described in cattle, 60 in human, 33 in mouse and 58 in other species. (c) The vast majority (80%) of miRNAs were expressed at low levels (2187 miRNAs expressed below 100 mapped reads on average), and only 53 miRNAs were highly expressed (above 10,000 mapped reads). (d) Twenty miRNAs account for 75% of the total miRNA mapped reads, including bta-mir-100, which alone

accounts for 30% of miRNA expression (40% of the mapped reads for the top twenty miRNAs). **(e)** miRNA location within several genomic features, including intergenic or genic regions (introns, exons, and other regulatory regions, such as promoter, TSS, TTS, 3p or 5p UTR). Percentages shown on each rightmost “Pie of Pie” chart are expressed relative to the genic miRNAs. (i) Global distribution of sperm miRNAs. (ii) Distribution of known mRNAs; (iii) Distribution of predicted miRNAs.

#### *Genomic distribution and correlation of expression*

The half of miRNAs (52%) are located within genic features (10 kb interval), including within introns (29%), exons (9%), or other regulatory regions such as the promoter, TSS, TTS, 3p or 5p UTR (13%). As illustrated in Fig 5e, known and predicted miRNAs showed nearly the same distribution, with known miRNAs being a little bit more frequently located within genic regions (54 vs 51% for predicted miRNAs). Based on *miRBase* standards (10 kb windows), 216 miRNA genomic clusters could be defined along the bovine genome (Additional Table S8), containing 532 miRNAs (i.e., 20% of total sperm miRNAs).

To highlight genomically clustered miRNAs that could in fact be co-expressed on the same primary transcript, expression correlations were calculated between each pair of the top 1580 miRNAs with mean expression level greater than 10 in at least one breed. While a majority (66%) of the 1.2 million correlation coefficients clustered near zero, significant correlations were observed and are reported in Additional Table S9. Only 45 significant strongly negative correlations ( $r < -0.7$  and adjusted p-value  $< 0.001$ ) involving 59 miRNAs were observed, compared to 3357 significant strongly positive correlations ( $r > 0.75$  and adjusted p-value  $< 0.001$ ) involving 755 miRNAs. As illustrated by the Circos plot (Fig 6), most correlations involved miRNA located on separate chromosomes.



**Fig 6. Circos plot depicting correlations between miRNAs.** Chromosomes are indicated on the first outer track. Genomic clusters are indicated on the second track by green lines. All significant correlations ( $|r| > 0.9$ , adjusted p-value  $< 0.00001$ ) were visualized, altogether involving 283 miRNAs. A subset of these miRNA is shown on the third track. Correlations between miRNA pairs are shown in the center of the plot, using blue and red lines for positive and negative correlations, respectively.

Correlations between genomically clustered miRNAs showed a distinct distribution (Additional Table S10 and Additional Fig S8), suggesting stronger correlations than for non-clustered miRNAs. However, only 77 out of 216 (about one-third) of the genomic clusters showed strong expression correlation between miRNAs. In particular, Cluster 170 had 10 strong correlations between 12 miRNAs (bta-miRNA-381, bta-miRNA-382, bta-miRNA-134, mmu-miRNA-409-3p, bta-miRNA-409a, bta-miRNA-369-3p, bta-miRNA-410, bta-miRNA-3956, hsa-miRNA-382-3p, bta-miRNA-323b-3p, bta-miRNA-656, bta-miRNA-369-5p) among the 19 miRNAs located in this 22 kb cluster. Likewise, clusters 1, 3, 5 and 7 showed strong correlations between at least 4 miRNAs (bta-miRNA-411a, bta-miRNA-323, bta-miRNA-543, bta-miRNA-495, bta-miRNA-411b, hsa-miRNA-329-3p for Cluster 1; bta-miRNA-431, bta-miRNA-127, bta-miRNA-432, mmu-miRNA-127-5p for Cluster 3; bta-miRNA-19a, bta-miRNA-20a, bta-miRNA-19b, bta-miRNA-18a, bta-miRNA-17-5p for Cluster 5; and bta-miRNA-302b, hsa-miRNA-302a-5p, bta-miRNA-302a, bta-mir-302d for Cluster 7).

When considering correlations above |0.5|, 43 clusters of co-expressed miRNA could be defined, ranging from two to 16 co-expressed miRNAs (Table 2).

**Table 2. Genomic clusters of miRNAs having strong expression correlations** (see Additional Tables S8 and S10 for detailed information. Empty cells correspond to none observed cluster in human article).

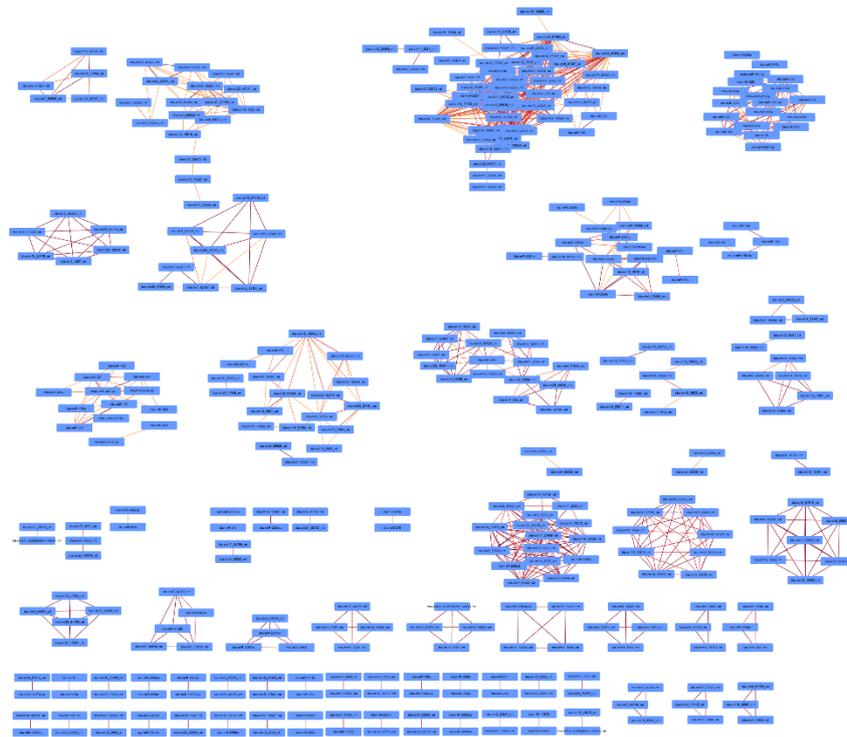
Cluster Number	Mean Correlation	Clustered miRNA putatively co-expressed	Co-expressed in humans (45)
1	0.75	bta-mir-379~bta-mir-411a~bta-mir-299~bta-mir-380~bta-mir-411b~bta-mir-323~bta-mir-329b~bta-mir-543~bta-mir-495~bta-mir-154c	
2	0.72	bta-mir-532~bta-mir-502a-1~bta-mir-502a-2~bta-mir-502b~bta-mir-500~bta-mir-660~bta-mir-502a-1	
3	0.80	bta-mir-431~bta-mir-127~bta-mir-432~bta-mir-136	
4	0.72	bta-mir-450b~bta-mir-450a-1~bta-mir-450a-2	Yes
5	0.91	bta-mir-17~bta-mir-18a~bta-mir-19a~bta-mir-20a~bta-mir-19b	Yes
6	0.60	bta-mir-363~bta-mir-19b-2~bta-mir-20b~bta-mir-18b~bta-mir-106a	Yes
7	0.89	bta-mir-302b~bta-mir-302a~bta-mir-302d	Yes
8	0.64	bta-mir-2443~bta-let-7b	Yes
10	0.65	bta-mir-10174~bta-mir-27b~bta-mir-24-1	
14	0.74	bta-mir-429~bta-mir-200a~bta-mir-200b	Yes
15	0.75	bta-mir-99b~bta-let-7e~bta-mir-125a	Yes
16	0.76	bta-mir-449c~bta-mir-449b~bta-mir-449a	

18	0.85	bta-mir-93~bta-mir-25	Yes
21	0.76	bta-mir-182~bta-mir-96~bta-mir-183	Yes
22	0.68	bta-mir-24-2~bta-mir-27a	Yes
23	0.75	bta-let-7a-1~bta-let-7f-1~bta-let-7d	
24	0.91	bta-mir-3431~bta-mir-224	
25	0.67	bta-mir-105a~bta-mir-767~bta-mir-105b	
29	0.65	bta-mir-16b~bta-mir-15b	Yes
30	0.70	bta-mir-181b-2~bta-mir-181a-2~bta-mir-181a-2~bta-mir-181b-2	Yes
34	0.74	bta-mir-16a~bta-mir-15a	Yes
41	0.89	bta-mir-34b~bta-mir-34c	
47	0.75	bta-mir-29b-2~bta-mir-29c	Yes
49	0.70	bta-mir-181b-1~bta-mir-181a-1	
56	0.80	bta-mir-132~bta-mir-212	
57	0.60	bta-mir-195~bta-mir-497	Yes
61	0.61	bta-mir-2904-1~bta-mir-2887-1	
70	0.64	bta-mir-7857-1~bta-mir-7857-2	
72	0.55	bta-mir-1-2~bta-mir-133a-2	
80	0.68	bta-mir-29a~bta-mir-29b-1	Yes
83	0.73	bta-mir-141~bta-mir-200c	No
91	0.50	bta-mir-374b~bta-mir-421	
93	0.91	bta-mir-222~bta-mir-221	No
111	0.65	bta-chr7_62503~bta-chr7_62509	
118	0.89	bta-chr9_69352~bta-chr9_69353	

125	0.78	bta-chr11_8402~bta-chr11_8403	
140	0.65	bta-chr14_15810~bta-chr14_15827	
145	0.99	bta-chr16_21540~bta-chr16_21541	
152	0.65	bta-chr17_25055~bta-chr17_25086	
169	0.51	bta-mir-6522~bta-chr21_37502	
170	0.72	bta-mir-1185~bta-mir-3956~bta-mir-381~bta-mir-487b~bta-mir-411c~bta-mir-487a~bta-mir-382~bta-mir-134~bta-mir-154a~bta-mir-154b~bta-mir-409a~bta-mir-412~bta-mir-369~bta-mir-410~bta-mir-323b~bta-mir-656	
183	0.78	bta-chr23_40757~bta-chr23_42270	
211	0.78	bta-mir-6526-3~bta-chrX_70042~bta-chrX_70050~bta-chrX_70048~bta-chrX_70052~bta-chrX_70054	

Intriguingly, most clusters gather either known or putative miRNAs, with the exception of Clusters 169 and 211, which comprise both known and predicted miRNAs.

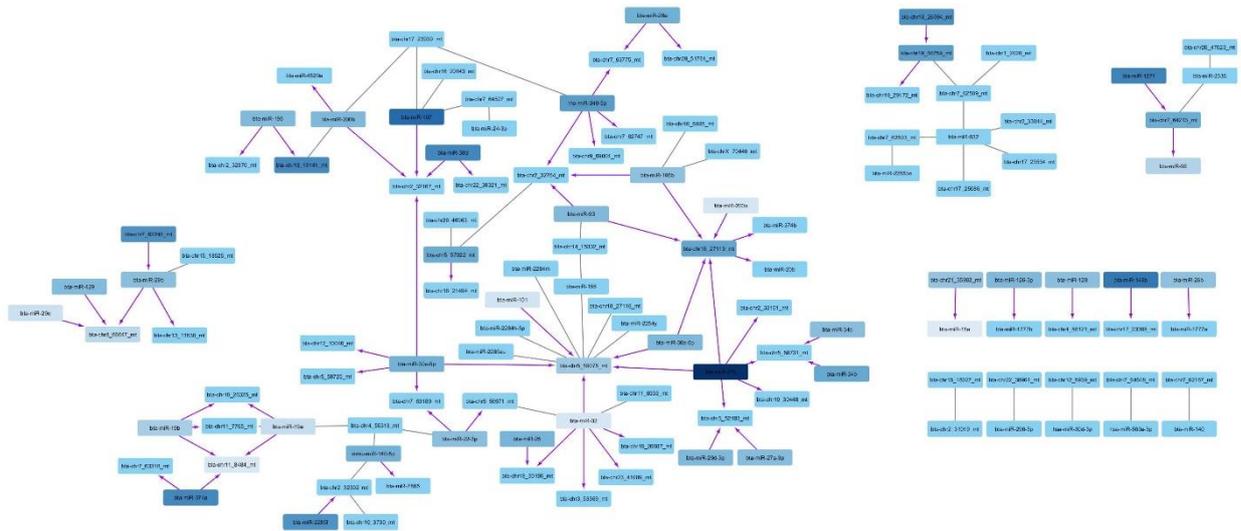
Regarding significant correlations between miRNAs that are not genomically clustered, we identified 77 correlation networks made up of between 2 to 61 positively correlated miRNAs (Fig 7).



**Fig 7. Correlation networks of positively correlated miRNAs.** Correlation networks of miRNA pairs with significant positive correlations above 0.9 as drawn by Cytoscape. Edges are colored according to the correlation level (0.9 in yellow to 1 in red).

Interestingly, five putative functional regulation networks were identified based on their correlation pattern (for instance the positive correlation between bta-miR-24-3p and bta-miR-107, while bta-miR-24-3p and bta-miR-107 are negatively correlated with bta-chr7-64527\_mt suggests that bta-miR-24-3p may regulate bta-chr7-64527\_mt, which may in turn regulate bta-miR-107, see Additional Fig S9).

Likewise, negatively correlated miRNAs defined 14 correlation networks, ranging from 2 to 76 miRNAs (Fig 8). Interestingly, among these, 41 “master” miRNAs were found to be negatively correlated with 46 master miRNAs, thus defining putative regulatory networks with their host genes.



**Fig 8. Correlation networks of negatively correlated miRNAs.** Partially oriented correlation networks of miRNA pairs with significant negative correlations less than -0.7 (any pair) or less than -0.5 (any miRNA paired with a miRNA located within its target genes), as drawn by Cytoscape. Nodes are colored according to the mean correlation level (from -0.5 (light blue) to -0.9 (dark blue)). Purple arrows indicate correlated pairs made up of a given miRNA (arrow start) and another miRNA located within a target gene of the first (arrow end). Grey edges without arrows indicate that linked miRNA are correlated, without any known functional interaction.

### *The repertoire isomiRs in bull sperm*

We also studied the diversity of isomiRs, i.e. changes in the mature miRNA sequence, including substitutions (Polymorphic isomiRs), cleavage variations (5p and 3p isomiRs), and additions at either the 5p or 3p end (see Additional Fig S10 for some examples of isomiRs and their nomenclature). Altogether, 192895 isomiRs were identified among the 2723 sperm miRNA (Additional Table S11). The number of isomiRs per miRNA varied greatly, ranging from 1 (bta-chr18\_27604\_mt) to 1439 (chi-miRNA-34b-3p), with on average  $71 \pm 126$  isomiRs per miRNA. As shown in Fig 9a, the majority (1882, i.e. 69%) of miRNAs were represented by fewer than 50 isomiRs, and only 4 miRNAs (0.15%) exhibited more than 1000 isomiRs. No correlation between the number of isomiRs and the expression level of miRNAs was observed.

In agreement with the levels of miRNA expression, most isomiRs were found to be expressed at low levels (Additional Table S12): only 2329 isomiRs (1.2%) exceeded a mean normalized expression of 100 mapped reads in at least one breed, and about 35785 isomiRs (20%) showed expression levels between 1 and 100 reads. The vast majority of isomiRs accounted for only a

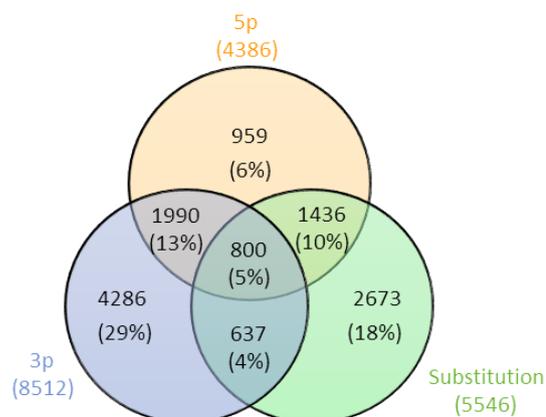
small proportion of miRNA expression: 167776 isomiRs (87%) contributed less than 1% to total miRNA expression, while only 2916 isomiRs (3%) from 2445 miRNAs contributed more than 25% of total miRNA expression. Of note, 101 canonical sequences described in miRBase (14% of known expressed miRNAs) were not detected at all in our data. In addition, the defined canonical sequence was found to be the most highly expressed for only 1269 miRNAs (47%), while the most expressed isomiR differed from either the canonical miRNA published in miRBase or the predicted canonical sequence identified by mirDeep2 in 75% and 46% of cases, respectively. For instance, the published canonical form of bta-miRNA-100 (AACCCGTAGATCCGAACTTGTG) accounted for only 10% of the miRNA read count, while bta-miRNA-100 isomiR AACCCGTAGATCCGAACTTGT (deletion of G in 3p) accounted for 30%. Overall, on average the most expressed sequences were found to account for 51% of total miRNA expression, ranging from 7 to 100%. The twenty most expressed isomiRs accounted for 49% of the total isomiR expression and are related to 10 miRNAs, including bta-miRNA-100 (6 isomiRs in the top twenty), the most abundant miRNA in bull sperm cells. On average, isomiRs accounted for 4.5% of the expression for each miRNA, ranging from 0.06 to 100%.

To avoid bias due to sequencing errors (0.1% of sequencing errors may produce up to 30,000 isomiR artefacts), we focused on 14765 isomiRs accounting for more than 1% of each miRNA's expression and expressed in at least half of the animals of at least one breed (Additional Table S13). Production of multiple isoforms appeared to be a general mechanism, as only 1984 sequences were identified as the canonical miRNAs: 3p, 5p and substitutions within the canonical sequence were detected for 81%, 57% and 77% of miRNAs, respectively. Altogether, 58% and 30% of isomiRs showed changes at their 3p and 5p ends, respectively, while 38% exhibited at least one substitution within the canonical sequence. Single 3p changes and substitutions were the most frequent variations (29% and 18% of isomiRs, respectively). About 5% of isomiRs were found to carry both 5p and 3p changes as well as substitutions (Fig 9b).

a) Distribution of isomiR number

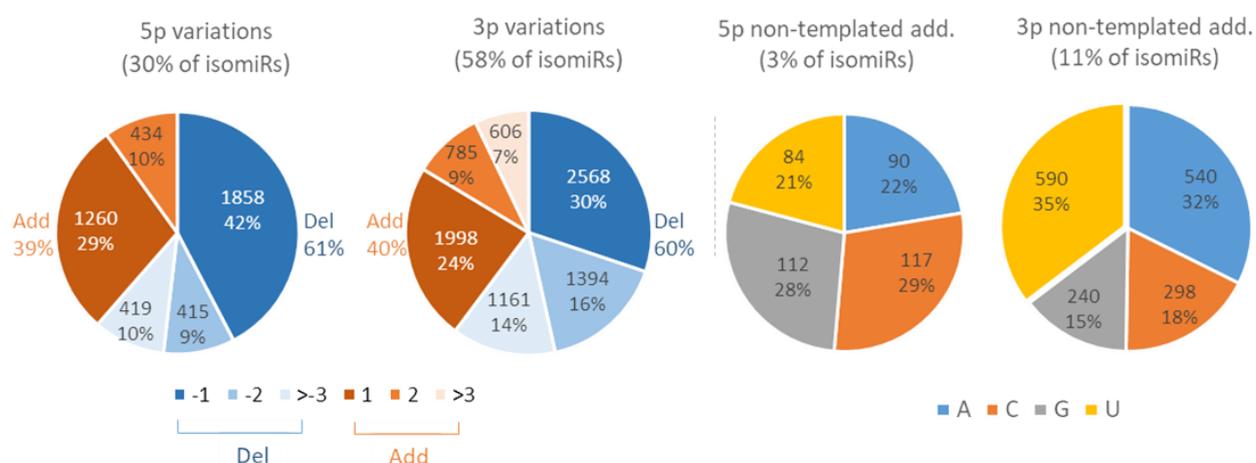
Range of isomiR number	Number of miRNA	%
2-50	1881	69
50-100	408	15
100-200	215	8
200-300	80	2.9
300-400	53	1.9
400-600	44	1.6
600-800	22	0.8
800-1000	15	0.6
> 1000	4	0.15

b) Overall changes carried by sperm isomiRs



**Fig 9. The diversity of isomiRs in bull sperm.** (a) 2722 out of 2723 sperm miRNA were represented by at least 2 isomiRs, with a vast majority (1882, i.e. 69%) represented by fewer than 50 isomiRs. Only 4 miRNAs (0.15%) exhibited more than 1000 isomiRs. (b) The number and percentage of isomiRs carrying 5p or 3p changes, a substitution, or a combination of these changes.

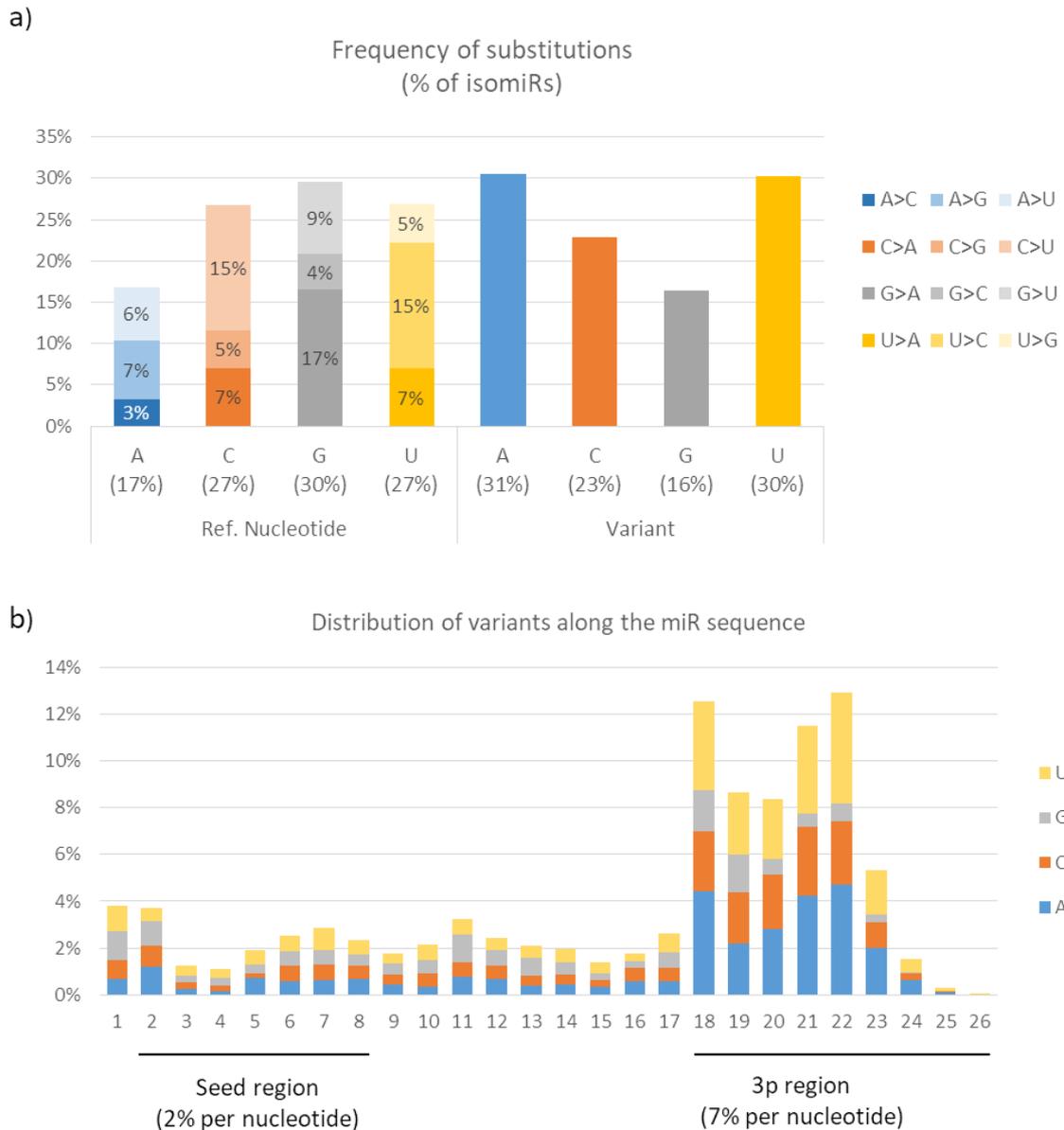
As illustrated in Fig 10, deletions accounted for ~60% of variations affecting 5p and 3p ends, with the addition or deletion of one nucleotide representing altogether 71% and 54% of 5p and 3p changes, respectively. Additions were usually consistent with the reference sequence, but 2071 isomiRs (14%) showed a non-templated nucleotide addition at either 5p (403) or 3p (1668) locations. Among them, only 86 substitutions (4%) may result from known genetic polymorphisms (Ensembl release 94). Addition of Cytidine (C) or Guanosine (G) were observed frequently at 5p (29% and 28%, respectively), whereas addition of Adenosine (A) and Uridine (U) were the most frequent 3p variations (32% and 35%, respectively).



**Fig 10. Frequency of end variations affecting isomiR sequences.** Percentages of addition and deletion at either 5p or 3p ends were computed based on unique sequence counts. Deletions were the most frequent variations, accounting for ~60% of variations affecting 5p and 3p ends. Single nucleotide changes were likewise the most frequent variation (71% and 54% of 5p and 3p changes). Regarding non-templated additions, additions of Cytidine (C) or Guanosine (G) were frequently observed at 5p (29% and 28%, respectively), whereas additions of Adenosine (A) and Uridine (U) were the most frequent 3p variations (32% and 35%, respectively).

Regarding polymorphic isomiRs, about 6% of variations may be related to known genetic polymorphisms, the remaining 94% likely being produced by miRNA editing. As shown in *Fig 11a*, G nucleotides were on average more frequently modified (30% of isomiR counts) than C/U (27%) or A (17%). However, when accounting for nucleotide usage within miRNAs (i.e. frequency of nucleotides within miRNA sequences), substitutions involving C and U appeared to be 50% and 10% more frequent than expected by chance. In contrast, G substitutions appeared 20% less frequently than expected (Additional Fig S11). Nucleotides were mainly replaced by A or U (Fig 11a). Altogether, G>A, U>C and C>U substitutions were the most frequent polymorphisms, each accounting for ~15% of all polymorphisms.

The distribution of variations along the miRNA sequence (see Fig 11b) clearly showed that substitutions are more likely to occur within the 3p region (61% of isomiRs, ~7% substitution per nucleotide) than the first 5p nucleotide (4%) or the seed region (16% of isomiRs, ~2% substitution per nucleotide). Distinct patterns of variation were observed when comparing 5p and 3p regions: substitutions at 5p were prone to be G or U nucleotides (33% and 28%, respectively), whereas substitutions to A or U nucleotides were observed more frequently at 3p (34% and 33%, respectively).



**Fig 11. Frequency of substitutions affecting isomiR sequences. (a)** Percentages of nucleotide substitutions were computed based on unique sequence counts. On average, G nucleotides were more frequently modified (30% of isomiR counts) than C/U (27%) and A (17%), being mainly replaced by A or U. Altogether, G>A, U>C and C>U substitutions were the most frequent polymorphisms, each accounting for ~15% of all polymorphisms. **(b)** The distribution of variations along the miRNA sequence clearly shows that substitutions are more likely to occur within the 3p region (61% of isomiRs, ~7% substitution per nucleotide) than the first 5p nucleotide (4%) or the seed region (16% of isomiRs, ~2% substitution per nucleotide). Substitutions at 5p were prone to be G or U nucleotides (33% and 28%, respectively), whereas substitutions to A or U nucleotides were observed more frequently at 3p (34% and 33%, respectively).

## Diversity of sncRNAs across breeds

Differential expression analysis was used to explore the expression patterns of sncRNAs (miR, piRNA, tRNA and rRNA) across breeds (Additional Table S14). Fig 12 provides, for each sncRNA class, the number of differentially expressed features (adjusted p-value < 0.05) observed for each comparison of breeds.

miRNA/isomiRs	Abondance	Normande	Holstein	Montbeliarde	Belgian Blue	Charolais
Abondance	x	36	45	103	13	15
Normand	129	x	0	76	17	18
Holstein	170	1	x	72	20	11
Montbeliard	175	187	118	x	29	38
Belgian Blue	33	76	54	32	x	8
Charolais	17	65	23	59	4	x

rRNA/tsRNA	Abondance	Normande	Holstein	Montbeliarde	Belgian Blue	Charolais
Abondance	x	0	2409	3529	1419	264
Normand	795	x	5	5849	63	1889
Holstein	716	20	x	137	49	205
Montbeliard	1005	1953	212	x	100	830
Belgian Blue	458	0	21	49	x	36
Charolais	239	189	198	237	2	x

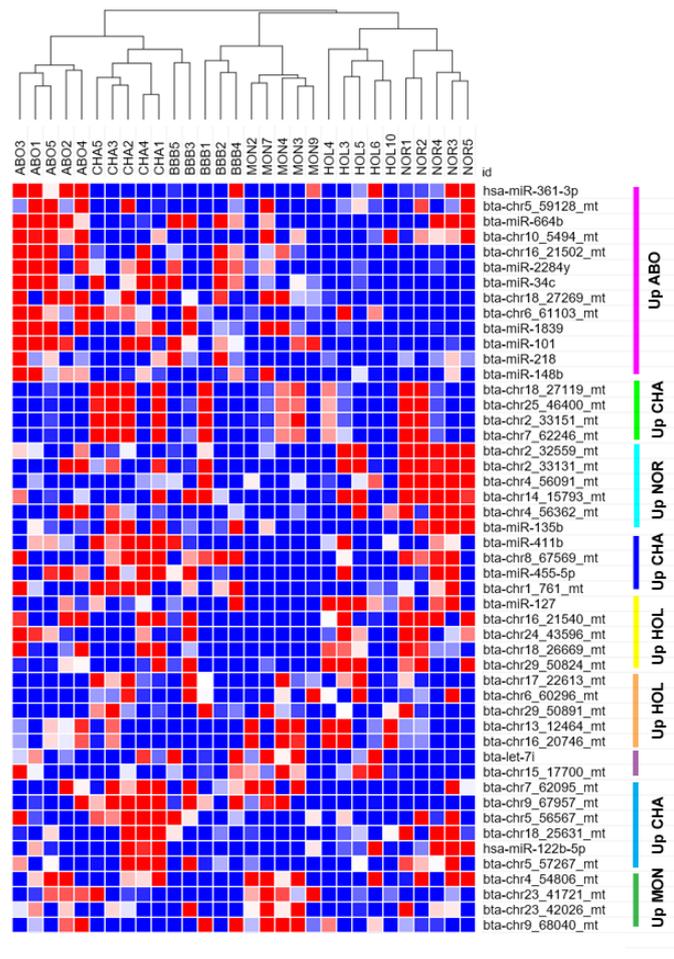
piRNA	Abondance	Normand	Holstein	Montbeliard	Belgian Blue	Charolais
Abondance	x	1502	2815	2044	1218	649
Normand		x	342	1652	817	518
Holstein			x	574	480	619
Montbeliard				x	581	589
Belgian Blue					x	148
Charolais						x

**Fig 12. Differential expression of sncRNAs across breeds.** Number of sncRNA per class showing significant differential expression between pairs of breeds (adjusted p-value < 0.05). Many sequences were found to be differentially expressed between at least two breeds, whatever the sncRNA class. Interestingly, the Abondance and Montbeliarde breeds showed the highest number of differentially expressed sncRNAs, whatever the class and the breed comparison. In contrast, the Belgian Blue vs Charolais as well as the Holstein vs Normand comparisons yielded the lowest number of differentially expressed sncRNAs.

Many sequences were found to be differentially expressed in each pairwise breed comparison, regardless of the sncRNA class. The Montbeliarde breed showed the highest number of differentially expressed sncRNAs, whatever the class and breed comparison. For instance, 63 miRNAs and 2089 rRNAs were found to be differentially expressed on average when considering comparisons between the Montbeliarde and other breeds, whereas only 22 and 720 differentially expressed miRNAs and rRNAs, respectively, were identified on average for other comparisons. In contrast, the Belgian Blue vs Charolais as well as the Holstein vs

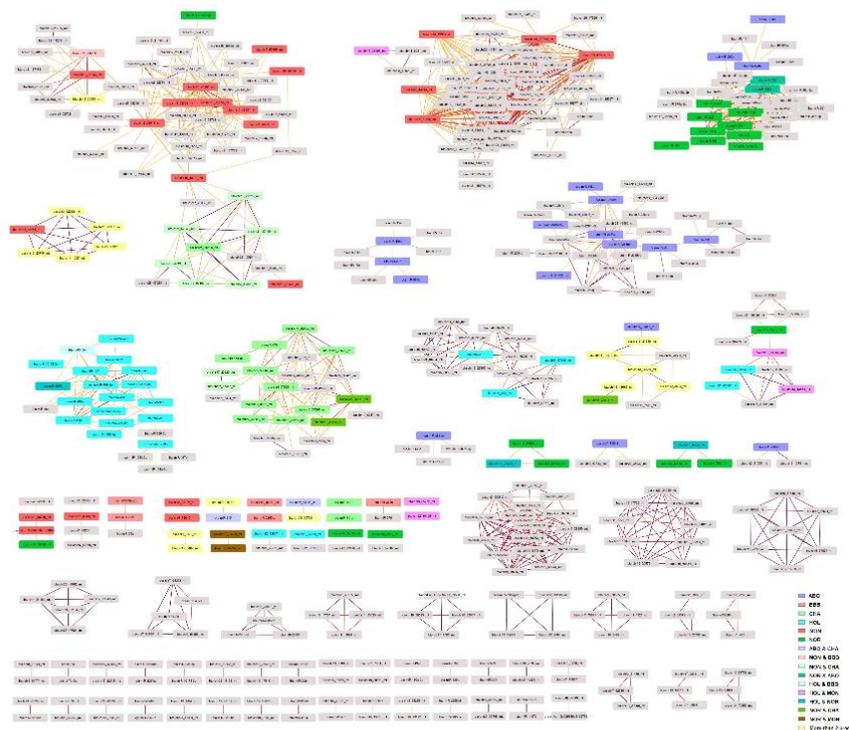
Normand comparisons yielded the smallest number of differentially expressed sncRNAs, whatever the class. For instance, no differentially expressed miRNAs and only one differentially expressed isomiR could be identified between Holstein and Normand. Likewise, only a few differential sequences were identified between Belgian Blue and Charolais, with only 8 miRNAs and 4 isomiRs (compared to 15 miRNAs and 76 isomiRs on average). A subset of five miRNA was chosen to perform a biological validation by qRT-PCR: two miRNA (bta-chr4\_54509 and bta-mir-148b) and two isomiRs (bta-isomir-26a-1\_1 and bta-isomir-26a-1\_2) differentially expressed between Abondance and Normande breeds, as well as one miRNA differentially expressed between Charolais and Normande (bta-chr16\_21053). Five additional bulls, which were not included in the RNA-Seq study, were chosen for each breed. On these five miRNAs, three confirm statistically the RNA-Seq results, one showed the same trend but failed to produce a statistically significant result and no differential expression was observed for the last one (Additional Table S15). Altogether, NGS results were confirmed in 80% of cases (60% with statistical significance).

Consistent with the observed differential expression, breeds could be distinguished by hierarchical clustering on miRNA, with Holstein and Normande being clustered close together, while in contrast the Abondance and Normande breed belong to the most distant clusters (Fig.13). The Belgian Blue breed appeared to be more diverse, with 2 bulls close to the Charolais breed and 3 bulls close to the Monbeliarde breed.



**Fig 13: Hierarchical clustering based on miRNA expression.** Breeds could be distinguished based on 294 differentially expressed miRNA, among which 50 are depicted in the diagram. Breed specific expression patterns could be observed, some miRNA being upregulated in one or a few breeds.

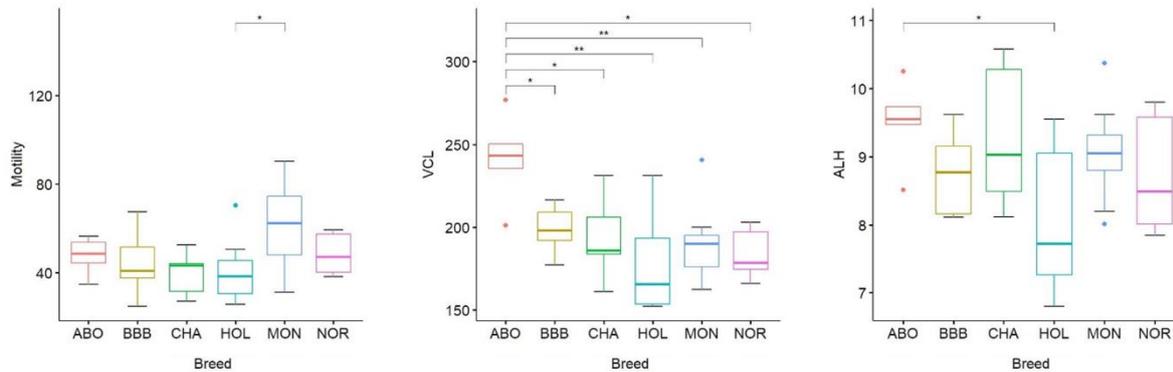
Noteworthy, breed specific expression patterns could be observed, with several groups of miRNAs being differentially expressed between one (or a few) breed and the others. Altogether, about half of differentially expressed miRNA could be associated to such specific patterns. Interestingly, these miRNAs were found to cluster in the previously described correlation networks (Fig 7), with miRNA having a similar pattern clustering within the same correlation networks, as illustrated by the node color code in each network (Fig 14). Thus, positive correlations between miRNA are, at least in part, indicative of coordinated expression in each breed.



**Fig 14. Correlation networks of positively correlated miRNAs.** Correlation networks of miRNA pairs with significant positive correlations above 0.9 (any pair) or above 0.75 (pair including a miRNA exhibiting differential expression across breed), as drawn by Cytoscape. Nodes were colored according to the miRNA-specific expression across breeds (i.e., up-regulation in one or a few breeds compared to the others).

Putative targets of differentially expressed known miRNAs were identified using Targetscan 7.2 and filtered on "weighted context++ score percentile" ( $\geq 95\%$ ), resulting in a set of 2826 targets used for a gene ontology analysis. Several enriched biological terms were identified, including animal organ morphogenesis, cell development, anatomical structure morphogenesis, and regulation of gene expression. Surprisingly, by focusing on the two most highly enriched terms (organ morphogenesis and cell development), 71% of the annotated genes with significantly differential expression among breeds were characterized by highly specific expression in the Montbéliarde breed compared to another. Moreover, the majority of miRNAs targeting these same genes were found to be under-expressed in the Montbéliarde sperm cells (about 90% of the cases).

Noteworthy, semen quality parameters were shown to vary across breeds, especially parameters related to motility such as curvilinear velocity (VCL,  $\mu\text{m/s}$ ), amplitude of lateral head displacement (ALH,  $\mu\text{m}$ ) and progressive motility (%) (Fig. 15 and Additional Table S16).



**Fig. 15: Distribution of motility related sperm quality parameters across breeds.** Boxplots illustrate the distribution of three parameters (progressive motility (%), curvilinear velocity (VCL,  $\mu\text{m/s}$ ) and amplitude of lateral head displacement (ALH,  $\mu\text{m}$ )), showing the minimum, first quartile, median, third quartile, and maximum for each breed. Significant pairwise comparison (Wilcoxon Rank Sum Tests) are depicted above boxplots (\*pvalue under 5%, \*\* pvalue under 5%).

Since a previous study reported differentially expressed miRNA and piRNA between high and low motile spermatozoa (30), association between quality parameters and breed differentially expressed miRNA and piRNA was assessed. Firstly, among the 196 differentially expressed sequences described in Capra et al (30), 156 were also identified in our dataset, corresponding to 121 different miRNA. On these, 22 (18%) were also differentially expressed between breeds (Additional Table S17). Then correlations were computed between quality parameters and miRNA/piRNA expression level in each breed (Additional Table S18). With only five sample per breed, it was not possible to take inter-individual variation into account and evaluate whether variation among breeds is larger than individual variation within each breed. Thus, the mean expression level was computed per breed for each miRNA/piRNA and correlations were estimated between the mean expression level and the mean of each functional parameter. When considering parameters related to motility (VCL, ALH, motility), 833 miRNA were found to be correlated to either VCL, ALH, motility, 129 being also differentially expressed between breeds and 91 being differentially expressed in Abundance breed, which shows significant difference in VCL and ALH parameters compared to the other breeds. In addition, 38 miRNA

belong to miRNA differentially expressed between High and Low motile spermatozoa identified by Capra et al (30). Likewise, 5464 piRNAs were found to be correlated to motility related parameters, 259 being differentially expressed between breeds and 250 being differentially expressed in Abundance breed. These piRNA belong to all clusters associated to High motile (64 clusters) and Low motile spermatozoa (15 clusters) identified by Capra et al (30).

## Discussion

The study of sperm RNAs is challenging, as spermatozoa harbor low amounts of highly fragmented RNA and are to some extent resistant to lysis using conventional methods. As a consequence, numerous procedures have been developed, often leading to highly variable RNA content and making it difficult to compare between samples or studies, especially across species. In our hands, the published protocols led to partial lysis of cattle sperm (evaluated by microscopic examination) and low RNA yield. In contrast, we were able to reproducibly apply our novel protocol to 40 sperm samples. On average, about 60 ng of total RNA can be recovered from about 30 million spermatozoa (2 straws), i.e. about 2 fg on average per frozen spermatozoa. This is consistent with the published estimate of 5 fg/sperm in swine (46), 2-20 fg/sperm in stallions (47), and 5-15 fg/sperm in humans (9, 48, 49). In agreement with previous studies, no 18S and 28S rRNA peaks could be detected by electrophoretic assay, indicative of low contamination by somatic cells (50). Most importantly, the conditions for RNA extraction and sequencing enabled us to establish a comprehensive repertoire of cattle sperm-borne sncRNA, which could serve as reference for later studies focusing on sperm quality characteristics and fertility.

*The overall small RNA content of bull sperm*

*Consistent with SpermBase data on human and mouse (19), the most abundant classes of sncRNA in bull sperm were rRNAs, piRNAs, miRNAs and tsRNAs, with rRNA being the most*

*highly represented class. However, this finding contrasts with Spermbase data, where tsRNA account for more reads than any other sncRNAs. Our findings also agree with previous reports suggesting that piRNAs are highly abundant sncRNAs in sperm [59].*

*Reads annotated as mRNA (0.4% of reads) represented degraded transcript fragments from 3250 genes, of which a substantial portion is involved in spermatogenesis. This information supports the hypothesis that these mRNA fragments are remnants from earlier stages of spermatogenesis rather than derived from contaminating somatic cells in our preparation.*

Though no 18S and 28S rRNA peaks could be detected by electrophoretic assay, about 25% of reads were annotated as rRNAs, mainly 18S (29%) and 28S (23%) rRNAs. However, consistent with a previous report in humans, they appeared to be fragmented, with particular patterns of peaks and read-poor regions, suggesting that this degradation is achieved by a selective RNA cleavage (51). Of note, mature spermatozoa are thought to be depleted in rRNAs, due at least in part to the expulsion of the cytoplasm as a droplet during maturation. The abundance of rRNA fragments in mature sperm suggests that expulsion through the cytoplasmic droplet is not complete and that cleavage may be of importance to avoid protein synthesis. Whether these rRNA fragments may also play a role at fertilization or during zygote development warrants future study.

Together with rRNAs, we identified piRNAs as a major constituent of sperm sncRNAs and observed a great diversity of piRNA-like sequences. In a previous study on cattle gonads, gametes, and embryos, Russel et al reported a distinct population of slightly shorter, 24-27 nt piRNA-like RNAs exhibiting many of the canonical characteristics of piRNAs but having weak or absent 1U bias (52). Consistent with these findings, piRNA-like RNAs identified in our study also appeared to be shorter, resulting mainly from a deletion at either ends. However, we still observed a 1U bias, although weaker than for piRNAs. No particular expression profile and no association with particular biogenesis classes could be observed. Thus, with the exception of their length, these piRNA-like RNAs showed all the same features of piRNAs and likely represent isoforms of piRNAs. Recently, different length variations of piRNAs have been discovered in several studies, and several exonucleases have been found to be required for piRNA end trimming in different species (53-55). To our knowledge, no bioinformatic resource is yet available for the detection and annotation of piRNA isoforms, with the exception of IsopiRBank (56), which unfortunately does not cover the bovine species. However, consistent

with statistics from IsopiRBank and general mechanisms of RNA editing, we found that length variations at the 3' end were the most frequent changes affecting piRNA isoforms.

Besides their established role in repressing transposable elements and protecting the integrity of the genome in germ cells (57), growing evidence suggests that piRNAs also function in regulating protein coding genes in germ cells and early embryo development (58) and play a role in sperm fertility (59). A noteworthy result of our study was the overwhelming majority of reads annotated as intergenic-derived piRNAs, with a significant four-fold enrichment compared to the whole piRBase database. In addition, the sperm piRNAs were predominantly 29-31 nt in length; showed preference for uracil at their 5p terminus; had no adenine enrichment at nucleotide 10; and appeared to derive from transposons in very small number (only 5%). Altogether, this suggests that the piRNA population identified in bull mature sperm is mainly produced by the PIWIL1-directed production pathway rather than the PIWIL2/PIWIL4-directed secondary pathway in which piRNAs are typically produced from transposons or other repeat-associated regions. This is consistent with previous findings in mice, where studies showed that i) in contrast to primordial testis, piRNAs are produced in adult testis independent of the ping-pong mechanism (60); and ii) cauda spermatozoa were enriched in piRNAs apparently derived from the primary PIWIL1-directed production pathway and possibly generated *in situ* in maturing spermatozoa (7). Elucidating the role of bull sperm piRNA and piRNA-like in sperm fertility and early embryo development warrants further investigation.

Regarding tsRNAs, we identified 5p-tRHs as the predominant subgroup in bull sperm, in good agreement with results in human, mouse and rabbit sperm data (19). Altogether, 5p-derived fragments represented 63% of all bull sperm tsRNAs, with tRF3s and 3p-tRHs being expressed at low levels (30x less expressed than 5p derived fragments). A similar pattern was also described in bull sera (61) but with a larger contrast, where 5p derived fragments were 4000x more expressed than 3p ones. This discrepancy may be related to biological differences between tissues, as suggested by the extensive differences in tRF expression observed between human biofluids (62). Alternatively, they may also reflect sperm changes in response to stress induced by the cryopreservation process. Indeed, tRHs are produced by specific cleavage in the anticodon loop of mature tRNA by angiogenin under various cellular stress conditions, and changes in tRF relative expression have been described under pathological

conditions (63). In addition, oxidative stress has been shown to promote tRNA fragmentation (64), and altered 5p tRHs profiles have been observed in mouse sperm after a high-fat diet (16). Whether tRFs expression profiles could be relevant biomarkers for sperm quality or bull management will require further work. Unlike many other studies, we identified i-tRFs as the second contributor (18%) to sperm tsRNAs, with more than 84% of these reads being produced by MT tRNAs, a disproportionately high contribution considering that there are only 22 MT tRNA sequences. Similar results have already been described in prostate cancer tumors, where i-tRFs derived from MT tRNAs are particularly abundant (65). In our data, Gly and Glu isoacceptors were predominant in ejaculated bovine sperm cells, in good agreement with results obtained on cauda epididymis sperm cells in mice (66). Interestingly, these isoacceptors have also been reported as over-represented in seminal plasma exosomes (67) and epididymosomes (66), which can transfer their tsRNA content to the sperm (68). This epididymal cell-sperm communication mechanism has been proposed to link the paternal environment (i.e., nutrient availability) to the mature sperm, providing an adaptive metabolic advantage to the offspring (69). Whether sperm tsRNAs derive from this mechanism or from the intrinsic ability of sperm to cleave tRNA is still unclear. Regarding the biological role of these sperm tsRNAs, several groups have provided compelling evidence in mice that they are involved in paternal epigenetic inheritance and early embryonic development through modification of embryonic gene expression (70, 71). For instance, tRF5- Gly-CCC, -TCC, and -GCC were shown to repress MERVL, an endogenous retro-element active in the preimplantation embryo and regulating the expression of genes involved in totipotency and zygote genome activation (72). Likewise, tRF- Gly-GCC and tRF-Glu-TTC were found to be downregulated in human sperm producing low quality embryo *in vitro* (73). Whether misregulation of some tsRNAs in sperm might contribute to poor sperm quality, abnormal early embryo development, and reduced fertility in cattle warrants future investigation.

#### *The bull sperm miRnome*

Based on miRBase, we identified 635 miRNAs, increasing by ~50% the known miRNA repertoire present in sperm. Indeed, 376 miRNAs have been identified in either human, mouse or rat sperm according to SpermBase (19), from which 216 have also been detected in the

present study. Our results partly overlap those of previous studies on cattle sperm, both in terms of identified miRNAs and expression levels. For instance, 28 out of the 30 most expressed miRNA published by Stowe et al. (74) are also expressed at high levels in the present study, with the exception of bta-miRNA-199a-5p, which is weakly expressed. Likewise, expression levels of 404 shared miRNAs were also in good agreement with those measured by Capra et al (30) (Additional Table S19).

Based on miRDeep2, 2022 additional miRNAs were predicted, associated with 1944 pre-miRNAs. Among these predicted miRNAs, 68 were also previously identified by Capra et al., and 5% were covered by both 5p and 3p miRNAs, increasing the likelihood of being true novel miRNAs. Of note, we report numerous known and predicted miRNA expressed at moderate and low levels. Our efficient extraction procedure, as well as the high sequencing depth in our study (35M reads vs ~4.5M per sample in previous studies), probably accounts for the wider panel of known and novel miRNAs detected in this study. Alternatively, discrepancies between studies may also be explained by different miRBase release versions or the use of different breeds. In particular, about 100 out of the 232 predicted miRNAs published by Capra et al. have been added to the latest miRBase release.

Regarding the genomic location of bovine sperm miRNAs, their distribution appears to be similar to the known distribution of cattle miRNAs published in miRBase and in good agreement with results in other species (75, 76), showing that miRNAs are frequently located within introns of protein-coding genes.

About 20% of sperm miRNAs have been located within genomic clusters (within 10 kb of another miRNA), in good agreement with estimates based on miRBase for cattle (22%) and humans (20%). We also explored whether genomically clustered miRNA are co-expressed to highlight physically linked miRNA, such as those produced from a same pri-miRNA transcript. Our analysis, based on statistical correlations, suggests that some, but not all (i.e. 20%), genomically clustered miRNAs may be co-expressed as a single polycistronic pri-miRNA transcript. Among the 43 clusters of co-expressed miRNA identified in this study, 15 were already described in humans (45). Of note, miRNA belonging to the 9 remaining human clusters were not expressed in sperm. Our study thus expands the list of known polycistronic pri-miRNA clusters, adding in particular 5 large clusters ranging from 4 to 16 miRNAs. Interestingly, 8 clusters comprise predicted miRNAs alone, and two clusters comprise both

known and predicted miRNA, strengthening their likelihood to be genuine miRNAs. Surprisingly, many strong positive and negative correlations were also observed between non clustered miRNA pairs. Such correlations may correspond to regulatory networks, including for instance miRNAs co-regulated by shared regulatory features (positive correlation), as well as miRNAs hosted by transcripts targeted by other miRNAs (negative correlation). Regarding positive correlations, our analysis highlighted 77 correlation networks that mostly gather miRNAs differentially expressed between breeds, suggesting that they reflect mainly inter-breed genetic differences. Regarding negative correlation networks and assuming that intragenic miRNAs are transcribed in parallel with their host transcripts, we also identified putative miRNA-miRNA regulatory networks, encompassing 41 “master” miRNAs negatively correlated with other miRNAs hosted within their predicted targets. In support of this hypothesis, co-expression of miRNAs and their host genes has already been shown by expression studies (77, 78), as well as eQTL studies that show a substantial fraction of eQTLs shared between intragenic miRNAs and their host genes (79). This mechanism is, however, far from general; many intronic miRNAs possess their own promoter and are transcribed independently of their hosting genes (80). Additional data will be required to further explore in cattle whether miRNAs and their host genes are co-expressed, and in vitro transfection studies may be helpful in validating some of the putative miRNA-miRNA regulatory networks identified in this study.

A large number and an impressive diversity of isomiRs were also observed, with the number of isomiRs per miRNA varying greatly, ranging from a few up to one thousand. Some miRNA, such as miRNA-100, miRNA-146a, miRNA-151-3p, miRNA-143, miRNA-331-3p, miRNA-23b, miRNA-24, miRNA-222 and miRNA-199a-3p, have been proposed to be more prone to variations in human tissues (35). Interestingly, these miRNAs also display a high number of isomiRs in our data (531 on average, compared to 97 for the whole miRNome). This suggests that common sequence features or conserved molecular mechanisms contribute to the isomiR profile across species. Surprisingly, published canonical sequences were found to be the most expressed in only half of the cases, and 14% of them were not detected at all in our data. Evolution of the criteria used to define canonical miRNAs throughout time (81), as well as tissue- and species-specific expression, may account for this situation.

### *IsomiRs: an underexplored diversity*

Focusing on 14765 relevant isomiRs in terms of expression levels and representation across breeds, 3p variations appeared twice as frequently as 5p variations, both in terms of isomiR counts and expression. More than half of isomiRs are produced by a single 3p change (34%) or substitution within the miRNA sequence (21%), while single 5p changes account for only 8% and 38% result from a combination of several changes. These findings are consistent with previous reports, showing for instance that a vast majority of observed miRNA carried changes at their ends, in agreement with the hypothesis of cleavage variations by either Drosha or Dicer (43). However, the high frequency of 3p changes, especially deletions, suggests that other mechanisms also contribute to the variation pattern. In particular, crystallographic studies have shown that 5p ends are tethered to Argonaute through multiple interactions and packed within the MID domain, whereas 3p ends may extend from the PAZ domain and may therefore be susceptible to 3'–5' exonucleolytic activities (82). We also described non-templated nucleotide additions, such as mono uridylation and adenylation, especially at 3p ends. This finding is consistent with other studies, showing that post-transcriptional modification of RNA and miRNA by 5p–3p nucleotidyl transferases, especially poly(A) polymerase and terminal uridylyl transferases, is a common mechanism, conserved in a variety of species (37, 83). In cattle sperm, although mono uridylation and adenylation affect a similar proportion of miRNA (32–35%), mono uridylation seems to be the predominant mechanism, accounting for 34% of expression of isomiRs carrying 3p non-templated additions, compared to only 25% for mono adenylation.

We also identified substitutions as a major driver of miRNA variation, accounting for more than one third of isomiRs. The vast majority of these polymorphic isomiRs are likely produced by miRNA editing, since only 6% of these variations co-localize with known SNPs. However, this percentage is probably underestimated due to limited availability of whole genome sequences from the Abondance, Charolais, Normand, and Belgian Blue breeds. Whole genome sequencing of bulls included in this study may help in the future to distinguish more precisely genetic polymorphism from variations introduced by editing. Several mechanisms have been described as major editing processes, including adenine and cytosine deamination (leading to A>G or C>U transitions), nucleotide insertion in the mature miRNA by nucleotidyl transferases, and RNA-specific adenosine deaminase (ADAR) A>I deamination (detected as

A>G transitions). Consistent with these data, our results showed an elevated frequency of C>U changes, the C substitution being 50% more frequent than expected based on nucleotide usage within miRNA sequences. A>G substitutions were also observed at lower frequency, suggesting that cytosine deamination is more frequent in bull sperm than adenine deamination. In our data, G>A and U>C transitions were observed at similar frequencies as A>G and C>U transitions. This suggests that in addition to classic A>I and C>U editing events, some other molecular mechanisms also significantly shape the sperm miRnome, making the editing paradigm much more complicated than initially thought. In line with our results, G>A and U>C changes have been reported in B lymphocytes (84), and G>A editing of WT1 transcripts has been associated with the APOBEC3A protein, a member of the APOBEC/ADAR protein family (85). Interestingly, APOBEC3A has been shown to be expressed at high levels in mouse sperm heads (19) as well as human sperm, with upregulation between sperm achieving pregnancy and those that do not in intrauterine insemination (86). Whether the relative frequency of classic vs non-classic editing in sperm may also be associated with APOBEC3A and impact bull fertility in cattle warrants future investigation. Furthermore, ADAR A>I editing has been previously described as the major driver of seed polymorphisms in the mouse brain, accounting for about 40% of changes within the seed (87). This finding seems to be far from general, and in bull sperm G>A and C>U transitions are the most frequent changes within the seed. Polymorphisms within the seed and 5p variations together account for only 24% of isomiR expression. However, since polymorphisms as well as the seed region shift due to 5p variation may change the set of target mRNAs (39), these changes may be of functional importance, especially for isomiRs accounting for a large proportion of miRNA expression (e.g. 74 isomiRs polymorphic within the seed as well as 94 5p isomiRs having a relative expression above 50%). Overall, our findings on bull sperm support the notion that the diversity of expressed isomiRs should be considered instead of focusing on canonical miRNAs (88).

#### *Diversity of sncRNA profiles among breeds*

Many sequences belonging to all sncRNA classes showed breed specific expression patterns and were found to be differentially expressed between breeds. For instance, the Abondance and Monbeliarde breeds showed the highest number of differentially expressed sncRNAs, whatever the sncRNA class and the breed comparison, while only a few differential sequences could be identified between Holstein and Normande. The origin of these differences remains

elusive and several hypothesis can be drawn. Indeed, breeds in this study differ in many aspects including genetic background (breed effect), local environment, nutrition and management, many of which may influence gene expression, making it difficult to decipher the breed and environment effect.

Consisted with a genetic origin of miRNA and piRNA differential expression between breeds, we observed changes in miRNA and piRNA expression according to sperm production parameters, which are under genetic control and are known to differ between breeds. In particular, Abondance bulls showed significant differences for motility related parameters and a large number of miRNA (833) and piRNA (5464) were found to be associated with these parameters, including miRNA (91) and piRNA (250) differentially expressed between Abondance and other breeds. Interestingly, a previous study already reported differentially expressed miRNA and piRNA between high and low motile spermatozoa (30). About 25% of these miRNAs and 13% of piRNA located in these piRNA clusters were shown to be also associated with motility parameters in this study. This apparently low overlap between studies may result either from low statistical power or from physiological differences. Indeed Capra et al compared two populations of high/normal and extremely low motile spermatozoa (i.e VCL 100 vs 50  $\mu\text{m/s}$ , ALH 3 vs 2.15 $\mu\text{m}$ , Motility 48 vs 3.8%), whereas our study analyzed semen in a normal range of variation (i.e. VCL 280-150 $\mu\text{m/s}$ , ALH 10-6 $\mu\text{m}$ , Motility 90-25%).

Alternatively, environmental conditions may also influence sncRNA expression. In this respect, it is worth to note that Holstein and Normande bulls, which were housed in the same semen production center, also showed the lowest number of differentially expressed sncRNAs. Likewise, Charolais and Abondance bulls were housed in the same center and cluster close together based on miRNA expression despite differences in terms of semen functional parameters.

Disentangling genetic and environmental effect will require a large number of bulls from several breeds to be raised altogether in the same semen production center to evaluate the breed effect, and to duplicate this design in several semen production center to measure the environmental effect.

## Methods

### *Semen collection*

Commercial sperm straws were obtained for ten Holstein and Montbéliarde bulls, as well as five bulls for each of the following breeds: Normande, Charolais, Abondance and Belgian Blue. Thus a total of 40 representative bulls (normal sperm morphology and motility, good fertility) was selected for this study. Semen collection was performed by French and Belgian semen production centers (Evolution-XY, Umotest, Auriva-Elevage and AWE) following the routine protocol. *Freshly ejaculated* semen was diluted in egg yolk extender (Optidyl, IMV Technologies) before freezing using the DigitCool program (IMV Technologies).

### *Total RNA isolation*

Two straws per ejaculates (30-40 million sperm cells) were thawed and pooled for RNA extraction. Sperm quality was assessed after thawing using IVOS (Hamilton Thorne Inc) and flow cytometry (EasyCyte 6HT, Guava Technologies Inc). Motility, kinetics, viability, mitochondrial potential activity and oxidation sensitivity were measured. A search for somatic cells was also performed using both visual examination under microscope (counting about 1000 sperm cells/sample) and flow cytometry (forward- and side- scatters), showing that contamination was below detectable levels. Remaining seminal plasma and extender were removed by washing the semen twice in 1ml of Phosphate Buffered Saline solution (PBS), followed by centrifugation 5 minutes at 2400 x g.

RNA extraction was performed according to an optimized Guanidium-Trizol total RNA extraction protocol, based on a Guanidium thiocyanate–phenol–chloroform extraction protocol (89). Sperm pellets were homogenized in 100µl of RLT buffer (Qiagen) supplemented with 1 µl beta-mercaptoethanol and incubated 15 minutes at room temperature. Then, 1 ml of Trizol (Invitrogen) was added and samples were incubated 5 minutes at room temperature. Complete lysis was ascertained by microscopic examination of 10µl aliquots. After addition of 100µl of chloroform and vigorous shaking for 15 seconds, samples were incubated 2 minutes at room temperature and centrifuged at 12 000 x g during 15 minutes at 4°C. The aqueous phase was transferred to a new collection tube and mixed with 100µl of chloroform by

vigorous shaking. After incubation for 2 minutes at room temperature and centrifugation at 12 000 x g during 15 minutes at 4°C, the aqueous phase was recovered and nucleic acids were precipitated overnight at -20°C using 1 volume of isopropanol supplemented with 25µl glycogen (Ambion AM9510, 5mg/ml). After centrifugation at 12 000g during 15 minutes at 4°C, pellets were washed with 75% ethanol and dried under vacuum. Dried pellets were then re-suspended in 12µl of RNase free water and incubated 1 hour at 4°C before quality control assessment.

#### *RNA quality controls and sequencing*

RNA concentration was assayed using the Qubit® Fluorometer (Molecular Probes by Life Technologies) using the Qubit® RNA HS Assay Kits (Q32852 Molecular Probes), according to the manufacturer's recommendations. During protocol development, quality and concentration were also evaluated by electrophoretic assays on an Agilent Bioanalyser 2100, using standard procedures.

In addition, RT-qPCR was used to quantify the expression of miRNA-125b-5p, which has been shown to be highly expressed in sperm (41). Total RNAs (5ng) were reverse transcribed in 10-µl reactions using the miRCURY LNA™ Universal RT microRNA PCR Starter kit (Exiqon). RNA spike-in control (Usp6) was also added during the RT-step. Triplicates were assayed in 10-µl qPCR reactions following the miRCURY LNA kit protocol, using a StepOnePlus Real time PCR System (Applied biosystems). Amplification curves were analyzed using the StepOne software v2.3 both for determination of Ct values and for melting curve analysis.

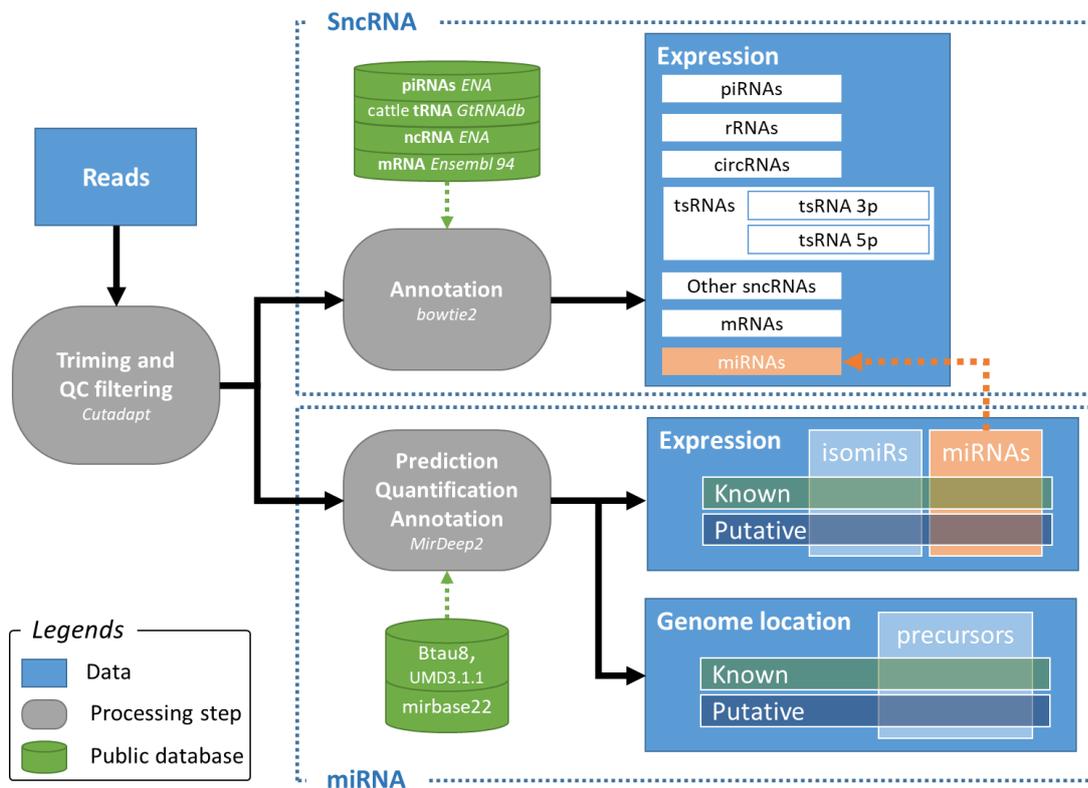
Library preparation and sequencing was done by the Exiqon service provider. Briefly, miRNA libraries were constructed using the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® kit, including a size selection and quality control by Bioanalyzer. Deep sequencing was performed on Illumina HiSeq2000, targeting 40M single-reads (50 bp) per sample.

#### *Sequence analysis*

Low quality sequences were removed and adaptors were trimmed (Cutadapt). Sequences shorter than 17nt were filtered and remaining sequences were then grouped by unique

sequences to obtain total read counts for each. A workflow mostly based on miRDeep2 (90, 91) was used to identify and quantify known as well as predicted miRNAs. The unique sequences were mapped to the reference bull genome (Btau8, UMD3.1.1). Precursors and miRNAs were identified using the miRDeep2 core module, miRDeep2.pl. Potential miRNA datasets were created by adding known miRNA in bull (miRBase v22, March 2018) to miRNA associated with predicted precursors with a miRDeep2 score  $\geq 0$ . The same operation was performed to create a data set of potential precursors. The quantifier.pl miRDeep2 module was then used to map unique reads, the set of potential miRNAs and all known miRNAs (miRBase v22) on the set of potential precursors enabling the annotation of miRNAs. The quantification results produced by the quantifier.pl module were then filtered with a custom perl script “parse\_miRDeep2\_outputs.pl” (<https://github.com/SmartBioInf/PAQmiR>) to eliminate any redundancy between known and predicted miRNAs.

Two separate expression files were generated from these results, providing either miRNA-level read counts (corresponding to similar sequences mapped to the same location), or isomiR-level read counts (corresponding to unique sequences). In parallel with this specific study of miRNAs, all unique sequences have been categorized by mapping to several databases and published data: cattle mRNA (Ensembl release 94, [www.ensembl.org](http://www.ensembl.org)), RFAM release 14 (<http://rfam.wustl.edu>), cattle ncRNA from ENA (<https://www.ebi.ac.uk/ena>), cattle tRNA from GtRNadb ([www.gtrnadb.ucsc.edu](http://www.gtrnadb.ucsc.edu)), human, mouse and cattle piRNAs from ENA, as well as (30) and (92). Fig 16 summarizes this analysis workflow.



**Fig 16. NGS data analysis workflow at a glance.**

An in-house pipeline (93) was used to annotate miRNA precursors relative to gene features (Ensembl release 94) and repeats (UCSC RepeatMasker file). The following criteria were applied for the delineation of promoters, transcription start sites (TSS) and transcription termination sites (TTS): promoter, -2000 to +100 bp relative to the TSS; TSS, -100 to +100 bp relative to the TSS; TTS: -100 to +100 bp relative to the TTS. The miRNAs located inside TTS, TSS, promoter, exon, intron, 3 or 5p UTR have been located in “genic” regions. On the other end, all miRNAs situated only on his own gene or in another place in the genome have been considered in “intergenic” regions.

Targetscan 7.2 (release March 2018) and Targetscan custom 5.2 were used to predict targets for known and predicted miRNAs, respectively. The gene ontology and KEGG pathways were explored using WebGestalt (94) (<http://www.webgestalt.org/>).

Correspondence between piRNA clusters published by Capra *et al* and our piRNA sequences was ensured by mapping piRNA on cluster sequences using blast (options -task “blastn-short”), keeping only perfect matches along the whole piRNA sequence.

RTqPCR validation of breed differential expression.

Two miRNA (bta-chr4\_54509 and bta-mir-148b) and two isomiRs (bta-isomir-26a-1\_1 and bta-isomir-26a-1\_2) differentially expressed between Abondance and Normande breeds were chosen, as well as one miRNA differentially expressed between Charolais and Normande (bta-chr16\_21053). LNA primers were purchased from Qiagen (sequences and references given in Additional Table S15). Additional bulls, which were not included in the RNA-Seq study, were chosen to perform the biological validation. Sperm RNA was produced for five bulls from each breed following the protocol described above. Total RNAs (5ng) were reverse transcribed in 10- $\mu$ l reactions using the miRCURY LNA™ Universal RT microRNA PCR Starter kit (Exiqon). Triplicates were assayed in 10- $\mu$ l qPCR reactions following the miRCURY LNA kit protocol, using a StepOnePlus Real time PCR System (Applied biosystems). Amplification curves were analyzed using the StepOne software v2.3 to compute Ct values and analyse the melting curve. Relative expression was computed with the qbase<sup>+</sup> software (Biogazelle).

### *Statistical analysis*

Expression similarity between miRNA pairs was assessed with the Pearson correlation matrix, using the *rcorr* and *p.adjust* functions from the Hmisc R package. Only miRNAs with mean normalized expression greater than 10 in at least one breed were included in the correlation study. The circos plot was drawn using the RCircos R package. Cytoscape 3.7.1 (95) was used to build correlation networks, using the Prefuse Force Directed layout. Between-class correspondence analysis was performed using the ade4 R package (with the *dudi.bca* function). DESeq2 (96) was used to normalize read counts and test for differential sncRNA expression between breeds, based on negative binomial generalized linear models. Statistical significance was evaluated based on Benjamini-Hochberg adjusted p-values (97), with a significance threshold of 5%. Multiple comparison tests between semen functional

parameters were performed using the `wilcox_test` function from the R `rstatix` package and boxplots were drawn using the `ggboxplot` function from the `ggpubr` package. Correlations between semen functional parameters and miRNA and piRNA expression was performed using the `rcorr` and `p.adjust` functions from the `Hmisc` R package. Correlations above  $|0.7|$  ( $pvalue \leq 0.1$ ) were kept for miRNA, while correlations above  $|0.8|$  ( $pvalue \leq 0.001$ ) were kept for piRNA.

### *Data availability*

The (trimmed) raw data have been stored at ENA, European Nucleotide Archive, under the primary accession number: PRJEB33940.

## Conclusion

Our study provides a comprehensive overview of cattle sperm sncRNAs, and our findings will pave the way for future work on the functional significance of sncRNAs in terms of semen quality and fertility. In particular, the role of rRNAs, tRFs and isomiRs at fertilization and during zygote development warrants future study. In particular, their individual expression levels, global expression profile, or frequency of particular editing events may prove to be relevant biomarkers for sperm quality or bull management, and thus of great interest for artificial insemination companies.

## Declarations

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

### **Availability of data and materials**

The (trimmed) raw data have been stored at ENA, European Nucleotide Archive, under the primary accession number: PRJEB33940.

### **Competing interests**

The authors have declared that no competing interests exist.

### **Funding**

This study was funded by a grant of “Agence Nationale de la Recherche” (ANR, ANR-13-LAB3-0008-0) and APIS-GENE (Sequamol).

### **Authors' contributions**

Eli Sellem was the main investigator, taking part to Conceptualization, Data curation, Formal analysis, Visualization, Writing – original draft and Writing – review & editing

Sylvain Marthey was involved in Data curation, Formal analysis – bioinformatics and Writing – review & editing

Andrea Rau and Luc Jouneau were involved in Data curation, Formal analysis – biostatistics and Writing – review & editing

Aurelie Bonnet, Chrystelle Le Danvic and Jean-Philippe Perrier took part to Investigations and Writing – review & editing

Sébastien Fritz and Mekki Boussaha provided Ressources and were involved in Formal analysis – Genetic polymorphisms vs editing

Hélène Kiefer was involved in Methodology, Investigations, Visualization, Writing – review & editing.

Hélène Jammes was involved in Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Laurent Schibler was involved in Conceptualization, Methodology, Funding acquisition, Supervision, Project administration, Formal analysis, Validation, Writing – original draft, Writing – review & editing.

### **Acknowledgements**

We gratefully acknowledge the French and Belgian semen production centers, Evolution XY, Umotest, Auriva-Elevage and AWE for providing the semen of all the bulls used in this study. We would also like to thank Didier Boichard (INRA, Jouy-en-Josas, France) for the fruitful discussions and advice around the choice of bulls.

## References

1. Grunewald S, Paasch U, Glander HJ, Andereg U. Mature human spermatozoa do not transcribe novel RNA. *Andrologia*. 2005;37(2-3):69-71.
2. Dadoune JP, Siffroi JP, Alfonsi MF. Transcription in haploid male germ cells. *Int Rev Cytol*. 2004;237.
3. Hilz S, Modzelewski AJ, Cohen PE, Grimson A. The roles of microRNAs and siRNAs in mammalian spermatogenesis. *Development*. 2016;143(17):3061-73.
4. Chuma S, Nakano T. piRNA and spermatogenesis in mice. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 2013;368(1609):20110338.
5. Das PJ, McCarthy F, Vishnoi M, Paria N, Gresham C, Li G, et al. Stallion sperm transcriptome comprises functionally coherent coding and regulatory RNAs as revealed by microarray analysis and RNA-seq. *PLoS one*. 2013;8(2):e56535.
6. Nixon B, Stanger SJ, Mihalas BP, Reilly JN, Anderson AL, Tyagi S, et al. The MicroRNA Signature of Mouse Spermatozoa Is Substantially Modified During Epididymal Maturation. *Biology of reproduction*. 2015;93(4):Article 91, 1-20.
7. Hutcheon K, McLaughlin EA, Stanger SJ, Bernstein IR, Dun MD, Eamens AL, et al. Analysis of the small non-protein-coding RNA profile of mouse spermatozoa reveals specific enrichment of piRNAs within mature spermatozoa. *RNA Biol*. 2017;14(12):1776-90.
8. Reilly JN, McLaughlin EA, Stanger SJ, Anderson AL, Hutcheon K, Church K, et al. Characterisation of mouse epididymosomes reveals a complex profile of microRNAs and a potential mechanism for modification of the sperm epigenome. 2016;6:31794.
9. Boerke A, Dieleman SJ, Gadella BM. A possible role for sperm RNA in early embryo development. *Theriogenology*. 2007;68:147-55.
10. Hosken DJ, Hodgson DJ. Why do sperm carry RNA? Relatedness, conflict, and control. *Trends Ecol Evol*. 2014;29(8):451-5.
11. Liu WM, Pang RT, Chiu PC, Wong BP, Lao K, Lee KF, et al. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(2):490-4.
12. Yuan S, Schuster A, Tang C, Yu T, Ortogero N, Bao J, et al. Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. *Development*. 2016;143(4):635-47.
13. Grandjean V, Fourre S, De Abreu DA, Derieppe MA, Remy JJ, Rassoulzadegan M. RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Scientific reports*. 2015;5:18193.
14. Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature*. 2006;441(7092):469-74.
15. Rodgers AB, Morgan CP, Leu NA, Bale TL. Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proceedings of the National Academy of Sciences*. 2015;112(44):13699-704.
16. Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science*. 2016;351(6271):397-400.
17. Skinner MK, Ben Maamar M, Sadler-Riggelman I, Beck D, Nilsson E, McBirney M, et al. Alterations in sperm DNA methylation, non-coding RNA and histone retention associate with DDT-induced epigenetic transgenerational inheritance of disease. *Epigenetics & chromatin*. 2018;11(1):8.
18. Goodrich RJ, Anton E, Krawetz SA. Isolating mRNA and small noncoding RNAs from human sperm. *Methods in molecular biology*. 2013;927:385-96.
19. Schuster A, Tang C, Xie Y, Ortogero N, Yuan S, Yan W. SpermBase: A Database for Sperm-Borne RNA Contents. *Biology of reproduction*. 2016;95(5):99.

20. Johnson GD, Sendler E, Lalancette C, Hauser R, Diamond MP, Krawetz SA. Cleavage of rRNA ensures translational cessation in sperm at fertilization. *Molecular human reproduction*. 2011;17(12):721-6.
21. Peng H, Shi J, Zhang Y, Zhang H, Liao S, Li W, et al. A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Res*. 2012;22(11):1609-12.
22. Keam SP, Hutvagner G. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. *Life (Basel)*. 2015;5(4):1638-51.
23. Sarker G, Sun W, Rosenkranz D, Pelczar P, Opitz L, Efthymiou V, et al. Maternal overnutrition programs hedonic and metabolic phenotypes across generations through sperm tsRNAs. *Proceedings of the National Academy of Sciences*. 2019;116(21):10547-56.
24. Chen Q, Yan W, Duan E. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nature reviews Genetics*. 2016;17(12):733-43.
25. Krawetz SA, Kruger A, Lalancette C, Tagett R, Anton E, Draghici S, et al. A survey of small RNAs in human sperm. *Human reproduction*. 2011;26(12):3401-12.
26. Kawano M, Kawaji H, Grandjean V, Kiani J, Rassoulzadegan M. Novel small noncoding RNAs in mouse spermatozoa, zygotes and early embryos. *PloS one*. 2012;7(9):e44542.
27. Nixon B, Stanger SJ, Mihalas BP, Reilly JN, Anderson AL, Tyagi S, et al. The microRNA signature of mouse spermatozoa is substantially modified during epididymal maturation. *Biology of reproduction*. 2015;93(4):91.
28. Das PJ, Paria N, Gustafson-Seabury A, Vishnoi M, Chaki SP, Love CC, et al. Total RNA isolation from stallion sperm and testis biopsies. *Theriogenology*. 2010;74(6):1099-106, 106e1-2.
29. Li Y, Li RH, Ran MX, Zhang Y, Liang K, Ren YN, et al. High throughput small RNA and transcriptome sequencing reveal capacitation-related microRNAs and mRNA in boar sperm. *BMC genomics*. 2018;19(1):736.
30. Capra E, Turri F, Lazzari B, Cremonesi P, Gliozzi TM, Fojadelli I, et al. Small RNA sequencing of cryopreserved semen from single bull revealed altered miRNAs and piRNAs expression between High- and Low-motile sperm populations. *BMC genomics*. 2017;18(1):14.
31. Du Y, Wang X, Wang B, Chen W, He R, Zhang L, et al. Deep sequencing analysis of microRNAs in bovine sperm. *Molecular reproduction and development*. 2014;81(11):1042-52.
32. Kuchenbauer F, Morin RD, Argiropoulos B, Petriv OI, Griffith M, Heuser M, et al. In-depth characterization of the microRNA transcriptome in a leukemia progression model. *Genome research*. 2008;18(11):1787-97.
33. Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, et al. Complexity of the microRNA repertoire revealed by next-generation sequencing. *Rna*. 2010;16(11):2170-80.
34. Han BW, Hung JH, Weng Z, Zamore PD, Ameres SL. The 3'-to-5' exonuclease Nibbler shapes the 3' ends of microRNAs bound to *Drosophila* Argonaute1. *Current biology : CB*. 2011;21(22):1878-87.
35. Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, et al. Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome research*. 2011;21(9):1450-61.
36. Neilsen CT, Goodall GJ, Bracken CP. IsomiRs--the overlooked repertoire in the dynamic microRNAome. *Trends in genetics : TIG*. 2012;28(11):544-9.
37. Burroughs AM, Ando Y, de Hoon MJL, Tomaru Y, Nishibu T, Ukekawa R, et al. A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome research*. 2010;20(10):1398-410.
38. Guo L, Yang Q, Lu J, Li H, Ge Q, Gu W, et al. A comprehensive survey of miRNA repertoire and 3' addition events in the placentas of patients with pre-eclampsia from high-throughput sequencing. *PloS one*. 2011;6(6):e21072-e.
39. Tan GC, Chan E, Molnar A, Sarkar R, Alexieva D, Isa IM, et al. 5' isomiR variation is of functional and evolutionary importance. *Nucleic acids research*. 2014;42(14):9424-35.

40. Telonis AG, Loher P, Jing Y, Londin E, Rigoutsos I. Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic acids research*. 2015;43(19):9158-75.
41. Govindaraju A, Uzun A, Robertson L, Atli MO, Kaya A, Topper E, et al. Dynamics of microRNAs in bull spermatozoa. *Reproductive biology and endocrinology : RB&E*. 2012;10:82.
42. Fagerlind M, Stalhammar H, Olsson B, Klinga-Levan K. Expression of miRNAs in Bull Spermatozoa Correlates with Fertility Rates. *Reproduction in domestic animals = Zuchthygiene*. 2015;50(4):587-94.
43. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome research*. 2008;18(4):610-21.
44. Yamtich J, Heo SJ, Dhahbi J, Martin DI, Boffelli D. piRNA-like small RNAs mark extended 3'UTRs present in germ and somatic cells. *BMC genomics*. 2015;16:462.
45. Chaulk SG, Ebhardt HA, Fahlman RP. Correlations of microRNA:microRNA expression patterns reveal insights into microRNA clusters and global microRNA expression patterns. *Molecular bioSystems*. 2016;12(1):110-9.
46. Yang CC, Lin YS, Hsu CC, Wu SC, Lin EC, Cheng WT. Identification and sequencing of remnant messenger RNAs found in domestic swine (*Sus scrofa*) fresh ejaculated spermatozoa. *Animal reproduction science*. 2009;113(1-4):143-55.
47. Das PJ, Paria N, Gustafson-Seabury A, Vishnoi M, Chaki SP, Love CC, et al. Total RNA isolation from stallion sperm and testis biopsies. *Theriogenology*. 2010;74:1099-106.e2.
48. Miller D, Ostermeier GC, Krawetz SA. The controversy, potential and roles of spermatozoal RNA. *Trends Mol Med*. 2005;11(4):156-63.
49. Lesson6 : Fertilization and foetal development.
50. Cappallo-Obermann H, Schulze W, Jastrow H, Baukloh V, Spiess AN. Highly purified spermatozoal RNA obtained by a novel method indicates an unusual 28S/18S rRNA ratio and suggests impaired ribosome assembly. *Molecular human reproduction*. 2011;17(11):669-78.
51. Johnson GD, Sandler E, Lalancette C, Hauser R, Diamond MP, Krawetz SA. Cleavage of rRNA ensures translational cessation in sperm at fertilization. *Molecular human reproduction*. 2011;17(12):721-6.
52. Russell S, Patel M, Gilchrist G, Stalker L, Gillis D, Rosenkranz D, et al. Bovine piRNA-like RNAs are associated with both transposable elements and mRNAs. *Reproduction*. 2017;153(3):305-18.
53. Lim SL, Qu ZP, Kortschak RD, Lawrence DM, Geoghegan J, Hempfling AL, et al. HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse. *PLoS genetics*. 2015;11(10):e1005620.
54. Feltzin VL, Khaladkar M, Abe M, Parisi M, Hendriks GJ, Kim J, et al. The exonuclease Nibbler regulates age-associated traits and modulates piRNA length in *Drosophila*. *Aging cell*. 2015;14(3):443-52.
55. Wang H, Ma Z, Niu K, Xiao Y, Wu X, Pan C, et al. Antagonistic roles of Nibbler and Hen1 in modulating piRNA 3' ends in *Drosophila*. *Development*. 2016;143(3):530-9.
56. Zhang H, Ali A, Gao J, Ban R, Jiang X, Zhang Y, et al. IsopiRBank: a research resource for tracking piRNA isoforms. *Database (Oxford)*. 2018;2018.
57. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature reviews Molecular cell biology*. 2011;12(4):246-58.
58. Rojas-Ríos P, Simonelig M. piRNAs and PIWI proteins: regulators of gene expression in development and stem cells. *Development*. 2018;145(17):dev161786.
59. Gou LT, Kang JY, Dai P, Wang X, Li F, Zhao S, et al. Ubiquitination-Deficient Mutations in Human Piwi Cause Male Infertility by Impairing Histone-to-Protamine Exchange during Spermiogenesis. *Cell*. 2017;169(6):1090-104 e13.
60. Beyret E, Liu N, Lin H. piRNA biogenesis during adult spermatogenesis in mice is independent of the ping-pong mechanism. *Cell Res*. 2012;22(10):1429-39.

61. Casas E, Cai G, Neill JD. Characterization of circulating transfer RNA-derived RNA fragments in cattle. *Frontiers in genetics*. 2015;6:271.
62. Godoy PM, Bhakta NR, Barczak AJ, Cakmak H, Fisher S, MacKenzie TC, et al. Large Differences in Small RNA Composition Between Human Biofluids. *Cell reports*. 2018;25(5):1346-58.
63. Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Martin DI. Deep Sequencing of Serum Small RNAs Identifies Patterns of 5' tRNA Half and YRNA Fragment Expression Associated with Breast Cancer. *Biomark Cancer*. 2014;6:37-47.
64. Mishima E, Inoue C, Saigusa D, Inoue R, Ito K, Suzuki Y, et al. Conformational change in transfer RNA is an early indicator of acute cellular damage. *Journal of the American Society of Nephrology : JASN*. 2014;25(10):2316-26.
65. Magee RG, Telonis AG, Loher P, Londin E, Rigoutsos I. Profiles of miRNA Isoforms and tRNA Fragments in Prostate Cancer. *Scientific reports*. 2018;8(1):5314.
66. Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science*. 2016;351(6271):391-6.
67. Vojtech L, Woo S, Hughes S, Levy C, Ballweber L, Sauteraud RP, et al. Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. *Nucleic acids research*. 2014;42(11):7290-304.
68. Sharma U, Sun F, Conine CC, Reichholz B, Kukreja S, Herzog VA, et al. Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm. *Developmental cell*. 2018;46(4):481-94 e6.
69. Watson Erica D. Transferring Fragments of Paternal Metabolism to the Offspring. *Cell metabolism*. 2016;23(3):401-2.
70. Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science (New York, NY)*. 2016;351(6271):391-6.
71. Zhang Y, Zhang X, Shi J, Tuorto F, Li X, Liu Y, et al. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nature cell biology*. 2018;20(5):535-40.
72. Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, et al. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell*. 2018;174(2):391-405 e19.
73. Hua M, Liu W, Chen Y, Zhang F, Xu B, Liu S, et al. Identification of small non-coding RNAs as sperm quality biomarkers for in vitro fertilization. *Cell Discovery*. 2019;5(1):20.
74. Stowe HM, Calcaterra SM, Dimmick MA, Andrae JG, Duckett SK, Pratt SL. The bull sperm microRNAome and the effect of fescue toxicosis on sperm microRNA expression. *PloS one*. 2014;9(12):e113163.
75. Steiman-Shimony A, Shtrikman O, Margalit H. Assessing the functional association of intronic miRNAs with their host genes. *Rna*. 2018;24(8):991-1004.
76. Hinske LC, Galante PA, Kuo WP, Ohno-Machado L. A potential role for intragenic miRNAs on their hosts' interactome. *BMC genomics*. 2010;11:533.
77. Elton TS, Selemon H, Elton SM, Parinandi NL. Regulation of the MIR155 host gene in physiological and pathological processes. *Gene*. 2013;532(1):1-12.
78. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. *Genome research*. 2004;14(10A):1902-10.
79. Budach S, Heinig M, Marsico A. Principles of microRNA Regulation Revealed Through Modeling microRNA Expression Quantitative Trait Loci. *Genetics*. 2016;203(4):1629.
80. Steiman-Shimony A, Shtrikman O, Margalit H. Assessing the functional association of intronic miRNAs with their host genes. *RNA (New York, NY)*. 2018;24(8):991-1004.
81. Desvignes T, Batzel P, Berezikov E, Eilbeck K, Eppig JT, McAndrews MS, et al. miRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants. *Trends in genetics : TIG*. 2015;31(11):613-26.

82. Elkayam E, Kuhn CD, Tocilj A, Haase AD, Greene EM, Hannon GJ, et al. The structure of human argonaute-2 in complex with miR-20a. *Cell*. 2012;150(1):100-10.
83. Kim B, Ha M, Loeffl L, Chang H, Simanshu DK, Li S, et al. TUT7 controls the fate of precursor microRNAs by using three different uridylation mechanisms. *The EMBO journal*. 2015;34(13):1801-15.
84. Li M, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, et al. Widespread RNA and DNA sequence differences in the human transcriptome. *Science*. 2011;333(6038):53-8.
85. Niavarani A, Currie E, Reyal Y, Anjos-Afonso F, Horswell S, Griessinger E, et al. APOBEC3A Is Implicated in a Novel Class of G-to-A mRNA Editing in WT1 Transcripts. *PLoS one*. 2015;10(3):e0120089.
86. García-Herrero S, Meseguer M, Martínez-Conejero JA, Remohí J, Pellicer A, Garrido N. The transcriptome of spermatozoa used in homologous intrauterine insemination varies considerably between samples that achieve pregnancy and those that do not. *Fertility and sterility*. 2010;94(4):1360-73.
87. Trontti K, Vaananen J, Sipila T, Greco D, Hovatta I. Strong conservation of inbred mouse strain microRNA loci but broad variation in brain microRNAs due to RNA editing and isomiR expression. *Rna*. 2018;24(5):643-55.
88. Desvignes T, Batzel P, Berezikov E, Eilbeck K, Eppig JT, McAndrews MS, et al. miRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants. *Trends in genetics : TIG*. 2015;31(11):613-26.
89. Gilbert I, Bissonnette N, Boissonneault G, Vallée M, Robert C. A molecular analysis of the population of mRNA in bovine spermatozoa. *Reproduction*. 2007;133:1073-86.
90. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic acids research*. 2012;40(1):37-52.
91. Le Guillou S, Marthey S, Laloe D, Laubier J, Mobuchon L, Leroux C, et al. Characterisation and comparison of lactating mouse and bovine mammary gland miRNomes. *PLoS one*. 2014;9(3):e91938.
92. Rosenkranz D. piRNA cluster database: a web resource for piRNA producing loci. *Nucleic acids research*. 2016;44(D1):D223-30.
93. Perrier JP, Sellem E, Prezelin A, Gasselin M, Jouneau L, Piumi F, et al. A multi-scale analysis of bull sperm methylome revealed both species peculiarities and conserved tissue-specific features. *BMC genomics*. 2018;19(1):404.
94. Wang J, Vasaiakar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic acids research*. 2017;45(W1):W130-W7.
95. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003;13(11):2498-504.
96. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 2014;15(12):550.
97. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B*. 1995;57(1):289-300.

## Additional figures S1 to S11 (voir les Annexes)

**Additional Fig S1. RNA quality control.** Bioanalyzer profil and RTqPCR with miRNA-125b-5p primers were performed to validate RNA size, quality and concentrations, starting from 5 ng of total RNA, as estimated by Qubit® RNA HS Assay Kit. **(a)** Typical electrophoretic profiles were obtained, with the expected dome around 20 nucleotides and no evidence of 18S and 28S rRNAs. **(b)** Consistent amplification results were obtained (Ct in the range 20-21), indicating that the Qubit –estimated concentration was reliable and that no inhibitors remain in the RNA preparation. Single peak melting curves were also obtained, indicating that a single, specific product has been produced.

**Additional Fig S2. NGS data quality controls.** **(a)** Phred Quality score distribution over all sequences for the 10 first samples, showing Q score above 30 for more than 97% of reads. **(b)** Read length distribution for the 10 first samples, showing two main peaks at 18-26nt (microRNA or siRNA) and 28-32nt (piRNA or tsRNAs).

**Additional Fig S3. Reads annotated as mRNA fragments.** About 24992 reads have been identified as mRNA fragments. **(a)** The vast majority of these reads were 50nt in length. **(b)** IGV profiles of two genes (AKAP1 and PRM1) covered by well distributed unique reads having high count levels.

**Additional Fig S4. Reads annotated as rRNA.** Distribution of reads across 18S **(a)** and 28S **(b)** rRNAs show a particular pattern made of several peaks and read-poor sub regions, suggesting that rRNAs are fragmented by selective RNA cleavage.

**Additional Fig S5. Biogenesis of tRFs and classification of fragments.** The cloverleaf structure of a tRNA typically contains a D-Loop, an Anticodon-Loop, a variable loop, a T-loop, and an amino acid acceptor stem. **(a)** Several endonucleases can cleave tRNAs at specific sites, generating tRFs of different categories: cleavages of an RNase at the D-loop or T-loop of a tRNA can generate a 5' or 3' tRFs, respectively. Dicer (DCR) endonucleases have been reported to cleave tRNA at D-Loop, T-Loop and the amino acid acceptor stem. RNase Z cleavage has been shown to produce tRF3s in a Dicer-independent manner. Moreover, Angiogenin (ANG), a member of the RNase A superfamily, was shown to cleave the Anticodon-Loop to produce tRNA halves upon stress stimuli. ANG is also able to cleave the T-loop of tRNAs. Other unknown RNases might also participate in tRFs generation. **(b)** Multiple tsRNAs alignments along the tRNA<sup>4174</sup>-LeuAAG sequence (D-Loop in red, T-Loop in yellow, Variable-Loop in green and amino acid acceptor stem in violet), illustrating the five tRFs categories: tRF5 and tRF3 (~15–32nt fragments), 5'-tRHs and 3'-tRHs (30-35nt) and i-tRFs. Reads that didn't fall into these categories were classified as other, possibly including full-length tRNAs (limited to the first 50th nucleotides due to the sequencing protocol). Post-transcriptional modifications and/or genetic polymorphisms may also affect tRFs (nucleotide depicted in red). Criteria used to define the categories are depicted below the multiple alignment. Full line indicates the mandatory region that should be covered by the tRF, while dot line indicate the range that can be covered by the tRF.

**Additional Fig S6. Frequency of tRFs isotypes.** The percentage of read counts was computed for each tRF and each associated anticodon, according to the tRF length. Only the most expressed tRFs are reported in the histogram. The left axis (0-40%) refers to 5p-tRHs, which are the most expressed tsRNAs. The right axis (0-10%) refers to the other tsRNAs.

**Additional Fig S7. Distribution of cleavage sites along the anticodon-Loop,** for both 5p-tRHs and 3p-tRHs, showing a bias towards 5p of the anticodon and a high frequency of cleavage at the 4<sup>th</sup> and 7<sup>th</sup> nucleotides.

**Additional Fig S8. Distribution of Pearson correlation for all and clustered miRNAs.** Correlation coefficients were computed for the top 1580 miRNAs having a mean expression level above 10 in at least one breed (all miRNA) and clustered miRNAs. Correlations of a miRNA with itself were omitted. Stronger correlations were observed between genomically clustered miRNA compared to non-clustered miRNAs.

**Additional Fig S9. Five putative functional regulation networks.** A search was performed to identify specific patterns of correlation (above |0.7|) indicative of putative functional regulation (miR-x -> miR-y -> miR-z implies negative correlation between miR-x and miR-y as well as miR-y and miR-z, while a positive correlation is expected between miR-x and miR-z).

**Additional Fig S10. Classification and examples of isomiRs for bta-chr29\_51574.** Multiple alignments along the Genome and the Pre-miRNA sequence are shown, illustrating the diversity of changes occurring in isomiRs and clarifying the nomenclature used in the text.

**Additional Fig S11. Frequency of substitution and nucleotide usage in miRNA sequences.** The frequency of each nucleotide was computed based on all canonical miRNA sequences and compared to the frequency of substitution in isomiRs. Ratio above or under 1 are suggestive of non-random substitutions. Substitutions involving C and U appeared to be 50% and 10% more frequent than expected by chance, while G substitutions appeared 20% less frequent than expected.

## Additional tables S1 to S19

**Additional table S1:** The Sperm miRnome. Catalogue of miRNAs of bovine frozen sperm.

**Additional table S2:** Sperm rRNAs. Catalogue of rRNAs of bovine frozen sperm.

**Additional table S3:** Sperm tRNAs. Catalogue of tRNAs of bovine frozen sperm.

**Additional table S4:** Sperm piRNAs. Catalogue of piRNAs of bovine frozen sperm.

**Additional table S5:** Other sperm small RNAs. Catalogue of other small non coding RNA (without rRNA, tRNA and piRNAs)

**Additional table S6:** Sperm mRNAs. List of reads corresponding to mRNA fragments of ejaculated bovine sperm.

**Additional table S7:** Table of miRNAs expression in bovine frozen sperm.

**Additional table S8:** miRNA genomic clusters. List of genomic clusters identified in bovine sperm.

**Additional table S9:** Expression correlations between miRs.

**Additional table S10:** Expression correlations between miRs located in genomic clusters

**Additional table S11:** Expression of sperm isomiRs across 6 cattle breeds.

**Additional table S12:** Distribution of isomiRs expression levels and contribution to miRNA expression (%)

**Additional table S13:** Details (expression and changes) about isomiRs accounting for more than 1% of each miRNA expression level

**Additional table S14:** DeSeq2 results. List of sncRNAs differentially expressed between breeds.

**Additional table S15:** RTqPCR validation: list of LNA primers, normalized expression and statistics.

**Additional table S16:** Sperm functional parameters across breeds

**Additional table S17:** Comparison of DEmiR across breeds and DEmir published in Capra et al.

**Additional table S18:** Correlation of sperm parameters with miRNAs and piRNAs expression across breeds.

**Additional table S19:** Comparison of miRNA content and expression with Capra et al.

# Chapitre II : L'origine des sncRNA du spermatozoïde bovin éjaculé

## 1) Éléments de contexte

Nous avons établi précédemment le catalogue exhaustif des sncRNA du spermatozoïde. Au regard du nombre important de séquences identifiées, la mise en place de différents « filtres » s'avère nécessaire afin de concentrer nos actions sur les séquences les plus prometteuses. À ce titre, l'origine tissulaire des sncRNA, potentiellement en lien avec leur fonction, constitue une information utile pour développer une stratégie de priorisation en entonnoir. La question est donc de savoir si les sncRNA du spermatozoïde éjaculé sont des vestiges de la spermatogenèse (donc pas totalement éliminés lors de la spermiogenèse) ou s'ils sont acquis au cours de leur transit dans les voies post-testiculaires (potentiellement utiles pour la fertilité et le développement embryonnaire précoce).

La dynamique d'expression des différents sncRNA portés par le spermatozoïde bovin, tout au long de son voyage dans le tractus mâle et au-delà, a fait l'objet d'une publication en cours de soumission dans le journal *Epigenetics & Chromatin* et vous est présentée ci-dessous.

## 2) Résumé de l'étude

Trois éjaculats de taureaux Holstein ont été récoltés en centre de production de semence le jour de leur réforme. Ce même jour, leurs testicules et épидидymes ont été prélevés à l'abattoir, afin de collecter les spermatozoïdes du parenchyme testiculaire, de la tête, du corps et de la queue de l'épididyme. L'ARN extrait des gamètes mâles suivant leur localisation tissulaire a ensuite été séquencé.

L'analyse du contenu du spermatozoïde de chacune des sections (testicule, tête, corps et la queue de l'épididyme, spermatozoïde éjaculé sans plasma séminal et spermatozoïde éjaculé) a permis d'établir la dynamique d'expression des sncRNA tout au long du tractus génital. Que

ce soit sur le plan de la diversité ou d'expression, les profils changent suivant l'organe ou la section de l'organe considéré. La plus importante modification du contenu spermatique est observée entre le testicule et l'épididyme. Post spermiogenèse, dans le spermatozoïde formé et localisé dans le parenchyme testiculaire, les piRNA représentent plus de 80% des petits ARN non codants. Leur expression s'effondre dès l'entrée du spermatozoïde dans l'épididyme, où ils ne représentent plus que la moitié de l'expression des sncRNA dans la tête et moins de 20% dans la queue de l'épididyme. Plusieurs hypothèses ont été émises pour expliquer cette élimination massive des piRNA, telles que leur dégradation dans le spermatozoïde parenchymaire par le système ubiquitine – AP/C, ou indirectement par leur implication dans l'élimination des ARNm de la spermatogenèse ou par leur évacuation dans la gouttelette cytoplasmique. Cette dynamique d'expression des piRNA n'est pas spécifique aux spermatozoïdes bovins, puisqu'elle a déjà été décrite chez la souris dans les mêmes proportions<sup>194,203</sup>.

Inversement, les expressions post-testiculaire des miRNA, tsRNA et rsRNA augmentent : L'expression relative des miRNA passe de 1% dans le testicule à 38% dans la semence, quand celle de tsRNA passe de 2% à 16% et celle des rsRNA de 11% à 19%. L'analyse de l'expression des iso accepteurs met une fois de plus l'accent sur la sur-représentation des séquences portant une Glycine.

Un regroupement des sncRNA en fonction de leur profil d'expression a été réalisé par une approche de *Kmeans* clustering, permettant une analyse au-delà des tendances générales, mettant en particulier en lumière les apports à chaque section du tractus génital. Ces travaux mettent en évidence l'ensemble des séquences de miRNA, tRFs ou rRFs gagnées par le gamète en fin de transit dans l'épididyme ou au contact du plasma séminal. Acquis post-spermatogenèse, ces éléments ne contribuent donc pas à la formation du gamète, mais peuvent avoir un rôle spécifique à jouer dans sa maturation et sa fécondance, ou dans le développement embryonnaire. Dans cet ordre d'idée, il est intéressant de constater que les cibles des miRNA des différents groupes (« testiculaire », « épидидymaire » et « éjaculé ») mettent en avant des voies biologiques en accord avec la localisation spermatique dans le tractus : L'ubiquitination et le métabolisme énergétique pour les spermatozoïdes situés dans le parenchyme testiculaire ou dans l'épididyme ; des voies du développement de l'embryon

et de régulation de la balance « pluripotence – différenciation » pour ceux qui sont situés dans la queue de l'épididyme ou éjaculés.

Les connaissances acquises vont nous permettre de prioriser *in fine* les séquences d'intérêt pour à la fois identifier des biomarqueurs et développer de nouveaux programmes de recherche pour démontrer la fonctionnalité de ces séquences et élucider leurs mécanismes d'action. Des expériences de micro-injection de sncRNA dans le zygote pourront par exemple être réalisées pour valider le rôle de quelques sncRNA de la queue de l'épididyme dans le développement embryonnaire, alors que l'on recherchera plutôt une association entre paramètres spermatiques et expression des sncRNA du testicule ou de la tête de l'épididyme.

3) Article en cours de soumission dans *Epigenetics & Chromatin* (janvier 2021)

## Dynamics of cattle sperm sncRNAs during maturation, from testis to ejaculated sperm.

Eli Sellem<sup>1\*</sup>, Sylvain Marthey<sup>2</sup>, Andrea Rau<sup>2</sup>, Luc Jouneau<sup>3,4</sup>, Aurelie Bonnet<sup>1</sup>, Chrystelle Le Danvic<sup>1</sup>, Benoit Guyonnet<sup>5</sup>, H  l  ne Kiefer<sup>3,4</sup>, H  l  ne Jammes<sup>3,4</sup> and Laurent Schibler<sup>1</sup>

<sup>1</sup> R&D Department, ALLICE, 149 rue de Bercy, 75012 Paris, France.

<sup>2</sup> Universit   Paris-Saclay, AgroParisTech, INRAE, GABI, 78350 Jouy-en-Josas, France.

<sup>3</sup> Universit   Paris Saclay, UVSQ, INRAE, BREED, 78350 Jouy en Josas, France.

<sup>4</sup> Ecole Nationale V  t  rinaire d'Alfort, BREED, 94700 Maisons-Alfort, France.

<sup>5</sup> R&D Department,   volution XY, rue Eric Tabarly, 35538, Noyal sur Vilaine, France

\* Corresponding author

## Abstract

### Background:

During epididymal transit, spermatozoa go through several functional maturation steps, resulting from interactions with epididymal secretomes specific of each region. In particular, the sperm membrane is under constant remodelling, with sequential attachment and shedding of various molecules provided by the epididymal lumen fluid and epididymosomes, which also deliver sncRNA cargo to sperm. As a result, the sperm sncRNAs payload change during the transit from the epididymis caput to the distal cauda. This work was designed to study the dynamics of cattle sperm sncRNAs from spermatogenesis to final maturation.

## Results:

Comprehensive catalogues of sperm sncRNAs were obtained from testicular parenchyma, epididymal caput, corpus and cauda, as well as ejaculated semen from three Holstein bulls. The primary cattle sncRNA sperm content is markedly remodelled as sperm mature along the epididymis. Expression of piRNAs, which are abundant in testis parenchyma, decreases dramatically at epididymis. Conversely, sperm acquires progressively miRNAs, rsRNAs and tsRNAs along epididymis, with regional specificities. For instance, miRNAs and tsRNAs are enriched in epididymis cauda and ejaculated sperm, while rsRNA expression peaks at epididymis corpus. In addition, epididymis corpus contains mainly 20nt long piRNAs, instead of 30nt in all other locations. Beyond the bulk differences in abundance of sncRNAs classes, K-means clustering was performed to study their spatiotemporal expression profile, highlighting differences in specific sncRNAs and providing insights into their putative biological role at each maturation stage. For instance, Gene Ontology analyse using miRNA targets highlighted enriched processes such as cell cycle regulation, response to stress and ubiquitination processes in testicular parenchyma, protein metabolism in epididymal sperm and embryonic morphogenesis in ejaculated sperm.

## Conclusions:

Our findings confirm that the sperm sncRNAome does not simply reflect a legacy of spermatogenesis. Instead, sperm sncRNA expression shows a remarkable level of plasticity resulting from the combination of multiple factors such as loss of the cytoplasmic droplet, interaction with epididymosomes and more surprisingly, the putative *in situ* production and/or modification of sncRNAs by sperm. Given the suggested role of sncRNA in epigenetic transgenerational inheritance, our detailed spatiotemporal analysis may pave the way for the study of sperm sncRNAs role in embryo development.

## Keywords:

Sperm epigenetics, sncRNAs, bovine, miRNA, piRNA, tRFs, rRFs, epididymis location, testicular parenchyma sperm

## Introduction

Spermatozoa are produced from germ cells in the seminiferous tubules of the male testes by a complex stepwise process called spermatogenesis. Spermatogenesis is a tightly regulated developmental process involving thousands of genes and proteins [1] as well as small non-coding RNAs (sncRNAs), especially microRNAs (miRNAs) and piwi-interacting RNAs (piRNAs) [2]. For instance, the selective knock-out of *Dicer1* at onset of male germ cell development leads to infertility, due to multiple cumulative defects at the meiotic and post-meiotic stages of spermatogenesis [3], highlighting the role of microRNAs (miRNAs) and endogenous small interfering RNAs (endo-siRNAs). Likewise, miR-100, the miR-29 family and miR-34c have been shown to be required for spermatogonial stem cells proliferation [4], regulation of meiosis [5] and germinal expression profiles [6], respectively. Likewise, piRNA are known to play roles in spermatogenesis, as evidenced by the mitosis, meiosis, chromatin compaction, flagella elongation and fertility defects in mutants lacking Piwi [7].

Beyond spermatogenesis, spermatozoa go through several temporal functional maturation steps at various locations in order to acquire their fertilizing capacity. In particular, sperm gain motility and acrosomal function whilst in the epididymis. During epididymal transit, the sperm membrane is under constant remodelling, with sequential attachment and shedding of various molecules provided by the epididymal lumen fluid [8] and extracellular vesicles called epididymosomes, including proteins [9, 10] and lipids [11]. In addition, sncRNA cargo are also delivered to sperm by epididymosomes [12, 13]. The head (caput), body (corpus), and tail (cauda) make up the three main regions of the epididymis, which can be distinguished by specific epithelial morphology, luminal diameter, protein [14, 15] and lipid content [11, 16], as well as expression pattern and epididymosome content [17-20]. In addition, miRNA repertoires contained within epididymosomes were shown to differ from those of their parent epithelial cells, suggesting an active sorting of miRNAs released into the intraluminal fluid [21].

Several biological roles have been proposed for epididymosomes, including intercellular communication throughout the epididymis [21], modulation of sperm motility during epididymal transit, protection against oxidative stress, acquisition by spermatozoa of surface proteins essential for fertilization, tagging defective sperm for elimination [22], and supporting embryo development [12, 19, 23]. Epididymal sperm maturation is thus the result of complex and sequential interactions with different epididymal secretomes specific of each

region. As a result, the sperm proteome [24], lipid composition [25] and sncRNAs payload [18, 19] change during the transit from the epididymis caput to the distal cauda.

Given the suggested role of sncRNA in epigenetic transgenerational inheritance phenomenon [26], this work was designed to study the dynamics of cattle sperm sncRNAs from spermatogenesis to final maturation. Comprehensive catalogues of sperm sncRNAs were obtained from testicular parenchyma, epididymal caput, corpus and cauda to infer the potential origin of sncRNAs observed in ejaculated sperm in our previous work [27].

## Methods

### Semen collection

Semen was collected from three Holstein bulls following the standard protocol used by the semen production center (Evolution-XY, France). A batch was frozen following the routine straw production protocol, while another batch was centrifuged to separate seminal plasma from spermatozoa before nitrogen freezing. On the same day, bulls were slaughtered to collect the testis and epididymis and isolate sperm from testicular parenchyma as well as three regions of the epididymis (caput, corpus and coda).

To do so, several small pieces of each regions were incubated 20 min in a Petri dish filled with 3 ml Phosphate Buffered Saline solution (PBS). The supernatant was collected and centrifuged 5 minutes at 2400 x g to recover the cellular pellet, which was washed twice with 1 ml PBS and once with 1 ml water to remove somatic cells. Somatic cells contamination was assessed by microscopy. The final sperm pellet was stored in nitrogen.

### Total RNA isolation, quality control and sequencing

RNA extraction was performed as previously described [27]. Briefly, sperm pellets were homogenized in 100µl of RLT buffer (Qiagen) supplemented with 1 µl beta-mercaptoethanol and incubated 15 minutes at room temperature. Then, 1 ml of Trizol (Invitrogen) was added and samples were incubated 5 minutes at room temperature before adding 100 µl of

chloroform. After vigorous shaking for 15 seconds, samples were incubated 2 minutes at room temperature and centrifuged at 12 000 x g during 15 minutes at 4°C.

The aqueous phase was transferred to a new collection tube and mixed with 100µl of chloroform by vigorous shaking. After incubation for 2 minutes at room temperature and centrifugation at 12 000 x g during 15 minutes at 4°C, the aqueous phase was recovered and nucleic acids were precipitated overnight at -20°C using 1 volume of isopropanol supplemented with 25 µl glycogen (Ambion AM9510, 5mg/ml). After centrifugation at 12 000 x g during 15 minutes at 4°C, pellets were washed with 75% ethanol and dried under vacuum. Dried pellets were then re-suspended in 12µl of RNase free water and incubated 1 hour at 4°C before quality control assessment.

As described previously [27], RNA concentration was assayed using the Qubit® fluorometer (Life Technologies), with the RNA HS Assay Kits (Q32852 Life Technologies) according to the manufacturer's recommendations. In addition, RTqPCR was used to detect and quantify miR-125-5p which has been shown to be highly expressed in bovine sperm. To exclude any contamination by somatic RNAs, RT-qPCR was performed to assess expression of the Epididymal Secretory Glutathione Peroxidase gene (*GPX5*), which is known to be expressed by epididymis epithelium cells. Primers were designed from the Bovine sequence (UCSC genome browser) using Primer3web (<https://primer3.ut.ee>). Their efficiency and specificity were evaluated on bovine genomic DNA.

Library preparation and sequencing was performed by Qiagen, starting from 42 ng total RNA and using the QIAseq miRNA Library Kit® and Unique Molecular Indexes (UMI). UMIs are random oligonucleotide barcodes used to distinguish identical copies arising from distinct molecules from those arising through PCR amplification of the same molecule. UMI deep sequencing was performed on Illumina HiSeq2000, targeting 20 M single-reads (75 bp) per sample.

#### RT-qPCR validation of differential expression among regions

Expression of three miRNAs (bta-miR-100: AACCCGTAGATCCGAACTTGT, bta-miR-16a: TAGCAGCACGTAAATATTGGCG and bta-chr27-30001: CGCCGGGGCGGGTTCCGGAGG) as well as two bta-miR-191 isomiRs (CAACGGAATCCCAAAGC and CAACGGAATCCCAAAGCAG) was assessed by RT-qPCR on 3 testis parenchyma sperm and 3 ejaculated sperm samples to

evaluate their differential expression. LNA primers were supplied by Qiagen. Duplicates were assayed in 10- $\mu$ l / 5ng qPCR reactions following the miRCURY LNA kit protocol, using a StepOnePlus Real time PCR System (Applied biosystems). Amplification curves were analysed using the StepOne software v2.3 to compute Ct values and analyse the melting curve. Relative expression was computed with the qbase<sup>+</sup> software (Biogazelle).

### Bioinformatics and biostatistics

Our previously described bioinformatic pipeline [27] was slightly modified to take UMI technology into account: de-duplication was performed first based on UMI, to remove reads arising through PCR amplification of the same molecule. One exemplar was conserved by UMI, associated with the mean quality score calculated at each position. UMIs and adaptors were trimmed from the remaining sequences (UMI-tools), which were then filtered to remove reads shorter than 17nt and sequences containing at least one nucleotide having a Phred quality score under 25. Remaining sequences were then stacked by unique sequences to compute total read counts and mapped on the bovine reference genome (BosTau9). A workflow mostly based on miRDeep2 [28, 29] was used to identify and quantify known (miRBase v22) as well as predicted miRNAs. Expression levels were computed both at the miRNA and isomiR levels. In parallel, all unique sequences were annotated using several public databases including Ensembl release 94 (<http://www.ensembl.org>), RFAM release 14 (<http://rfam.wustl.edu>), cattle ncRNA from ENA (<https://www.ebi.ac.uk/ena>), cattle tRNA from GtRNadb ([www.gtrnadb.ucsc.edu](http://www.gtrnadb.ucsc.edu)), human, mouse and cattle piRNAs from ENA, as well as [30] and [31]. Targetscan 7.2 (release March 2018) was used to predict miRNAs targets. Functional annotations and enrichment analysis were done using Gorilla [32] (<http://cbl-gorilla.cs.technion.ac.il/>) and WebGestalt [33] (<http://www.webgestalt.org/>). GO results were summarized to produce scatterplots and interactive graphs of enriched terms grouped by semantic similarity using Revigo [34].

The DESeq2 R package [35] was used to normalize read counts and test for differential sncRNA expression between regions. Statistical significance was evaluated based on Benjamini-Hochberg adjusted p-values, with a significance threshold of 5%. The proportion of each sncRNA classes was computed per region. Between-class analysis were performed using the

ade4 R package (dudi.bca function). Kmeans clustering (Lloyd algorithm, 100 random starts and 1000 iterations) was performed using the R (*kmeans()* function) after data normalization (*normalize()* function from R package “som”). Graphs were plotted using the ggplot2 library.

#### Data availability

Fastq files were made available to the scientific community at European Nucleotide Archive (ENA), under the primary accession number: PRJEB41989

## Results

#### Total RNA preparation and UMI NGS sequencing

At least 20 million spermatozoa could be recovered from each sampling region. Contamination with somatic cells was negligible, as confirmed by microscopy (less than 1 somatic cell per 1000 sperm cells) examination and by the *GPX5* ddCTs obtained from the RNA expression comparisons of sperm versus somatic cells (control sample).

About 23 to 1970 ng of total RNA could be obtained from sperm, according to the sampling region. Consistent amplifications of bta-miR-125 (Ct in the range 24-26 starting from 5 ng of total RNA) and single peak melt curve were obtained.

Sequencing resulted in 544 262 139 raw sequence reads, with on average  $32.0 \pm 8.0$  million reads per library. After de-duplication based on UMIs and filtering, 216 202 555 reads were kept, with on average  $12.7 \pm 8.7$  million reads per libraries, corresponding to 1 088 318 unique reads. On average, about 80% of sequences could be mapped unambiguously to the cattle reference genome, while 20% were outmapped (more than five genomic locations, e.g. mapped to repetitive sequences) or unmapped (Table 1).

	<b>Mapped (80%)</b>	<b>Outmapped or Unmapped (20%)</b>	<b>Reads</b>
<b>Reads no.</b>	172 351 212	43 851 343	216 202 555
<b>Mean</b>	10 138 307	2 579 491	12 717 798
<b>Std. deviation</b>	7 588 059	1 158 567	8 697 569

Table 1: Mapping statistics across the 17 libraries.

As described in the Material & Method section, reads were annotated using several databases and classified as micro-RNAs (miRNAs), Piwi-interacting RNAs (piRNAs or piRNA-like), transfer RNA fragments (tRFs), ribosomal RNAs fragments (rRNAs or rRFs), circular RNAs (circRNAs), mRNAs and other RNAs. Sperm sncRNA comprises a wide variety of sncRNA classes, mainly miRNAs, piRNAs, tRFs and rRNAs (Supplementary Table S1-S4). Reads annotated as mRNA accounted for only 0.8% (range 3000 to 17.000 reads, according to the sampling region), highlighting the absence of somatic cells RNA contaminations. Only a small proportion of reads (5.6%) remained unannotated.

*The sncRNA signature discriminates between sperm sampled along the male tract.*

A between class analysis (Figure 1) showed great differences in sperm sncRNA content according to the sampling region, making it possible to discriminate between sperm samples collected at various developmental stage along the male tract. As exemplified by between-class inertia percentages (ratio ranging from 48% to 57%) according to the sncRNA family, a large proportion of variance could be attributed to the sampling region. Interestingly, testis parenchyma and epididymis caput on the one hand as well as epididymis cauda and total ejaculated sperm on the other hand, appeared to be close together whatever the sncRNA family. Epididymis corpus and ejaculated sperm depleted from seminal plasma showed peculiarities in terms of miRNAs, tRNAs and rRNAs content and clearly distinguished from the two other groups. No differences could be highlighted based on piRNA content, except for testis parenchyma and epididymis caput. These observations strongly suggest that sperm sncRNA content is highly dynamic, sncRNA families exhibiting particular trends along the male tract.

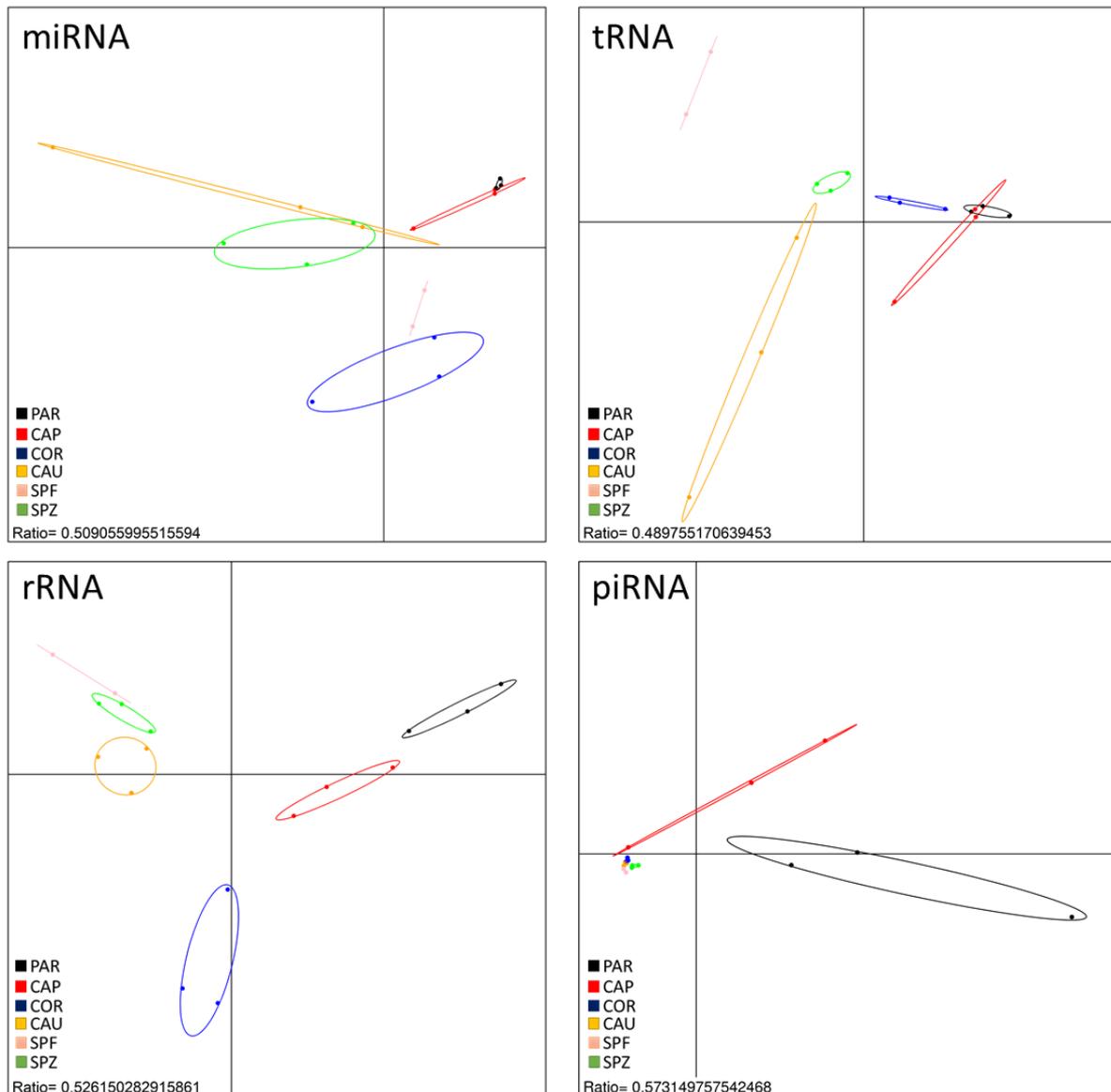


Fig 1: sncRNA signature discriminates between sperm sampled along the male tract. Samples could be distinguished by BCA analysis, whatever the sncRNA family, except for piRNA which failed to discriminate between epididymis corpus, cauda and ejaculated sperm. Interestingly, testis parenchyma and epididymis caput on the one hand as well as epididymis cauda and total ejaculated sperm on the other hand, appeared to be close together, while epididymis corpus and ejaculated sperm depleted from seminal plasma exhibited a particular profile.

Dynamics of cattle sperm sncRNAs from spermatogenesis to final maturation.

As illustrated by Figures 2, sperm sncRNA content varies greatly between the different regions. In particular, piRNAs were shown to account for 84% (12 106 755 normalized count) of all sncRNAs contained in testis parenchyma sperm and decrease along the epididymis from 57%

(5 064 065 normalized count) at caput to 13% (1 837 823 normalized count) at cauda. A slight increase in proportion was observed in ejaculated sperm (18%, 1 104 327 normalized count).

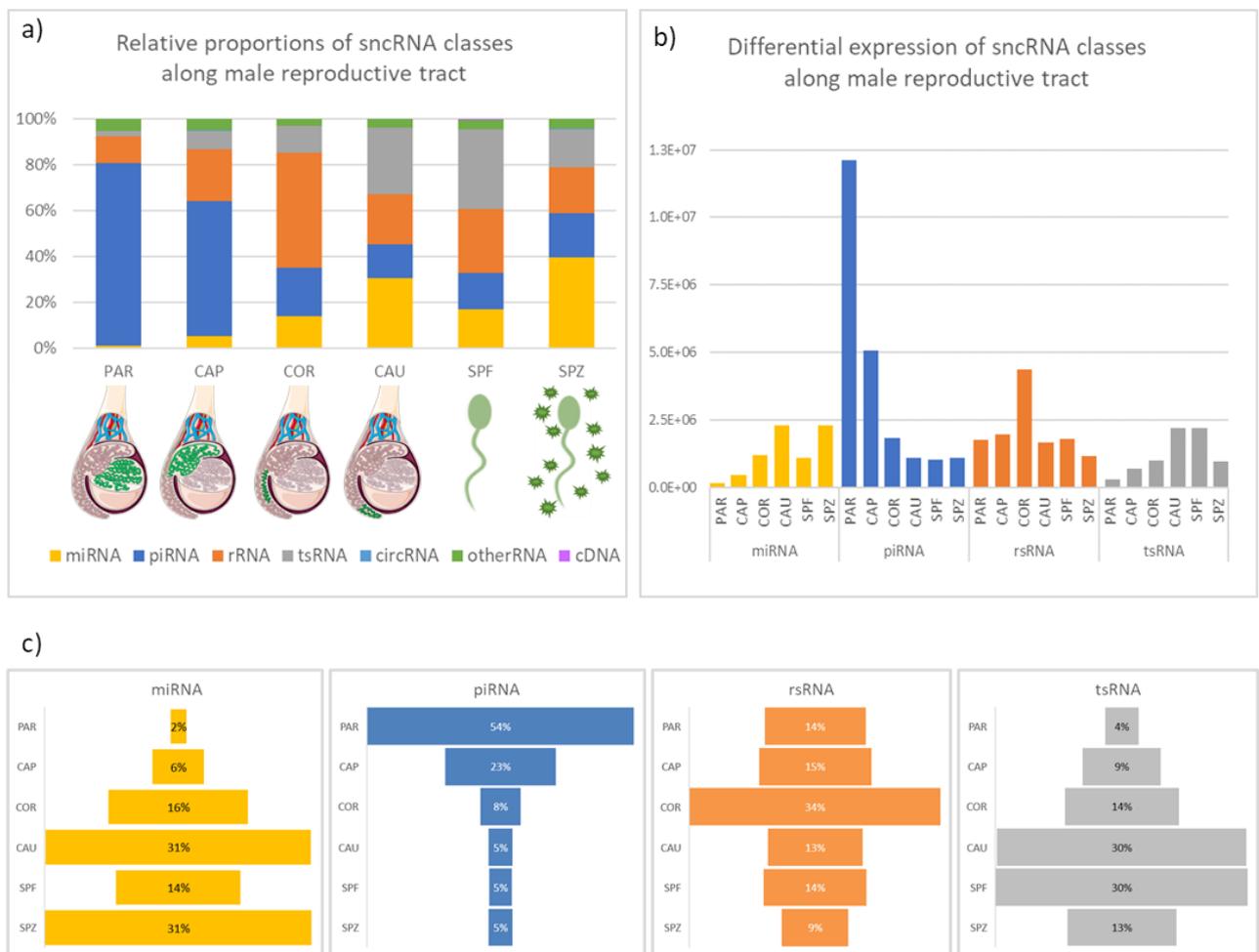


Fig 2 : Changes in sncRNA expression along the male reproductive tract. Relative proportions (%) of sperm sncRNA were computed for each of the six sampling regions: PAR (testis parenchyma), CAP (caput), COR (corpus) and CAU (cauda) epididymis, ejaculated sperm depleted from seminal plasma (SPF) and total ejaculated spermatozoa (SPZ). Bargraphs clearly illustrates opposite trends between piRNAs whose proportion decrease from parenchyma (84%) to cauda (13%) and ejaculated sperm (18%), and miRNAs and tRNAs, which follow an overall upward trend from a few % in parenchyma to 38% and 32% in ejaculated sperm, respectively. Interestingly, differences were observed between ejaculated sperm depleted from seminal plasma and total ejaculated sperm, suggesting that seminal plasma may contain piRNAs and miRNAs. b) Expression level of each sncRNA class along the male tract. b) Total normalized expression of each sncRNA class for each sampling region. c) Proportions of each sncRNA class expression for the six sampling regions.

The mean Fold Change between PAR and SPZ was computed for downregulated piRNA and compared among the four PAR expression quartile groups. As illustrated in Table 2, highly expressed piRNAs exhibited a more pronounced decrease in expression.

Quartile group	Mean PAR expression	Mean Fold Change PAR vs SPZ
1	1.8	7
2	4.1	14
3	10.1	28
4	96.9	76

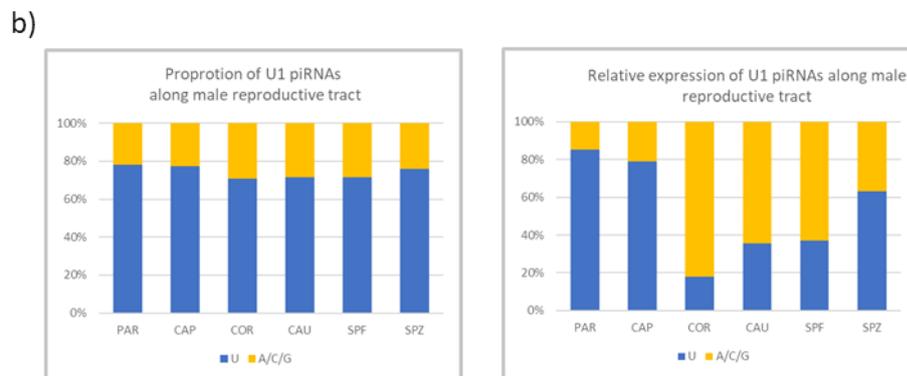
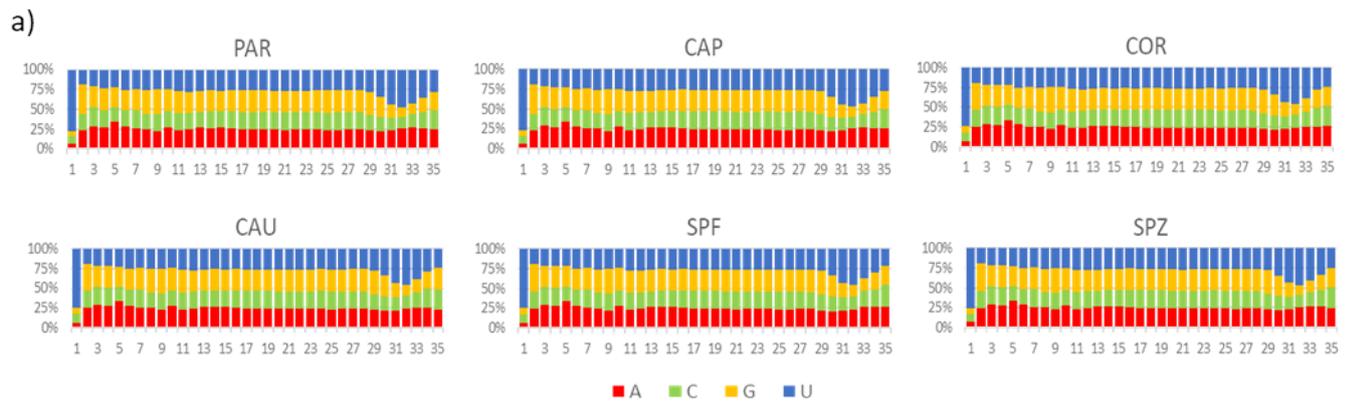
Table 2: Expression ratio for downregulated piRNAs between parenchyma and ejaculated sperm

Interestingly, a great variation of piRNA number was observed across regions, with thrice as many sequences in PAR and CAP and twice in SPZ as in CAU.

	PAR	CAP	COR	CAU	SPF	SPZ
Total expression	12042620	5039827	1832432	1095920	1023744	1097840
Number of piRNAs	724019	721321	235040	201673	179805	385354

Table 3: Abundance and diversity of sperm piRNA along the male tract.

The nucleotide composition of the piRNA pool was assessed for each of the six sampling regions to gain insight into the mechanism through which piRNA were generated during sperm transit. Notably, whatever the region, more than 75% of piRNA were determined to bear an uracil (U) residue at their 5' end (U1 piRNAs) and no enrichment for an A residue at piRNA position 10 (A10) was observed (Figure 3a). Noteworthy, U1 piRNAs account for about 80% of piRNAs whatever the region, but a great variation in their relative expression was observed across regions. Indeed, while they account for about 80% of piRNA expression in PAR and CAP, this ratio was almost inverted in COR (Figure 3b) and the expression of piRNAs was shown to be dominated by non U1 piRNAs in COR, CAU and SPF.



**Fig 3: Nucleotide frequency at each position for all detected piRNA in each sampling region. a)** For each region, the frequency of each nucleotide was computed along the piRNA sequence and plotted as bar charts (from the 1st nucleotide at 5' end to the 35th nucleotide). More than 75% of piRNA were shown to possess an U residue at their 5' end (U1), and no enrichment for an A residue at position 10 (A10) was observed. **b)** U1 piRNAs generated via the PIWIL1-directed primary production pathway account for about 80% of piRNAs whatever the region, but a great variation in their relative expression was observed across regions: while they account for about 80% of piRNA expression in PAR and CAP, this ratio was almost inverted in COR, expression of piRNAs being dominated by non U1 piRNAs in COR, CAU and SPF.

In contrast, an increase in miRNAs and tRNAs content was observed along the male tract. Indeed, while a low sperm miRNAs content was observed in testis parenchyma (1%, 140 591 normalized count), increasing miRNA expression was observed along the epididymis (5%, 13% and 27% In caput, corpus and cauda, respectively) to reach 38% in ejaculated sperm (2 291 805 normalized count). Likewise, sperm tRNAs content increase from 2% in testis (297 555 reads), 8% in epididymis caput (683 656 reads), 10% in corpus (997 131 reads) and 26% in cauda (2 178 437 reads).

Interestingly, a higher proportion of piRNAs and miRNAs was observed in total ejaculated sperm (18% and 38% for piRNAs and miRNAs, respectively), compared to ejaculated sperm depleted from seminal plasma (15% for both piRNAs and miRNAs), suggesting that seminal plasma may contain piRNAs and miRNAs. In contrast, the proportion of tRNAs appeared to be higher in ejaculated sperm depleted from seminal plasma, compared to total ejaculated sperm (32% versus 16%).

While piRNAs, miRNAs and tRNAs expression seemed to follow an overall downward or upward trend, rRNAs showed a particular profile, peaking at epididymis corpus (4 386 564 reads, 46% of sperm sncRNAs).

Other sncRNAs did not display great differences among regions, accounting for only a small proportion of sperm sncRNAs.

#### *In depth study of sperm sncRNA spatiotemporal dynamics*

Differential expression analysis was performed using DeSeq2 for all pairwise comparisons between sampling regions. Table 4 summarizes the number of differentially expressed sncRNA between each region (adjusted p-values <0.05, fold change > ±2.5).

	Testis	Epididymis			Ejaculated sperm	
	PAR	CAP	COR	CAU	SPF	SPZ
PAR	NS	4.436	15.6611	120.914	85.379	194.062
CAP		NS	52.471	77.264	65.132	82.921
COR			NS	5.209	9.471	21.074
CAU				NS	1.284	1.188
SPF					NS	106
SPZ						NS

Table 4: Number of sperm sncRNAs differentially expressed between the six sampling regions: testis parenchyma (PAR), epididymis caput (CAP), corpus (COR), cauda (CAU), ejaculated sperm depleted from seminal plasma (SPF) and total ejaculated sperm (SPZ).

Consistent with the aforementioned global trends, the number of differentially expressed sncRNA was shown to increase with the spatiotemporal distance between samples: while only 106 differentially expressed sncRNAs were observed between ejaculated sperm with or without seminal plasma, comparisons with sperm sampled at epididymis cauda, corpus and caput highlighted 1.188, 21.074 and 82.921 and differentially expressed sncRNAs, respectively. Likewise, 194.062 differentially expressed sncRNAs were observed between ejaculated sperm and sperm sampled at the testis parenchyma. The RTqPCR expressions level for five isomiRs over-expressed in SPZ compared to PAR, confirm the NGS results (table 5).

isomiR	NGS_FC	NGS_adj_pvalue	qPCR_FC	qPCR_pvalue
<b>bta-miR-16a TAGCAGCACGTAAATATTGGCG</b>	47.3	<0.05	2.8	<0.05
<b>bta-miR-100 AACCCGTAGATCCGAACTTGT</b>	194.5	<0.05	4.4	<0.05
<b>bta-miR-191 CAACGGAATCCCAAAGCAG</b>	330.9	<0.05	10.7	<0.05
<b>bta-chr27_30001 CGCCGGGGCGGGTTCCGGAGG</b>	604.3	<0.05	3.3	<0.05
<b>bta-miR-191 CAACGGAATCCCAAAGCAG</b>	757.1	<0.05	28.5	<0.05

Table 5: Fold changes SPZ versus PAR and pvalues of five differentially expressed miRNAs, measured by NGS and RTqPCR. The RTqPCR analysis confirmed a significant over-expression of the five isomiRs in SPZ compared to PAR.

Consistent with the above findings, piRNA account for the majority of these differentially expressed sncRNA. For instance, they account for 67% of the 194 062 sncRNA differentially expressed between testis parenchyma (PAR) and ejaculated sperm (SPZ), while rRNAs, tsRNAs and miRNAs account for 16%, 4% and 3%, respectively. As illustrated in Figure 4, 83% of differentially expressed sncRNAs are under-expressed in SPZ, mainly piRNA (64%) and rRNA (11%). Among over-expressed sncRNAs in SPZ, rRNAs, piRNAs, tsRNAs and miRNAs account for 32%, 17%, 15% and 14%, respectively.

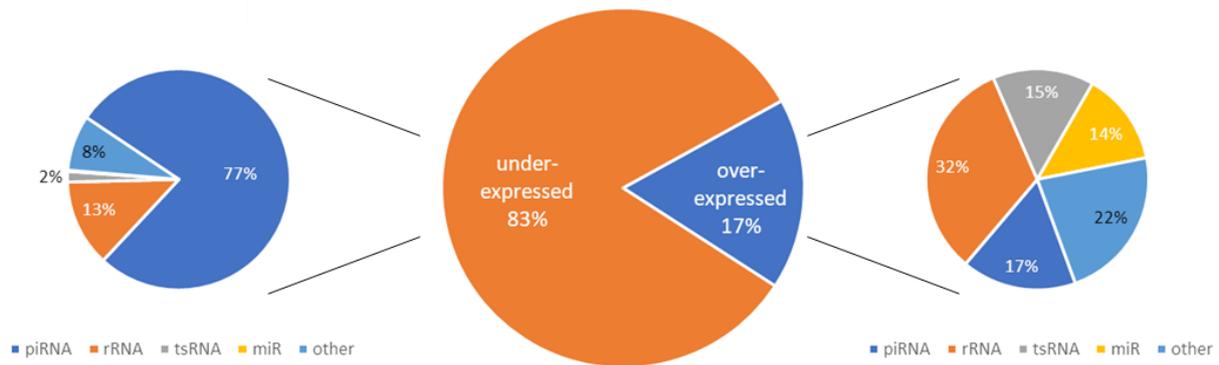


Fig 4: Proportion of sncRNAs families among differentially expressed sncRNA between testis parenchyma and ejaculated sperm. Differentially expressed sncRNA are mostly under-expressed (83%, middle pie chart). Among them 77% are piRNAs, 13% rRNAs and 2% tRNAs (left pie chart). Among over-expressed sncRNAs, 32% are rRNAs, 17% piRNAs, 15% tRNAs and 14% miRNAs (right pie chart).

While differentially expressed tsRNA sequences covered isoforms associated with 21 amino acids, four isoacceptors contributed to 69% of all identified tsRNA (Supplementary Table S5): Glycine (29% of read counts), Glutamine (17%), Methionine (13%) and Serine (10%). Differences in relative expression were also observed across regions, as illustrated in Figure 5a. For instance, Alanine, Histidine and Threonine are mainly expressed in epididymis corpus, while Arginine, Glycine, Isoleucine and Methionine are mostly expressed in epididymis cauda. Glutamine and Glutamate isoacceptors account for 30% of expressed tsRNA in ejaculated sperm, while they account only for 8% in epididymis caput. Glycine isoacceptor is the most expressed tsRNA whatever the region, ranging from 20% in ejaculated sperm to 35% in epididymis cauda. Likewise, tsRNA subgroups showed particular expression profiles, as exemplified by tRF5s, tRF3s and 3'-tRHs which are mainly expressed in sperm depleted from seminal plasma, while i-tRFs are mainly expressed in epididymis cauda. Altogether, i-tRFs accounted for 55% of differentially expressed tsRNA (Figure 5b).

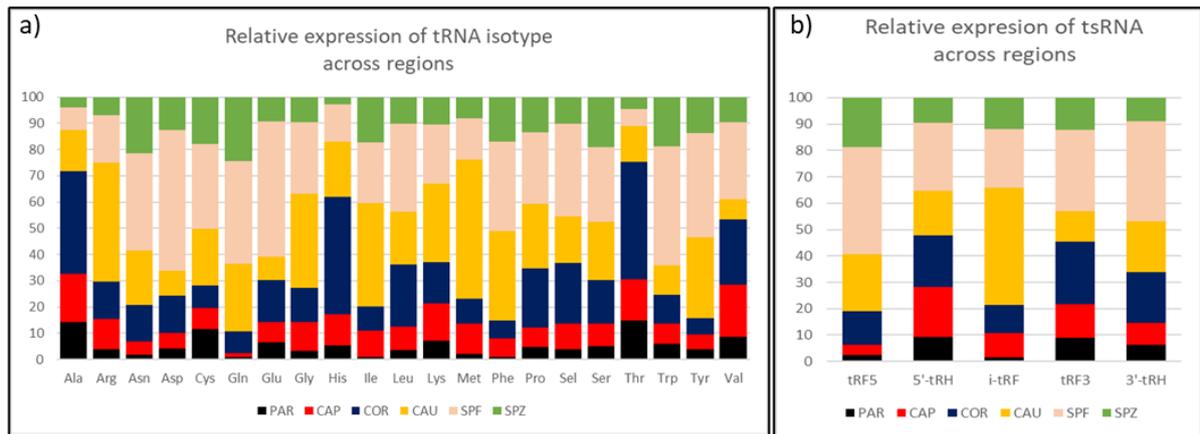


Fig 5: Dynamics of tsRNA across regions. **a)** Distribution across regions of isoacceptor read counts, showing specific expression pattern. For instance, Alanine, Histidine and Threonine are mainly expressed in epididymis corpus, while Arginine, Glycine, Isoleucine and Methionine are mostly expressed in epididymis cauda and Asparagine, Cysteine, Glutamate, Glutamine and Tryptophan are mostly expressed in sperm depleted from seminal plasma. **b)** Distribution across regions of tsRNA read counts, showing that tRF5s, tRF3s and 3'tRHs are mainly expressed in sperm depleted from seminal plasma, while i-tRFs are mainly expressed in epididymis cauda.

About 37% and 22% of differentially expressed rRNA reads were annotated as 28S and 18S, respectively (Supplementary Table S6). Specific expression patterns were observed, as illustrated in Figure 6: fragments derived from 28s rRNA account for the majority of read counts in parenchyma (63%) and epididymis caput (53%); likewise, fragments derived from 16s rRNA account for a large proportion of read counts in sperm depleted from seminal plasma (42%) and ejaculated sperm (48%), while 5.8S account for 40% of reads in epididymis corpus.

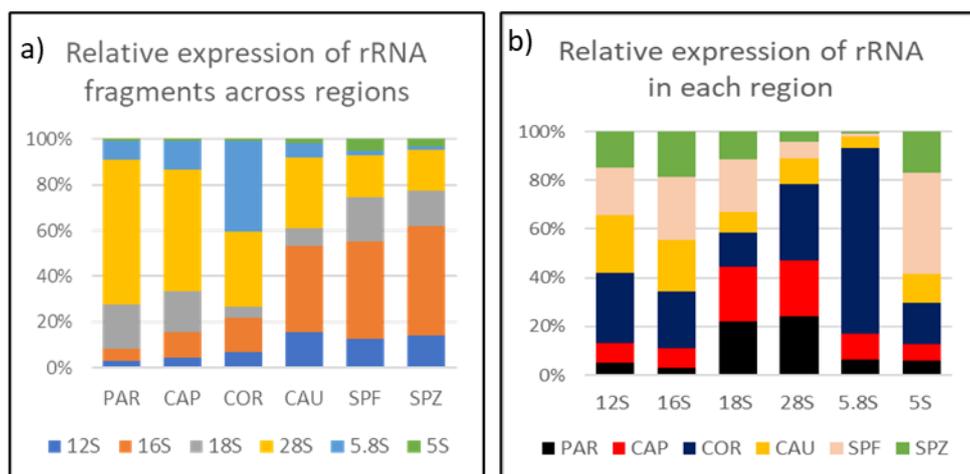


Fig 6: Dynamics of rRNA fragments across regions. **(a)** Proportions of read counts from each rRNA fragment subtype was computed per region, showing that fragments derived from 28s rRNA account

for the majority of read counts in parenchyma and epididymis caput, while fragments derived from 16s rRNA account for a large proportion of read counts in epididymis cauda, sperm depleted from seminal plasma and ejaculated sperm. **(b)** Proportions of read counts from each region was computed per rRNA fragment subtype, illustrating specific patterns of rRNA expression. For instance, 5.8S derived fragments are primarily expressed in epididymis corpus, while 5S are mostly expressed in sperm without seminal plasma.

### Biological processes and pathways associated with differentially expressed sncRNA

K-means clustering was performed with all sncRNA class to cluster differentially expressed sncRNAs in 14 groups according to their expression profile and search for particular expression patterns as well as enrichment in specific biological processes and pathways. As illustrated in Figure 7, a diversity of expression profiles was observed beyond the general trend.

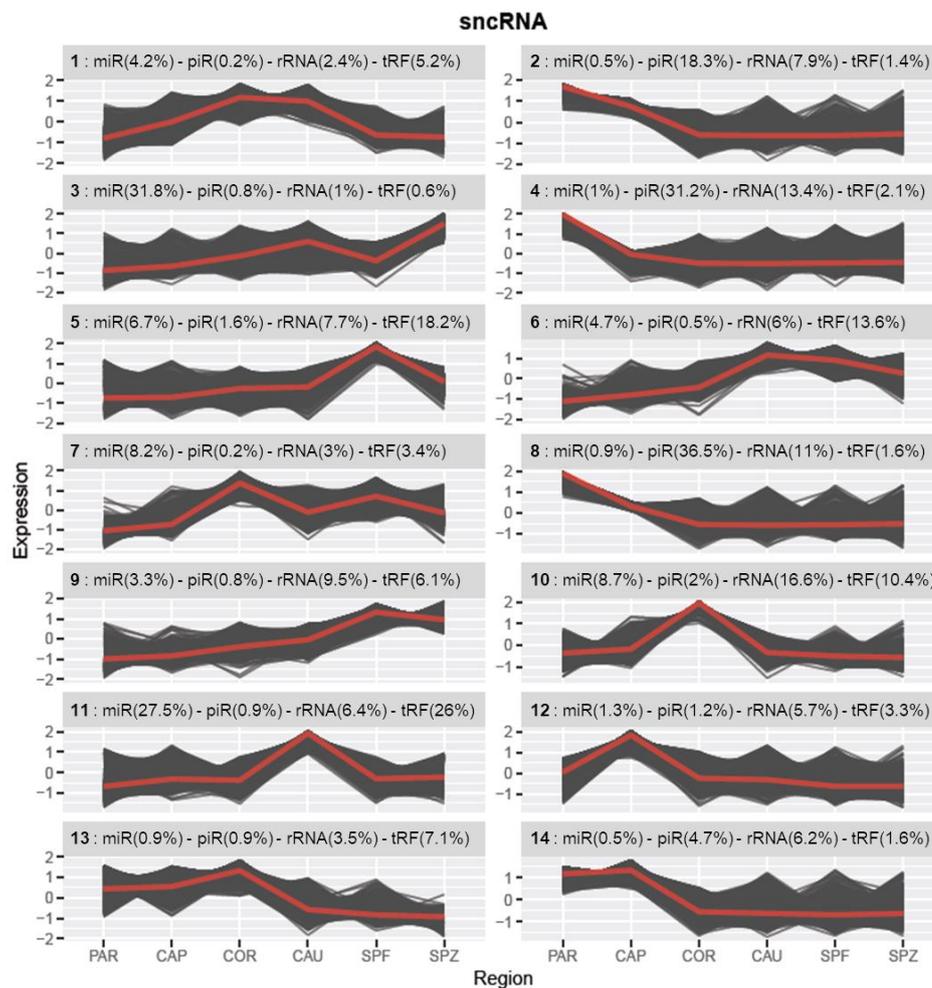


Fig 7: K-means clustering illustrates the diversity of expression profiles among sncRNAs. For each sncRNA class, expression plots were drawn for each cluster (cluster Name and % of sncRNA included

in the cluster are provided on top). Beyond the general trend a diversity of expression profiles was observed. For instance, miRNAs were shown to increase from parenchyma to epididymis cauda (43% of miRNAs) or semen (35% of miRNAs), in good agreement with the general trend observed in Figure 2. However, 3% of miRNA were found to decrease along the male tract (clusters 2,4,8,14). In contrast, while 86% of piRNAs were found to decrease in good agreement with the global trend (clusters 2;4,8), 15% of piRNAs showed a more complex profile, with a transient increased expression at either epididymis caput (6%), corpus (3.1%), cauda (2.5%) or ejaculated semen (3.5%).

For instance, miRNA showed mainly increasing trends, peaking at epididymis cauda (43%, clusters 1,5,6,11) or semen (35%, clusters 3,9). However, 6% of miRNA were found to decrease in relative expression along the male tract (clusters 2,4,8). In contrast, piRNA were found to decrease in relative expression (86%, clusters 2,4,8) or showed a transient increase in relative expression from parenchyma to epididymis caput (6% cluster 12,14), corpus (3.1% clusters 7,10,13) or cauda (2.5% clusters 1,3,6,11). Only 3.5% of piRNA were found to increase in relative expression. Likewise, predominant tsRNA expression was identified at CAP and SPF, with clusters 6 & 11 (peaking at CAU) and clusters 5 & 9 (peaking at SPF) gathering 64% of tsRNA. In contrast a more balanced distribution was observed for rsRNA : 33% of rsRNAs were found in downregulated clusters, 28% in clusters peaking at COR (clusters 1, 7, 10, 13), 12% in clusters peaking at CAU (clusters 6 and 11) and 18% showed an increased post epididymal expression (clusters 3, 5 and 9). It is noteworthy, among downregulated sncRNAs higher fold changes were observed for piRNAs between CAP and COR, compared to other sncRNA classes (Figure 8).

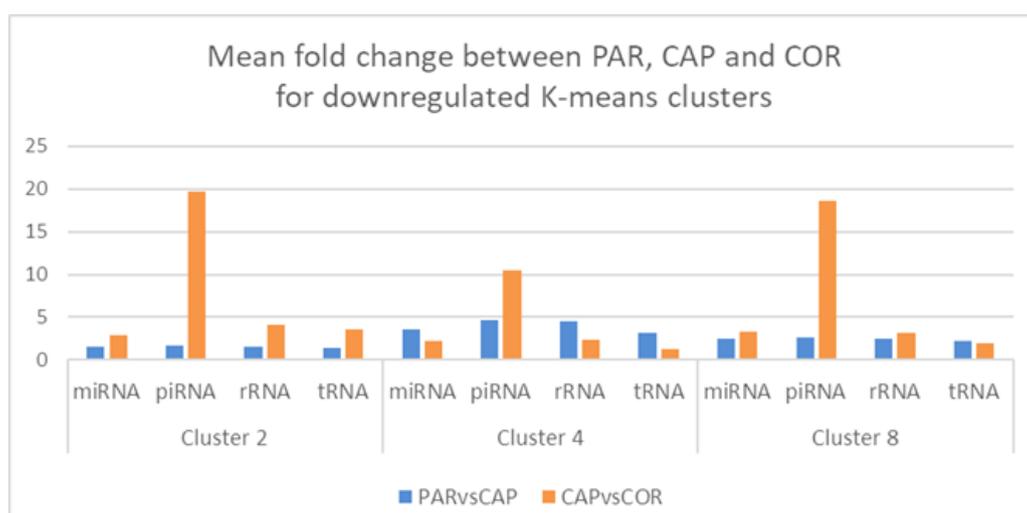


Fig 8: Expression fold changes observed between PAR and CAP, as well as CAP and COR for sncRNA belonging to downregulated K-means clusters. Total expression was computed at each maturation

stage (PAR, CAP and COR) for each sncRNA class and each cluster. PAR versus CAP and CAP versus COR fold changes we computed and plotted as bargraphs.

Overall, K-means tend to cluster the expression of specific sncRNA classes (Figures 9a and 9b). Indeed, clusters of downregulated sncRNAs (cluster 4 from PAR to CAP and cluster 2 & 8 from PAR to COR) were found to gather mostly piRNA (86% of all piRNAs) and rsRNA (33% of all rsRNAs). Overall, miRNA are enriched in clusters peaking at CAU (cluster 11, 27% of miRNAs) and SPZ (cluster 3, 32% of miRNAs), while tsRNA are enriched in cluster 11 (26% of tsRNAs). In contrast, rsRNAs show a similar distribution among clusters, except for clusters 1,7 & 13 (underrepresented) and 10 (enriched), which all peak at COR.

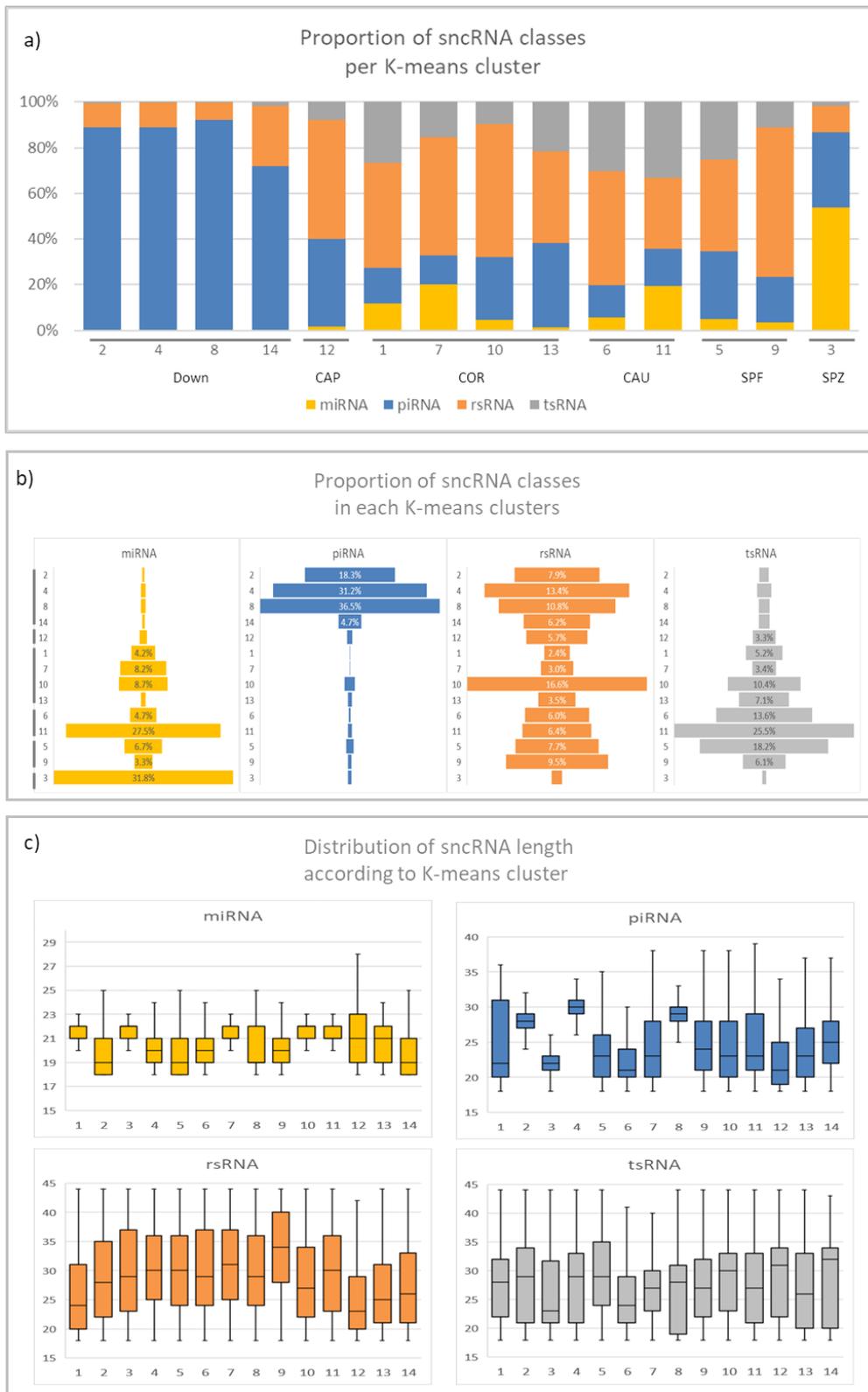


Fig 179: General characteristics of K-means clusters. a) The proportion of sncRNAs classes per K-means cluster and b) the proportion of each sncRNA class per K-means cluster highlighted enrichment of some cluster in particular sncRNA classes. Downregulated clusters (i.e. 2, 4, 8 and 14) gather mainly piRNAs and rsRNAs. Overall, tsRNAs are enriched in clusters peaking at CAU and miRNA at CAU and SPZ, while rsRNA account for a large proportion of clusters peaking at CAU and COR, as well as SPF. c) The

distribution of reads length in each K-means cluster revealed variations in miRNA and piRNAs length among clusters. Indeed, miRNA appeared to be larger in size within cluster 1, 7, 10 (COR), cluster 11 (CAU) and cluster 3 (SPZ). Likewise, piRNAs appeared to be longer in downregulated clusters 2, 4 and 8.

Interestingly, while sncRNA length distribution of each sncRNA class was consistent with the expected size for each sncRNA class, specific patterns were observed for some clusters. For instance, miRNA appeared to be longer within cluster 1, 7, 10 (COR), cluster 11 (CAU) and cluster 3 (SPZ). Greater length heterogeneity was also observed according to clusters, indicative of variable intensity of RNA edition mechanisms. However, the most striking variation was observed for piRNAs, whose length appeared to be 30nt in downregulated clusters 2, 4 & 8 and 25-30nt for all other clusters, except for cluster 3 (20-25nt).

In addition, some K-means tend to cluster the expression of specific sncRNA sub-groups (Figure 10 and supplementary Table S5-S6). For instance (Figure 10a), Glycine isoacceptors account for 57%, 72%, and 62% of expression within clusters 1, 7 and 12, respectively, but Glycine isoacceptors are mostly expressed within clusters 5 and 11 (29% and 32% respectively). Likewise, Isoleucine isoacceptors account for 72% of total expression within cluster 3 but this cluster account for only 8% of Isoleucine isoacceptors whole expression. Glutamine isoacceptors account for 34% and 58% of expression within cluster 6 and 9 and conversely, these two clusters account altogether for 86% of Glutamine isoacceptors expression.

Surprisingly, cluster 5 (peaking at SPF) accounts for more than one third of 9 isoacceptors expression (Asp, Cys, Glu, Leu, Phe, Sel, Trp, Tyr and Val), ranging from 30% to 80% according to isoacceptors. Specific patterns were likewise observed for tRFs (Figure 10b). For instance, tRF5s account for a large proportion of expression within clusters 5, 7 and 9 (43%, 60% and 74%, respectively), while clusters 12 and 14 are associated with 5'-tRHs (86% and 81% of expression, respectively). i-tRFs are mostly expressed in cluster 11 (46%), where they account for 85% of expression. Likewise, they account for 61% and 91% of expression within cluster 1 and 3, respectively. As for isoacceptors, cluster 5 account for a large proportion of tRFs expression: 31% of tRF5s, 46% of tRF3s and 54% of 3'-tRHs. Regarding rRNAs, 16S account for

most of the expression within cluster 6, 7, 9 and 11 (59%, 60%, 72% and 56%, respectively), while 28S account for about 73%, 64%, 72%, 78% and 60% of expression within cluster 1, 2, 4, 8 and 14 (Figure 10c). cluster 10 account for 31%, 20%, 29% and 86% of 12S, 16S, 28S and 5.8S expressions, respectively and cluster 5 for 67% of 5S and 29% of 18S expressions. Likewise, K-means clusters showed a diversity of piRNA length and nucleotide composition, suggesting changes in the piRNA biogenesis along male reproductive tract (Figure 10d). Indeed, downregulated clusters 2, 4, 8 were enriched (>80%) in U1 piRNAs bearing an uracil residue at their 5' end. In contrast, clusters with increased expression in epididymis corpus and cauda, as well as ejaculated sperm were dominated by non U1 piRNAs. No enrichment for an A residue was observed at position 10 and U1 piRNA were more frequent than A10 piRNAs whatever the cluster, except for clusters 1, 7 and 10 (clusters peaking at epididymis corpus) where the proportion of A10 piRNAs reached up to 40%. When considering expression levels (Figure 10e), contribution of U1 piRNA was still overriding in most clusters, except for clusters peaking at epididymis corpus. Noteworthy, less than 40% of cluster expression could be attributed to U1 or A10 piRNAs for some clusters peaking at epididymis corpus (1, 13), cauda (6, 11) and ejaculated sperm (5, 9). Furthermore, a different picture emerged for cluster 1, the expression ratio between A10/U1 being higher than the number ratio. Likewise, expression attributable to A10 piRNAs dramatically increased starting from CAP for cluster 10. These changes illustrate thus that even if they share a similar expression trend, U1 and A10 piRNAs belonging to a same cluster may be subjected to differential regulation.

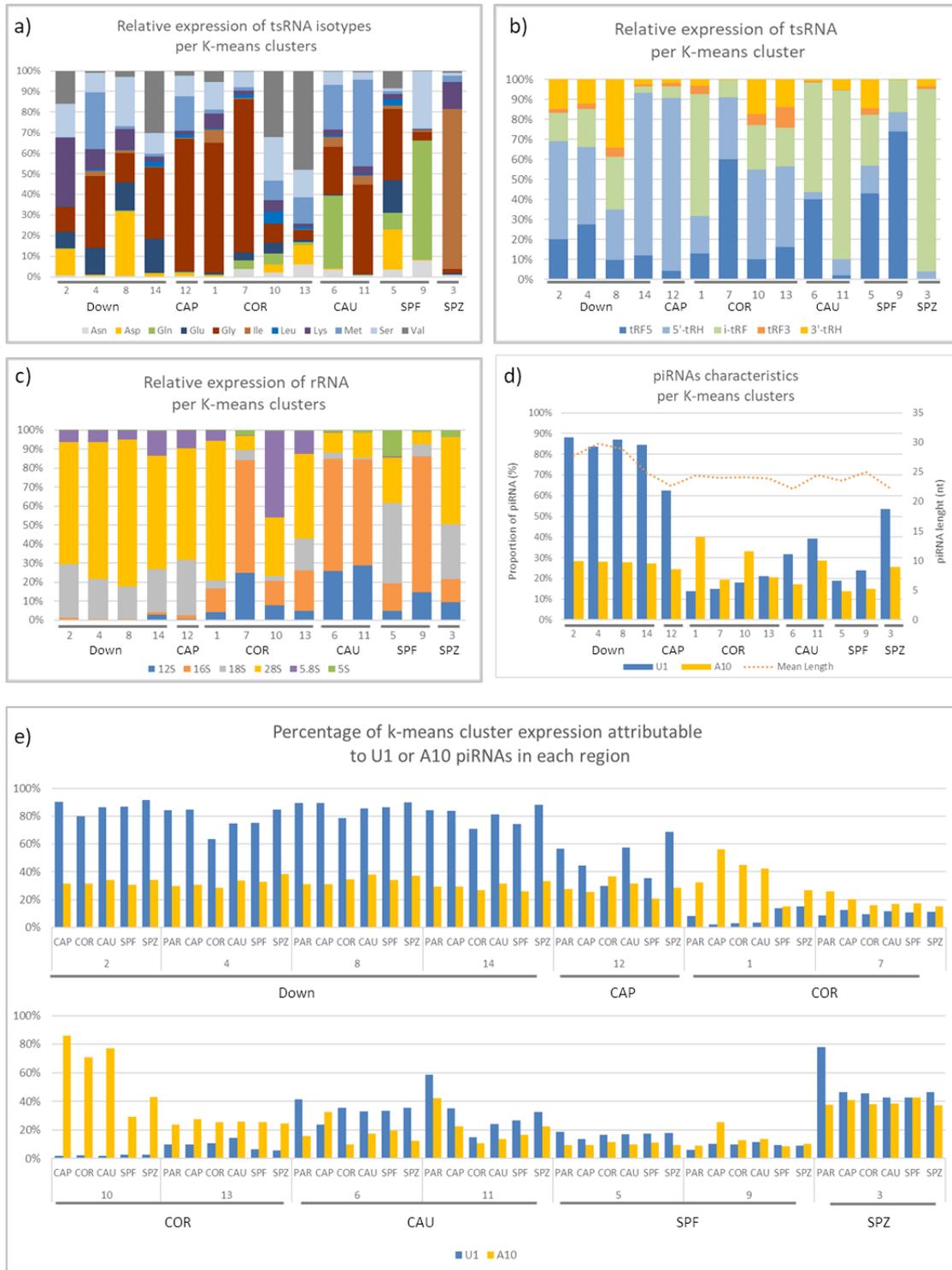


Fig 10: sncRNA expression across K-means clusters. The relative expression of sncRNA subgroups was computed for each cluster, showing differences in proportion of sncRNA sub-groups across K-means clusters. **a)** Glycine isoacceptors account for a large proportion of expression within clusters 1, 7 and 12 (57%, 72% and 62%, respectively), while isoleucine isoacceptors account for 72% of expression within cluster 3 **b)** Regarding tsRNAs, tRF5s account for a large proportion of expression within clusters 5, 7 and 9 (43%, 60% and 74%, respectively), while clusters 12 and 14 are associated with 5'-tRHs (86 and 81% of expression, respectively). Likewise, i-tRFs account for 61%, 91% and 85% expression within clusters 1, 3 and 11, respectively. **c)** Specific patterns were also observed for rRNAs. For instance, 16S

account for most of the expression within cluster 6, 7, 9 and 11 (59%, 60%, 72% and 56%, respectively), while 28S account for about 73%, 64%, 72%, 78% and 60% of expression within cluster 1, 2, 4, 8 and 14. **d)** piRNA characteristics across K-means clusters. About 80-90% of piRNA were shown to bear a U residue at their 5' end (U1 piRNAs) within downregulated clusters 2,4,8 and 14. In contrast, clusters with increased expression in COR and CAU, as well as SPF were dominated by non U1 piRNAs, while cluster 3 (SPZ) gathers 53% of 1U-piRNA. No gross enrichment for an A residue was observed at position 10, but piRNA length variations were observed among clusters. **e)** When considering expression levels, the proportion of expression due to A10 piRNA dramatically increased for clusters 1, 10 and 13.

Furthermore, a search for biological processes and molecular pathways associated with miRNA targets was performed for each K-means cluster, highlighting several enriched GO terms (Figure 11 and supplementary Tables S7-S8).

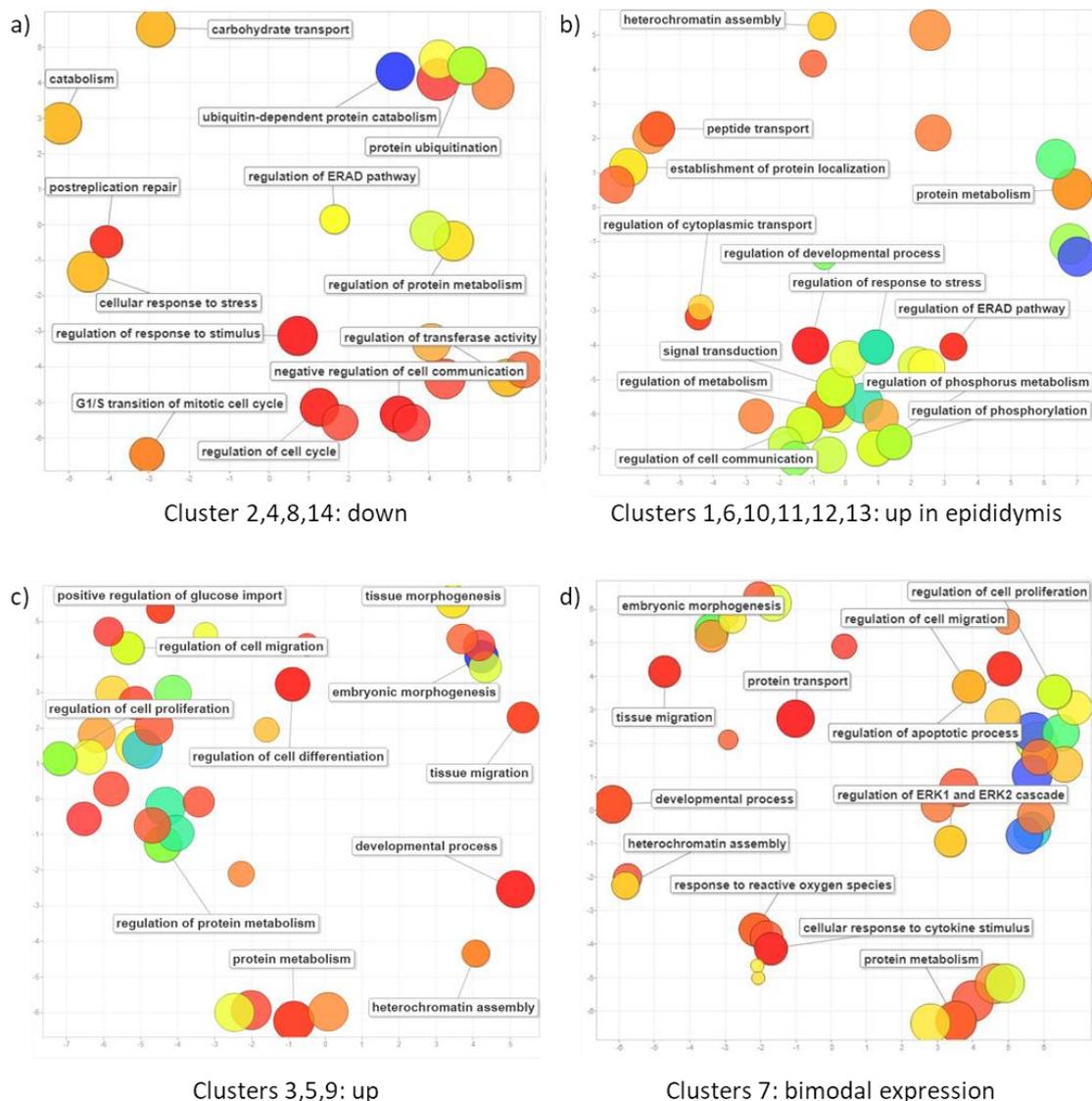


Fig 11: GO biological processes associated with K-means clusters. For each cluster, miRNA targets predicted by Targetscan were retrieved to explore biological processes and pathways. Gorilla was used to identify enriched GO terms and scatterplots were produced using Revigo, showing enriched terms grouped by semantic similarity. Relevant GO terms are shown for miRNA targets of **a)** downregulated clusters 2, 4, 8 and 14; **b)** clusters 1, 6, 10, 11, 12 and 13 whose expression peaks at epididymis; **c)** clusters 3, 5 and 9, whose expression increases from testis to ejaculated sperm; **d)** cluster 7 having a bimodal expression curve, peaking both at epididymis and ejaculated sperm. The size of each dot is proportional to the number of genes related to the given term (log size) and the color depicts the associated p-value (from red  $\leq 10E-3$  to blue  $\leq 1E-6$ ).

For instance, downregulated clusters 2, 4, 8 and 14 were shown to gather miRNAs targeting genes involved in cell cycle and its regulation, (G1/S transition, post-replication repair) as well as organelle or compounds degradation (ubiquitination and ERAD pathways) and response to stress. Functional annotation of genes targeted by clusters 1, 6, 10, 11, 12 and 13 whose expression peaks at epididymis, showed enrichment in protein metabolism related terms, as well as cytoplasmic transport (ER to Golgi vesicle-mediated transport and intracellular protein transport), response to stress (response to endoplasmic reticulum stress, regulation of intrinsic apoptotic signaling pathway), regulation of phosphorylation, cell differentiation and heterochromatin assembly. Genes targeted by clusters 3, 5, 9 whose expression increases in ejaculated sperm, were found to be involved in cell proliferation and migration, embryonic and tissues morphogenesis, regulation of transcription (regulation of transcription by RNA polymerase II, regulation of core promoter binding), heterochromatin assembly, as well as regulation of glucose import. Genes targeted by cluster 7, which exhibit a bimodal expression curve peaking at both epididymis and ejaculated sperm, showed functional annotations combining terms also enriched in clusters 3, 5, 9 and clusters 1, 6, 10, 11, 12, 13.

## Discussion

Sperm RNAs have long been considered as spermatogenic leftovers with no further function [36, 37], but a more dynamic picture has emerged recently and the sperm sncRNA profile has been shown to be dynamically modified as the cells migrate through the epididymis [20], possibly through direct transfer of sncRNA from epididymosomes to spermatozoa [13]. Our study was thus designed to study the dynamics of cattle sperm sncRNAs from spermatogenesis to final maturation.

This study was challenging technically as well as in terms of sample collection. However, enough spermatozoa could be collected, without somatic cells contamination, and good quality RNAs could be prepared and sequenced using our previously described protocol. As a result, comprehensive repertoires of cattle sperm sncRNA could be established at six maturation stages along the male tract. One caveat of our study is related to the fact that the subspace available for a given group of sncRNAs within a given sample is limited by the quantity of other sncRNAs being sequenced. To limit associated bias, between sample normalization has been performed, making it possible to compare relative expression between samples. However, this strategy prevents any conclusion on absolute quantification. For instance, it is not clear to what extent the increase in miRNA expression in epididymis results from increased expression of this sncRNA class or just decrease in piRNA expression. Anyway, our work provides a detailed picture of sperm sncRNA spatiotemporal dynamics.

The primary cattle sncRNA sperm content is dramatically remodelled as sperm mature along the epididymis.

The sperm sncRNAome showed a great plasticity level and huge differences in terms of sperm sncRNA content were observed, making it possible to discriminate between sperm samples collected at various maturation stage along the male tract.

One of the most striking findings of our study is the observed enrichment of piRNAs within testis parenchyma relative to the sperm fraction isolated from more distal (corpus and cauda)

regions of the epididymis as well as ejaculated sperm. Conversely, an increase in miRNAs and tRNAs content was observed from testis parenchyma to ejaculated sperm, while rRNAs showed a particular expression profile, peaking at epididymis corpus. Overall, our data are consistent with previous studies in mouse and recapitulate prior findings, both in terms of sncRNA content and global expression trends [12, 19, 38-42]. However, data unavailability as well as species and/or methodological differences precludes detailed comparison with these previous findings. For instance, Sharma et al [38] identified tRFs as the primary small RNA population in epididymal sperm but reads mapping to rRNAs were discarded before normalization. Here, rRNA were kept in the analysis and were identified as the major sncRNA population in epididymis corpus. Yet, we also report an increase in tRFs expression in epididymis. In addition, among 1.6 million sncRNA sequences described by Chu et al, only 56316 perfectly match our bovine sequences (71% rRFs, 15% piRNA, 9% tRFs and 5% miRNAs). The remaining sequences may reflect both species specificity and the plethora of isoforms produced through RNA editing mechanisms. In this respect, concordance between mouse and bovine increases when considering canonical miRNAs instead of isomiRs represented by individual reads. For instance, about 45% of bovine sperm miRNAs were also described in mouse, including several key miRNAs involved in cell differentiation, proliferation, spermatogenesis, or embryo development (miR-10, miR-29, miR-34, miR-100, miR-148, miR-191...).

Beyond the bulk differences in abundance of broad classes of small RNAs, K-means clustering was performed to cluster sncRNAs based on their spatiotemporal expression profiles, highlighting specific patterns of expression for individual sncRNAs and confirming that the sperm sncRNAome does not simply reflect a legacy of spermatogenesis. Indeed, four different expression profiles were observed for downregulated sncRNAs, differing by the timing and/or intensity of expression decrease. Likewise, four profiles were associated with increased expression at epididymis corpus, two at epididymis cauda, three at ejaculated sperm... Altogether, our data suggest at least three major waves of drastic changes in sperm sncRNA content, first between testis and epididymis corpus, then at epididymis cauda and later in ejaculated sperm. Indeed, the transition from testis parenchyma to epididymis corpus is characterized by a shift from piRNAs to other sncRNA classes, following a drastic decrease in expression of sncRNA belonging to clusters 2, 4, 8 & 14, together with a slight increase in

miRNA and tsRNA expression (clusters 1, 7, 10, 12, 13) and a transient burst in rsRNA expression (cluster 7, 10). The second wave of change between epididymis corpus and cauda is characterized by a decrease in rsRNA expression associated with an increase in miRNA and tsRNA expression. The third wave in ejaculated sperm is characterized by an increase in miRNA and a decrease in tsRNA expression. Interestingly, a higher proportion of piRNAs and miRNAs was observed in total ejaculated sperm (18% and 38% for piRNAs and miRNAs, respectively), compared to ejaculated sperm depleted from seminal plasma (15% for both piRNAs and miRNAs), suggesting that seminal plasma may contain piRNAs and miRNAs and may be involved in this third wave.

Encapsulation within the cytoplasmic droplet and ejection upon epididymal entry has been proposed as a likely mechanism to explain clearance of testicular piRNAs [38]. Since the bovine cytoplasmic is lost between epididymis caput and corpus, expression of piRNA belonging to clusters 2, 12 & 14 is in good agreement with such a mechanism. However, these piRNA account only for 24.2% of differentially expressed piRNA and a greater drop in expression was observed between parenchyma and epididymis caput compared to caput and corpus. Thus, other mechanisms may also be at work to account for these piRNAs belonging to clusters 4 & 8 and accounting for 67% of piRNAs. In addition, higher expression fold changes between caput and corpus were observed for piRNAs compared to other sncRNA classes, suggesting that cytoplasmic droplet ejection alone cannot account for the observed changes.

It has been proposed that epididymosomes may serve as cargo to supplement maturing spermatozoa with sncRNA, especially miRNA and tRFs [13, 40], during transit through the epididymis. The increased expression observed for miRNA and tRFs along epididymis is in good agreement with this hypothesis, given that several cattle sperm miRNAs were already identified in cattle epididymosomes [43]. Indeed, despite technical differences (miRNA array vs NGS), 33 and 5 miRNAs among 204 epididymal sperm miRNAs were identified as shared with miRNA extracted from caput and cauda epididymosomes, respectively (Figure 12). Interestingly, among the 33 shared miRNAs from caput epididymosomes, 76% are overexpressed in cauda sperm, suggesting a delay for epididymosomes to deliver their content within sperm transiting in epididymis lumen.

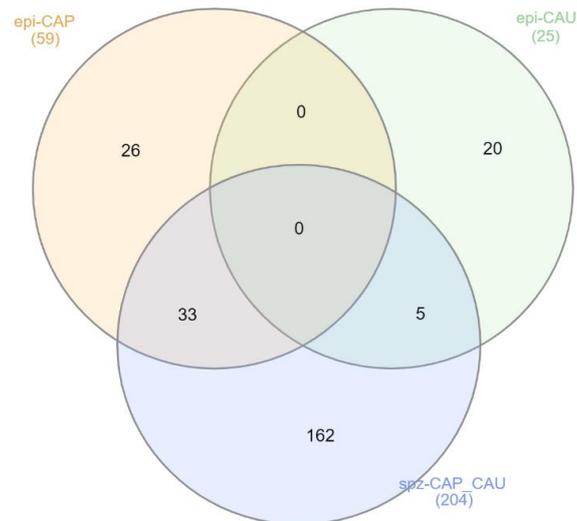


Fig 12: Differentially expressed miRNAs are shared between bovine epididymosomes and epididymal sperm. Comparison between published miRNAs expressed in caput (epi-CAP) or cauda (epi-CAU) and 204 miRNAs differentially expressed in caput or cauda sperm (spz-CAP\_CAU) identified 33 and 5 common miRNAs, respectively.

Likewise, a few percent of piRNA escape the general decreasing trend, as illustrated by cluster 12 peaking at epididymis caput or clusters 1, 7, 10 and 13 peaking at epididymis corpus. Interestingly, piRNA belonging to these cluster tend to be shorter (28nt on average vs 31nt for parenchyma piRNAs). Since somatic piRNAs and piRNA-like have been described to be shorter [44], it could be tempting to speculate that these piRNAs also originated from interactions with epididymosomes. However, their nucleotide composition differs from parenchyma pachytene piRNAs (mostly non U1 piRNAs, compared to 80% of U1-piRNAs in parenchyma), while somatic piRNAs are thought to be similar to pachytene piRNAs, although having a distinct signature associated with a yet to be identified biogenesis pathway [45]. In addition, epididymosomes seem to harbour only small amounts of piRNA [46] and co-incubation of epididymosomes and spermatozoa *in vitro* does not result in substantial piRNA transfer [12]. Thus, these piRNAs do not features typical characteristics of both somatic and pachytene piRNA and their origin remains unclear. They may only reflect false positives present in primary databases [47], may feature isopiRs produced by 5' edition of piRNAs generated by the primary or secondary production pathway, or may be produced *in situ* by sperm following a particular unknown biogenesis pathway [46]. Given that epididymal sperm nuclear gene

expression is completely repressed, this later hypothesis seems unlikely. However, piRNA may be produced by the mitochondrial genome [48] and may be derived from rsRNA or tsRNA [49]. Thus, it could be assumed that in situ production may rely on PIWI processing of rsRNA or tsRNA, eventually expressed by mitochondria, and/or mitochondrial expression of piRNAs. Further studies, including more detailed bioinformatics analysis, as well as NGS sequencing of bovine caput, corpus and cauda epididymosomes will be required to explore further this hypothesis.

Noteworthy, although some authors support that PIWI proteins and piRNAs are depleted in late spermatogenesis stages [50], our findings agree with others reporting that piRNAs exist in ejaculated spermatozoa and may play important roles during maturation, fertilization or embryonic genome activation [51, 52].

A search for sncRNA molecular function, including *in-silico* search for miRNA targets and GO enrichment analysis, was performed for each K-means cluster to provide insights into their putative biological role at each maturation stage.

#### Sperm sncRNA, a vestige of spermatogenesis

Downregulated clusters 2, 4, 8 & 14 show a progressive decrease in expression and may thus gather sncRNAs remnant from spermatogenesis. These clusters are enriched in pachytene piRNA, which are known to be crucial for transposon silencing [53] and for the global mRNA decay in spermatocytes and spermatids as well as the massive turnover of mRNAs in late spermiogenesis [54, 55]. Enriched GO terms associated with predicted miRNA targets highlight relevant biological processes for spermatogenesis, such as cell proliferation (regulation of cell cycle and G1/S transition), protein ubiquitination and response to stress. For instance, spermatogenesis requires extensive protein synthesis, folding, maturation trafficking to appropriate cellular location to differentiate spermatogonia to spermatozoa during mitosis and meiosis within the testes, as well as destroying unfolded/misfolded proteins. Thus, ER stress and UPR signalling cascades play a key role in spermatogenesis, and both extrinsic factors, such as toxicants, and intrinsic factors, such as hyperthermia or oxidative stress, have been shown to trigger ER stress and impair spermatogenesis [56, 57]. Likewise, the ubiquitin–

proteasome pathway plays a key role at several stages of spermatogenesis (meiosis, acrosome biogenesis, and spermatozoa maturation) to remove many proteins and organelles and facilitate the formation of condensed sperm. Here, UBE2D3, UBE2I and UBE2J1 were identified as targets of cluster 4 miRNAs. Knock-out mice lacking ubiquitylation enzymes such as UBE2J1 suffer from male sterility associated with flagella and acrosomal defects [58]. Interestingly, piRNAs have been shown to induce the ubiquitination and degradation of MIWI through the APC/C proteasome pathway [50], a process essential for maturation from late spermatid to sperm.

#### Epididymal sncRNA modifications: a role for sperm maturation?

Since spermatozoa are considered as transcriptionally and translationally silent, a role in sperm maturation seems to be unlikely. However, two studies have proven that contrary to the accepted dogma, mRNA are translated by mitochondrial-type ribosomes during capacitation and new proteins are thus synthesized as part of the capacitation process or to replace degraded proteins [59, 60]. Moreover, active transcription and translation have been observed in sperm mitochondria and proved to be essential for mitochondrial activity and regulation of high-speed linear motility through ATP level in low glucose condition [61]. In this respect, assuming that such mechanisms are also operational during maturation, sperm sncRNA may have a role in regulating energy metabolism, in line with the acquisition of progressive motility by spermatozoa during epididymal transit. For instance, mTOR, PI3K-AKT and AMP KEGG pathway were found to be targeted by K-means clusters 1, 6, 10, 11, 12 and 13, whose expression trends peak at epididymis. In particular, mTOR pathway has been associated with regulation of mitochondrial function (glycolysis activation, lipid synthesis, autophagy) [62] and human sperm motility [63, 64]. Tyrosine phosphorylation has also been linked to energy production by mitochondria [65]. Interestingly, GO terms related to phosphorylation were also associated with clusters 1, 6, 10, 11, 12 and 13. In addition, phosphorylation processes are also crucial since motility is dependent on intracellular cAMP and cascades of phosphorylations and dephosphorylations of dynein arms.

Spermatozoa are highly susceptible to oxidative stress due to their membrane containing high concentrations of polyunsaturated fatty acids and their lack of significant antioxidant

protection [66]. In this respect, it is worth mentioning that tsRNAs have been proposed to be involved in the oxidative stress response [68]. Furthermore, enriched GO terms include Regulation of response to stress.

#### Enrichment in specific sncRNAs of ejaculated sperm: a role in embryo development?

Sperm RNAs can be delivered to the oocyte during fertilization and remain stable until the onset of embryonic genome activation [67]. Treatment of mature sperm to remove sperm-carried RNAs results in a decrease of blastocyst formation rate [68] and some studies suggest that sperm RNA may act as an additional source of paternal hereditary information, which may be essential for fertilization and/or influence early embryo development [69] and the long-term phenotype of the offspring [70, 71]. Recently, sperm sncRNAs gained at epididymis cauda were shown to be crucial for embryonic development after the blastocyst stage in Mice [23]. Interestingly, Clusters 3, 5 & 9 which gather sncRNA peaking in ejaculated sperm, were found to be associated with GO terms related to embryo development such as developmental process, embryonic morphogenesis as well as cell proliferation, differentiation, and migration. In particular, several target genes are involved in the HIPPO signalling pathways that control embryonic cell fate during establishment of pluripotency. Likewise target genes HOXB1, HAND1 or the SMAD family have been associated with self-renewal of pluripotent stem cells and differentiation toward ectoderm, mesoderm, or endoderm lineages [72, 73]. Regulation and maintenance of pluripotency relies also on regulation of chromatin organization [74]. Interestingly, Heterochromatin Assembly belong to enriched GO terms and bta-miR-100, one highly expressed miRNA in ejaculated sperm, has been proposed as one of the main factors associated with the initiation of pluripotency [75], acting through repression of SMARCA5, a protein involved in chromatin remodelling [76]. Consistent with this hypothesis, embryos produced using caput sperm, which lack miR-100, overexpress SMARCA5 at 4-cell stage and fail to develop after blastocyst stage [23].

In addition, among miRNA belonging to clusters peaking in ejaculated sperm, some have already been described as involved in embryo development. For instance, miRNA-34c, which belongs to the most expressed miRNAs in cluster 3, is known to enhance the germinal phenotype of cells already committed to this lineage during spermatogenesis [6], but is also a

key player in murine early embryo development [77-79]. Likewise, bta-miR-191, the second most expressed miRNA in our dataset has been associated with fertilisation rate and embryo quality in humans [80].

Beside miRNAs, rsRNA, tsRNAs and piRNAs are also gained post epididymis, either by cauda epididymosomes or prostasomes, which may be involved in trans-generational epigenetic inheritance and impact embryo development. Altogether, increasing evidence suggest that sperm RNA-encoded information is decoded in early embryos to control offspring phenotypes [71], as illustrated by altered metabolic pathways in early embryo and induced metabolic disorders in F1 offspring produced by injection into zygotes of sperm tsRNA (mainly tRF5s) from males subjected to a high-fat diet [81]. Noteworthy, tRF5s are the most enriched tsRNA subpopulation in clusters 5 and 9. Likewise, tRNA-Gln-TTG, which are enriched exclusively in cluster 9, have been shown to participate in the early cleavage of porcine preimplantation embryos [82] and injection of tRNA-Gly-GCC fragments (one sperm sncRNAs upregulated by a low-protein diet) into zygotes results in repression of MERVL-regulated transcripts in two-cell embryos and slowdown of embryo development [40].

Altogether, these findings suggest both distinct role of sncRNA at each maturation stage and coordinated actions of several sncRNA classes at a given maturation stage. In a context of transcriptional and translational inactivity, several hypothetical mechanisms can be proposed to explain sncRNA effects, including regulation of ribosome biogenesis to adapt the mitochondrial translational machinery, use of alternative pathways to produce in situ sncRNA (e.g mt-piRNAs and recycling rsRNAs/tsRNAs to produce piRNAs), followed by chain reactions mediated by the interplay with DNA methylation, histone marks and transposon elements as well as induced metabolic changes [71].

Assuming that sncRNA expressed at a given maturation stage may be of importance for the transition towards the next stage, and sncRNA enriched in ejaculated sperm may have a role in capacitation, fertilization or embryo development, K-means clustering appears as a valuable tool to group sncRNA according to their expression profile and putative functions. This may help filtering sncRNA for biomarkers studies (e.g. relevant sncRNA associated with semen fertility), as well as identifying sncRNAs of interest and designing relevant experiments for future functional studies, with the aim to validate their biological role at each maturation stage.

## Conclusions

Mature sperm carry thousands of sncRNAs, including miRNAs, piRNAs, rRNAs and tsRNAs, whose function remains unclear, although growing evidence suggest that they play a role in sperm maturation and are involved in paternal epigenetic inheritance. Here, we confirm that cattle sperm sncRNA are far from being just remnants of the spermatogenic cycle, in good agreement with previous studies in mouse. Rather, we show that the sncRNA profile is dynamically modified as the sperm transit through the epididymis, due to the combination of multiple factors such as loss of the cytoplasmic droplet, intracellular degradation, and interaction with epididymosomes. In addition, we identified piRNAs whose production is unlikely to be attributed to epididymosomes, suggesting *in situ* production and/or modification of sncRNAs by sperm, a thrilling finding given that epididymal sperm nuclear gene expression is supposed to be completely repressed. Further studies will be required to decipher the underlying mechanisms. In addition, sncRNA expression profiles were used to provide insight into the putative role of sperm sncRNAs. Our preliminary *in silico* analysis suggests a role during sperm maturation for epididymal sncRNA and embryo development for sncRNA expressed in ejaculated sperm. This finding warrants further investigations, that may benefit from these results to identify sncRNA of interest and design relevant experiments.

## Declarations

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

### **Availability of data and materials**

The (trimmed) raw data have been stored at ENA, European Nucleotide Archive, under the primary accession number: PRJEB41989.

### **Competing interests**

The authors have declared that no competing interests exist.

### **Authors' contributions**

Eli Sellem was the main investigator, taking part to Conceptualization, Data curation, Formal analysis, Visualization, Writing – original draft and Writing – review & editing

Sylvain Marthey was involved in Data curation, Formal analysis –bioinformatics and Writing – review & editing

Andrea Rau and Luc Jouneau were involved in Data curation, Formal analysis – biostatistics and Writing – review & editing

Aurelie Bonnet and Chrystelle Le Danvic took part to Investigations and Writing – review & editing

Hélène Kiefer was involved in Methodology, Investigations, Visualization, Writing – review & editing.

Hélène Jammes was involved in Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

Laurent Schibler was involved in Conceptualization, Methodology, Funding acquisition, Supervision, Project administration, Formal analysis, Validation, Writing – original draft, Writing – review & editing.

## Bibliography

1. Chalmel F, Rolland AD. Linking transcriptomics and proteomics in spermatogenesis. *Reproduction*. 2015;150(5):R149. doi: 10.1530/rep-15-0073.
2. de Mateo S, Sassone-Corsi P. Regulation of spermatogenesis by small non-coding RNAs: role of the germ granule. *Seminars in cell & developmental biology*. 2014;29:84-92. Epub 04/19. doi: 10.1016/j.semcd.2014.04.021. PubMed PMID: 24755166.
3. Romero Y, Meikar O, Papaioannou MD, Conne B, Grey C, Weier M, et al. Dicer1 depletion in male germ cells leads to infertility due to cumulative meiotic and spermiogenic defects. *PloS one*. 2011;6(10):e25241. Epub 2011/10/15. doi: 10.1371/journal.pone.0025241. PubMed PMID: 21998645; PubMed Central PMCID: PMC3187767.
4. Huang YL, Huang GY, Lv J, Pan LN, Luo X, Shen J. miR-100 promotes the proliferation of spermatogonial stem cells via regulating Stat3. *Molecular reproduction and development*. 2017;84(8):693-701. Epub 2017/06/02. doi: 10.1002/mrd.22843. PubMed PMID: 28569396.
5. Hilz S, Fogarty EA, Modzelewski AJ, Cohen PE, Grimson A. Transcriptome profiling of the developing male germ line identifies the miR-29 family as a global regulator during meiosis. *RNA Biol*. 2017;14(2):219-35. Epub 2016/12/17. doi: 10.1080/15476286.2016.1270002. PubMed PMID: 27981880; PubMed Central PMCID: PMC5324742.
6. Bouhallier F, Allioli N, Laval F, Chalmel F, Perrard MH, Durand P, et al. Role of miR-34c microRNA in the late steps of spermatogenesis. *Rna*. 2010;16(4):720-31. doi: 10.1261/rna.1963810. PubMed PMID: 20150330; PubMed Central PMCID: PMC3187767.
7. Weick EM, Miska EA. piRNAs: from biogenesis to function. *Development*. 2014;141(18):3458-71. doi: 10.1242/dev.094037. PubMed PMID: 25183868.
8. Zhou W, De Iulii GN, Dun MD, Nixon B. Characteristics of the Epididymal Luminal Environment Responsible for Sperm Maturation and Storage. *Frontiers in endocrinology*. 2018;9:59. doi: 10.3389/fendo.2018.00059. PubMed PMID: 29541061; PubMed Central PMCID: PMC5835514.
9. Nixon B, De Iulii GN, Hart HM, Zhou W, Mathe A, Bernstein IR, et al. Proteomic Profiling of Mouse Epididymosomes Reveals their Contributions to Post-testicular Sperm Maturation. *Molecular & Cellular Proteomics*. 2019;18(Supplement 1):S91-S108. doi: 10.1074/mcp.RA118.000946.
10. Sullivan R, Frenette G, Girouard J. Epididymosomes are involved in the acquisition of new sperm proteins during epididymal transit. *Asian journal of andrology*. 2007;9(4):483-91. Epub 2007/06/26. doi: 10.1111/j.1745-7262.2007.00281.x. PubMed PMID: 17589785.
11. Girouard J, Frenette G, Sullivan R. Comparative proteome and lipid profiles of bovine epididymosomes collected in the intraluminal compartment of the caput and cauda epididymidis. *International journal of andrology*. 2011;34(5 Pt 2):e475-86. Epub 2011/08/31. doi: 10.1111/j.1365-2605.2011.01203.x. PubMed PMID: 21875428.
12. Sharma U, Sun F, Conine CC, Reichholf B, Kukreja S, Herzog VA, et al. Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm. *Developmental cell*. 2018;46(4):481-94.e6. Epub 2018/07/31. doi: 10.1016/j.devcel.2018.06.023. PubMed PMID: 30057273; PubMed Central PMCID: PMC6103849.
13. Reilly JN, McLaughlin EA, Stanger SJ, Anderson AL, Hutcheon K, Church K, et al. Characterisation of mouse epididymosomes reveals a complex profile of microRNAs and a potential mechanism for modification of the sperm epigenome. *Scientific reports*. 2016;6:31794. doi: 10.1038/srep31794. PubMed PMID: 27549865; PubMed Central PMCID: PMC4994100.
14. Dacheux JL, Belleanne C, Guyonnet B, Labas V, Teixeira-Gomes AP, Ecroyd H, et al. The contribution of proteomics to understanding epididymal maturation of mammalian spermatozoa. *Systems biology in reproductive medicine*. 2012;58(4):197-210. doi: 10.3109/19396368.2012.663233. PubMed PMID: 22788532.
15. Skerget S, Rosenow MA, Petritis K, Karr TL. Sperm Proteome Maturation in the Mouse Epididymis. *PloS one*. 2015;10(11):e0140650. doi: 10.1371/journal.pone.0140650. PubMed PMID: 26556802; PubMed Central PMCID: PMC4640836.

16. Rejraji H, Sion B, Prensier G, Carreras M, Motta C, Frenoux JM, et al. Lipid remodeling of murine epididymosomes and spermatozoa during epididymal maturation. *Biology of reproduction*. 2006;74(6):1104-13. doi: 10.1095/biolreprod.105.049304. PubMed PMID: 16510839.
17. Belleannée C, Thimon V, Sullivan R. Region-specific gene expression in the epididymis. *Cell and tissue research*. 2012;349(3):717-31. Epub 2012/03/20. doi: 10.1007/s00441-012-1381-0. PubMed PMID: 22427067.
18. Nixon B, De Luliis GN, Dun MD, Zhou W, Trigg NA, Eamens AL. Profiling of epididymal small non-protein-coding RNAs. *Andrology*. 2019;7(5):669-80. Epub 2019/04/26. doi: 10.1111/andr.12640. PubMed PMID: 31020794.
19. Chu C, Zhang YL, Yu L, Sharma S, Fei ZL, Drevet JR. Epididymal small non-coding RNA studies: progress over the past decade. *Andrology*. 2019;7(5):681-9. Epub 2019/05/03. doi: 10.1111/andr.12639. PubMed PMID: 31044548.
20. Nixon B, Stanger SJ, Mihalas BP, Reilly JN, Anderson AL, Tyagi S, et al. The MicroRNA Signature of Mouse Spermatozoa Is Substantially Modified During Epididymal Maturation. *Biology of reproduction*. 2015;93(4):Article 91, 1-20. doi: 10.1095/biolreprod.115.132209.
21. Belleannée C, Calvo É, Caballero J, Sullivan R. Epididymosomes Convey Different Repertoires of MicroRNAs Throughout the Bovine Epididymis<sup>1</sup>. *Biology of reproduction*. 2013;89(2). doi: 10.1095/biolreprod.113.110486.
22. Sullivan R, Saez F. Epididymosomes, prostasomes, and liposomes: their roles in mammalian male reproductive physiology. *Reproduction*. 2013;146(1):R21-35. doi: 10.1530/REP-13-0058. PubMed PMID: 23613619.
23. Conine CC, Sun F, Song L, Rivera-Perez JA, Rando OJ. Small RNAs Gained during Epididymal Transit of Sperm Are Essential for Embryonic Development in Mice. *Developmental cell*. 2018;46(4):470-80 e3. Epub 2018/07/31. doi: 10.1016/j.devcel.2018.06.024. PubMed PMID: 30057276; PubMed Central PMCID: PMC6103825.
24. Belleannée C, Belghazi M, Labas V, Teixeira-Gomes A, Gatti J, Dacheux J, et al. Purification and identification of sperm surface proteins and changes during epididymal maturation. *Proteomics*. 2011;1952-64.
25. Gervasi M, Visconti P. Molecular changes and signaling events occurring in spermatozoa during epididymal maturation. *Andrology*. 2017:204-18.
26. Sharma U. Paternal Contributions to Offspring Health: Role of Sperm Small RNAs in Intergenerational Transmission of Epigenetic Information. *Frontiers in cell and developmental biology*. 2019;7:215. Epub 2019/11/05. doi: 10.3389/fcell.2019.00215. PubMed PMID: 31681757; PubMed Central PMCID: PMC6803970.
27. Sellem E, Marthey S, Rau A, Jouneau L, Bonnet A, Perrier JP, et al. A comprehensive overview of bull sperm-borne small non-coding RNAs and their diversity across breeds. *Epigenetics & chromatin*. 2020;13(1):19. Epub 2020/04/02. doi: 10.1186/s13072-020-00340-0. PubMed PMID: 32228651; PubMed Central PMCID: PMC7106649.
28. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic acids research*. 2012;40(1):37-52. doi: 10.1093/nar/gkr688. PubMed PMID: 21911355; PubMed Central PMCID: PMC3245920.
29. Le Guillou S, Marthey S, Laloë D, Laubier J, Mobuchon L, Leroux C, et al. Characterisation and comparison of lactating mouse and bovine mammary gland miRNomes. *PloS one*. 2014;9(3):e91938. doi: 10.1371/journal.pone.0091938. PubMed PMID: 24658750; PubMed Central PMCID: PMC3962357.
30. Capra E, Turri F, Lazzari B, Cremonesi P, Gliozzi TM, Fojadelli I, et al. Small RNA sequencing of cryopreserved semen from single bull revealed altered miRNAs and piRNAs expression between High- and Low-motile sperm populations. *BMC genomics*. 2017;18(1):14. doi: 10.1186/s12864-016-3394-7. PubMed PMID: 28052756; PubMed Central PMCID: PMC5209821.

31. Rosenkranz D. piRNA cluster database: a web resource for piRNA producing loci. *Nucleic acids research*. 2016;44(D1):D223-30. doi: 10.1093/nar/gkv1265. PubMed PMID: 26582915; PubMed Central PMCID: PMC4702893.
32. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics*. 2009;10:48. doi: 10.1186/1471-2105-10-48. PubMed PMID: 19192299; PubMed Central PMCID: PMC2644678.
33. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic acids research*. 2017;45(W1):W130-W7. doi: 10.1093/nar/gkx356. PubMed PMID: 28472511; PubMed Central PMCID: PMC5570149.
34. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one*. 2011;6(7):e21800. doi: 10.1371/journal.pone.0021800. PubMed PMID: 21789182; PubMed Central PMCID: PMC3138752.
35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PubMed PMID: 25516281; PubMed Central PMCID: PMC4302049.
36. Dadoune JP, Siffroi JP, Alfonsi MF. Transcription in haploid male germ cells. *Int Rev Cytol*. 2004;237:1-56. doi: 10.1016/S0074-7696(04)37001-4. PubMed PMID: 15380665.
37. Pessot CA, Brito M, Figueroa J, Concha, II, Yanez A, Burzio LO. Presence of RNA in the sperm nucleus. *Biochemical and biophysical research communications*. 1989;158(1):272-8. Epub 1989/01/16. doi: 10.1016/S0006-291x(89)80208-6. PubMed PMID: 2463835.
38. Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science*. 2016;351(6271):391-6. doi: 10.1126/science.aad6780. PubMed PMID: 26721685; PubMed Central PMCID: PMC4888079.
39. Nixon B, Stanger SJ, Mihalas BP, Reilly JN, Anderson AL, Tyagi S, et al. The MicroRNA Signature of Mouse Spermatozoa Is Substantially Modified During Epididymal Maturation. *Biology of reproduction*. 2015. doi: 10.1095/biolreprod.115.132209. PubMed PMID: 26333995.
40. Sharma U, Conine CC, Shea JM, Boskovic A, Derr AG, Bing XY, et al. Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science*. 2016;351(6271):391-6. Epub 2016/01/02. doi: 10.1126/science.aad6780. PubMed PMID: 26721685; PubMed Central PMCID: PMC4888079.
41. Peng H, Shi J, Zhang Y, Zhang H, Liao S, Li W, et al. A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Res*. 2012;22(11):1609-12. Epub 2012/10/10. doi: 10.1038/cr.2012.141. PubMed PMID: 23044802; PubMed Central PMCID: PMC3494397.
42. Chu C, Yu L, Wu B, Ma L, Gou LT, He M, et al. A sequence of 28S rRNA-derived small RNAs is enriched in mature sperm and various somatic tissues and possibly associates with inflammation. *J Mol Cell Biol*. 2017;9(3):256-9. Epub 2017/05/10. doi: 10.1093/jmcb/mjx016. PubMed PMID: 28486659.
43. Belleanne C, Calvo E, Caballero J, Sullivan R. Epididymosomes convey different repertoires of microRNAs throughout the bovine epididymis. *Biology of reproduction*. 2013;89(2):30. Epub 2013/06/28. doi: 10.1095/biolreprod.113.110486. PubMed PMID: 23803555.
44. Perera BPU, Tsai ZT-Y, Colwell ML, Jones TR, Goodrich JM, Wang K, et al. Somatic expression of piRNA and associated machinery in the mouse identifies short, tissue-specific piRNA. *Epigenetics : official journal of the DNA Methylation Society*. 2019;14(5):504-21. doi: 10.1080/15592294.2019.1600389.
45. Ortogero N, Schuster AS, Oliver DK, Riordan CR, Hong AS, Hennig GW, et al. A novel class of somatic small RNAs similar to germ cell pachytene PIWI-interacting small RNAs. *The Journal of biological chemistry*. 2014;289(47):32824-34. doi: 10.1074/jbc.M114.613232. PubMed PMID: 25320077; PubMed Central PMCID: PMC4239631.
46. Hutcheon K, McLaughlin EA, Stanger SJ, Bernstein IR, Dun MD, Eamens AL, et al. Analysis of the small non-protein-coding RNA profile of mouse spermatozoa reveals specific enrichment of

- piRNAs within mature spermatozoa. *RNA Biology*. 2017;14(12):1776-90. doi: 10.1080/15476286.2017.1356569.
47. Tosar JP, Rovira C, Cayota A. Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. *Communications Biology*. 2018;1(1):2. doi: 10.1038/s42003-017-0001-7.
48. Larriba E, Rial E, del Mazo J. The landscape of mitochondrial small non-coding RNAs in the PGCs of male mice, spermatogonia, gametes and in zygotes. *BMC genomics*. 2018;19(1):634. doi: 10.1186/s12864-018-5020-3.
49. Barrenada O, Fernandez-Perez D, Larriba E, Brieno-Enriquez M, Del Mazo J. Diversification of piRNAs expressed in PGCs and somatic cells during embryonic gonadal development. *RNA biology*. 2020;17(9):1309-23. Epub 2020/05/08. doi: 10.1080/15476286.2020.1757908. PubMed PMID: 32375541.
50. Zhao S, Gou LT, Zhang M, Zu LD, Hua MM, Hua Y, et al. piRNA-triggered MIWI ubiquitination and removal by APC/C in late spermatogenesis. *Developmental cell*. 2013;24(1):13-25. doi: 10.1016/j.devcel.2012.12.006. PubMed PMID: 23328397.
51. Krawetz SA, Kruger A, Lalancette C, Tagett R, Anton E, Draghici S, et al. A survey of small RNAs in human sperm. *Human reproduction*. 2011;26. doi: 10.1093/humrep/der329.
52. Garcia-Lopez J, Alonso L, Cardenas DB, Artaza-Alvarez H, Hourcade Jde D, Martinez S, et al. Diversity and functional convergence of small noncoding RNAs in male germ cell differentiation and fertilization. *Rna*. 2015;21(5):946-62. doi: 10.1261/rna.048215.114. PubMed PMID: 25805854; PubMed Central PMCID: PMC4408801.
53. Zoch A, Auchynnikava T, Berrens RV, Kabayama Y, Schopp T, Heep M, et al. SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. *Nature*. 2020;584(7822):635-9. Epub 2020/07/17. doi: 10.1038/s41586-020-2557-5. PubMed PMID: 32674113.
54. Watanabe T, Cheng EC, Zhong M, Lin H. Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. *Genome research*. 2015;25(3):368-80. doi: 10.1101/gr.180802.114. PubMed PMID: 25480952; PubMed Central PMCID: PMC4352877.
55. Zhang P, Kang JY, Gou LT, Wang J, Xue Y, Skogerboe G, et al. MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Res*. 2015;25(2):193-207. doi: 10.1038/cr.2015.4. PubMed PMID: 25582079; PubMed Central PMCID: PMC4650574.
56. Guzel E, Arlier S, Guzeloglu-Kayisli O, Tabak MS, Ekiz T, Semerci N, et al. Endoplasmic Reticulum Stress and Homeostasis in Reproductive Physiology and Pathology. *International journal of molecular sciences*. 2017;18(4). Epub 2017/04/12. doi: 10.3390/ijms18040792. PubMed PMID: 28397763; PubMed Central PMCID: PMC5412376.
57. Karna KK, Shin YS, Choi BR, Kim HK, Park JK. The Role of Endoplasmic Reticulum Stress Response in Male Reproductive Physiology and Pathology: A Review. *World J Mens Health*. 2020;38(4):484-94.
58. Koenig PA, Nicholls PK, Schmidt FI, Hagiwara M, Maruyama T, Frydman GH, et al. The E2 ubiquitin-conjugating enzyme UBE2J1 is required for spermiogenesis in mice. *The Journal of biological chemistry*. 2014;289(50):34490-502. Epub 2014/10/17. doi: 10.1074/jbc.M114.604132. PubMed PMID: 25320092; PubMed Central PMCID: PMC4263858.
59. Gur Y, Breitbart H. Mammalian sperm translate nuclear-encoded proteins by mitochondrial-type ribosomes. *Genes & development*. 2006;20(4):411-6. Epub 01/31. doi: 10.1101/gad.367606. PubMed PMID: 16449571.
60. Gur Y, Breitbart H. Protein synthesis in sperm: dialog between mitochondria and cytoplasm. *Molecular and cellular endocrinology*. 2008;282(1-2):45-55. Epub 2008/01/30. doi: 10.1016/j.mce.2007.11.015. PubMed PMID: 18226442.
61. Zhu Z, Umehara T, Okazaki T, Goto M, Fujita Y, Hoque SAM, et al. Gene Expression and Protein Synthesis in Mitochondria Enhance the Duration of High-Speed Linear Motility in Boar Sperm. *Frontiers in physiology*. 2019;10(252). doi: 10.3389/fphys.2019.00252.
62. Morita M, Gravel SP, Hulea L, Larsson O, Pollak M, St-Pierre J, et al. mTOR coordinates protein synthesis, mitochondrial activity and proliferation. *Cell cycle*. 2015;14(4):473-80. Epub

- 2015/01/16. doi: 10.4161/15384101.2014.991572. PubMed PMID: 25590164; PubMed Central PMCID: PMC4615141.
63. Silva JV, Cabral M, Correia BR, Carvalho P, Sousa M, Oliveira PF, et al. mTOR Signaling Pathway Regulates Sperm Quality in Older Men. *Cells*. 2019;8(6). Epub 2019/06/27. doi: 10.3390/cells8060629. PubMed PMID: 31234465; PubMed Central PMCID: PMC6627782.
64. Mahran AM, Mosad E, Abdel-Raheem MA, Ahmed EH, Abdel Motaleb AA, Hofny ER. The correlation between mammalian target of rapamycin (mTOR) gene expression and sperm DNA damage among infertile patients with and without varicocele. *Andrologia*. 2019;51(9):e13341. Epub 2019/06/14. doi: 10.1111/and.13341. PubMed PMID: 31195431.
65. Pena FJ, Rodriguez Martinez H, Tapia JA, Ortega Ferrusola C, Gonzalez Fernandez L, Macias Garcia B. Mitochondria in mammalian sperm physiology and pathology: a review. *Reproduction in domestic animals = Zuchthygiene*. 2009;44(2):345-9. Epub 2009/01/16. doi: 10.1111/j.1439-0531.2008.01211.x. PubMed PMID: 19144010.
66. Aitken RJ, Gibb Z, Baker MA, Drevet J, Gharagozloo P. Causes and consequences of oxidative stress in spermatozoa. *Reproduction, fertility, and development*. 2016;28(1-2):1-10. doi: 10.1071/RD15325. PubMed PMID: 27062870.
67. Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA. Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature*. 2004;429(6988):154. doi: 10.1038/429154a. PubMed PMID: 15141202.
68. Guo L, Chao SB, Xiao L, Wang ZB, Meng TG, Li YY, et al. Sperm-carried RNAs play critical roles in mouse embryonic development. *Oncotarget*. 2017;8(40):67394-405. Epub 2017/10/06. doi: 10.18632/oncotarget.18672. PubMed PMID: 28978041; PubMed Central PMCID: PMC5620181.
69. Yuan S, Schuster A, Tang C, Yu T, Ortogero N, Bao J, et al. Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. *Development*. 2016;143(4):635-47. Epub 2016/01/01. doi: 10.1242/dev.131755. PubMed PMID: 26718009; PubMed Central PMCID: PMC4760322.
70. Hosken DJ, Hodgson DJ. Why do sperm carry RNA? Relatedness, conflict, and control. *Trends Ecol Evol*. 2014;29(8):451-5. Epub 2014/06/12. doi: 10.1016/j.tree.2014.05.006. PubMed PMID: 24916312.
71. Zhang Y, Shi J, Rassoulzadegan M, Tuorto F, Chen Q. Sperm RNA code programmes the metabolic health of offspring. *Nature reviews Endocrinology*. 2019;15(8):489-98. Epub 2019/06/27. doi: 10.1038/s41574-019-0226-2. PubMed PMID: 31235802; PubMed Central PMCID: PMC6626572.
72. Fei T, Zhu S, Xia K, Zhang J, Li Z, Han JD, et al. Smad2 mediates Activin/Nodal signaling in mesendoderm differentiation of mouse embryonic stem cells. *Cell research*. 2010;20(12):1306-18. Epub 2010/11/17. doi: 10.1038/cr.2010.158. PubMed PMID: 21079647.
73. Morikawa Y, Cserjesi P. Extra-embryonic vasculature development is regulated by the transcription factor HAND1. *Development*. 2004;131(9):2195-204. Epub 2004/04/10. doi: 10.1242/dev.01091. PubMed PMID: 15073150.
74. Lessard JA, Crabtree GR. Chromatin regulatory mechanisms in pluripotency. *Annual review of cell and developmental biology*. 2010;26:503-32. doi: 10.1146/annurev-cellbio-051809-102012. PubMed PMID: 20624054.
75. Assou S, Cerecedo D, Tondeur S, Pantesco V, Hovatta O, Klein B, et al. A gene expression signature shared by human mature oocytes and embryonic stem cells. *BMC genomics*. 2009;10:10. doi: 10.1186/1471-2164-10-10. PubMed PMID: 19128516; PubMed Central PMCID: PMC2628676.
76. Thompson PJ, Norton KA, Niri FH, Dawe CE, McDermid HE. CECR2 is involved in spermatogenesis and forms a complex with SNF2H in the testis. *Journal of molecular biology*. 2012;415(5):793-806. doi: 10.1016/j.jmb.2011.11.041. PubMed PMID: 22154806.
77. Donkin I, Barres R. Sperm epigenetics and influence of environmental factors. *Molecular metabolism*. 2018;14:1-11. Epub 2018/03/12. doi: 10.1016/j.molmet.2018.02.006. PubMed PMID: 29525406; PubMed Central PMCID: PMC6034033.

78. Liu WM, Pang RT, Chiu PC, Wong BP, Lao K, Lee KF, et al. Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(2):490-4. doi: 10.1073/pnas.1110368109. PubMed PMID: 22203953; PubMed Central PMCID: PMC3258645.
79. Le Blevet E, Muronova J, Ray PF, Arnoult C. Paternal epigenetics: Mammalian sperm provide much more than DNA at fertilization. *Molecular and cellular endocrinology*. 2020:110964. Epub 2020/08/02. doi: 10.1016/j.mce.2020.110964. PubMed PMID: 32738444.
80. Xu H, Wang X, Wang Z, Li J, Xu Z, Miao M, et al. MicroRNA expression profile analysis in sperm reveals hsa-mir-191 as an auspicious omen of in vitro fertilization. *BMC genomics*. 2020;21(1):165. Epub 2020/02/19. doi: 10.1186/s12864-020-6570-8. PubMed PMID: 32066367; PubMed Central PMCID: PMC7027243.
81. Chen Q, Yan M, Cao Z, Li X, Zhang Y, Shi J, et al. Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science*. 2016;351(6271):397-400. Epub 2016/01/02. doi: 10.1126/science.aad7977. PubMed PMID: 26721680.
82. Chen X, Zheng Y, Lei A, Zhang H, Niu H, Li X, et al. Early cleavage of preimplantation embryos is regulated by tRNA(Gln-TTG)-derived small RNAs present in mature spermatozoa. *The Journal of biological chemistry*. 2020;295(32):10885-900. Epub 2020/06/04. doi: 10.1074/jbc.RA120.013003. PubMed PMID: 32487749; PubMed Central PMCID: PMC7415976.

## Additional tables S1 to S8

**Additional table S1:** The Sperm miRnome. Catalogue of miRNAs of bovine frozen sperm.

**Additional table S2:** Sperm rRNAs. Catalogue of rRNAs of bovine frozen sperm.

**Additional table S3:** Sperm tRNAs. Catalogue of tRNAs of bovine frozen sperm.

**Additional table S4:** Sperm piRNAs. Catalogue of piRNAs of bovine frozen sperm.

**Additional table S5:** Dynamics of tsRNAs

**Additional table S6:** Dynamics of rsRNAs

**Additional table S7:** miRNA targets per k-means cluster

**Additional table S8:** Table S8: Enriched GO terms per k-means cluster

# Chapitre III : Les sncRNA comme biomarqueur de la fertilité mâle ?

## 1) Éléments de contexte

Comme évoqué en introduction, le contrôle qualité de la semence est un enjeu important pour la filière insémination, de façon à garantir une bonne fertilité du taureau et de ses éjaculats, et peut être demain s'assurer de l'absence de désordres épigénétiques pour garantir l'expression du potentiel génétique de la descendance. Différentes approches ont été explorées pour établir un prédicteur de la fertilité, basées principalement sur divers paramètres en lien avec la fonctionnalité ou la physiologie des spermatozoïdes. Mais tous ces paramètres et prédicteurs se sont avérés insuffisants et seule leur combinaison par des approches de régression « stepwise » ou « logistic » a permis d'améliorer la prédiction, qui n'explique toutefois que 40% de la variabilité de la fertilité observée<sup>1</sup>. D'où l'idée d'explorer un éventuel lien entre fertilité et expression des sncRNA du spermatozoïde, idée fondée sur les avancées scientifiques, détaillées dans le Chapitre II, montrant un lien entre sncRNA spermatiques et paramètres fonctionnels de la semence (motilité ou anomalie du spermogramme<sup>237,239,240</sup>), fertilité des mâles<sup>241,242</sup> ou développement de la descendance<sup>167,223,225</sup>.

Dans les chapitres précédents, nous avons réalisé l'inventaire des sncRNA du spermatozoïde bovin, en précisant leur origine tissulaire. L'étape suivante, objet de ce chapitre, consiste à étudier le lien entre l'expression de ces sncRNA et la fertilité des taureaux, afin de mettre en évidence des biomarqueurs utilisables en contrôle qualité de routine, éventuellement combinés aux paramètres fonctionnels ou d'autres biomarqueurs « omics ».

## 2) Plan expérimental

### 2.1) Constitution de la bio-banque d'éjaculats

Les semences de 154 taureaux Holstein (N=54) et Montbéliard (N=100) ont été fournies par trois centres de production de semence (figure 35) : Évolution XY (Saint Aubin du Cormier, France) pour les éjaculats de taureaux Holstein et Umotest (Roulans, France) et Evalura (Crançot, France), pour les taureaux Montbéliards. Afin de limiter d'éventuelles variations liées à l'évolution de pratiques d'élevages au cours du temps, l'année de naissance des taureaux de l'étude a été limitée à la période 2011-2015.

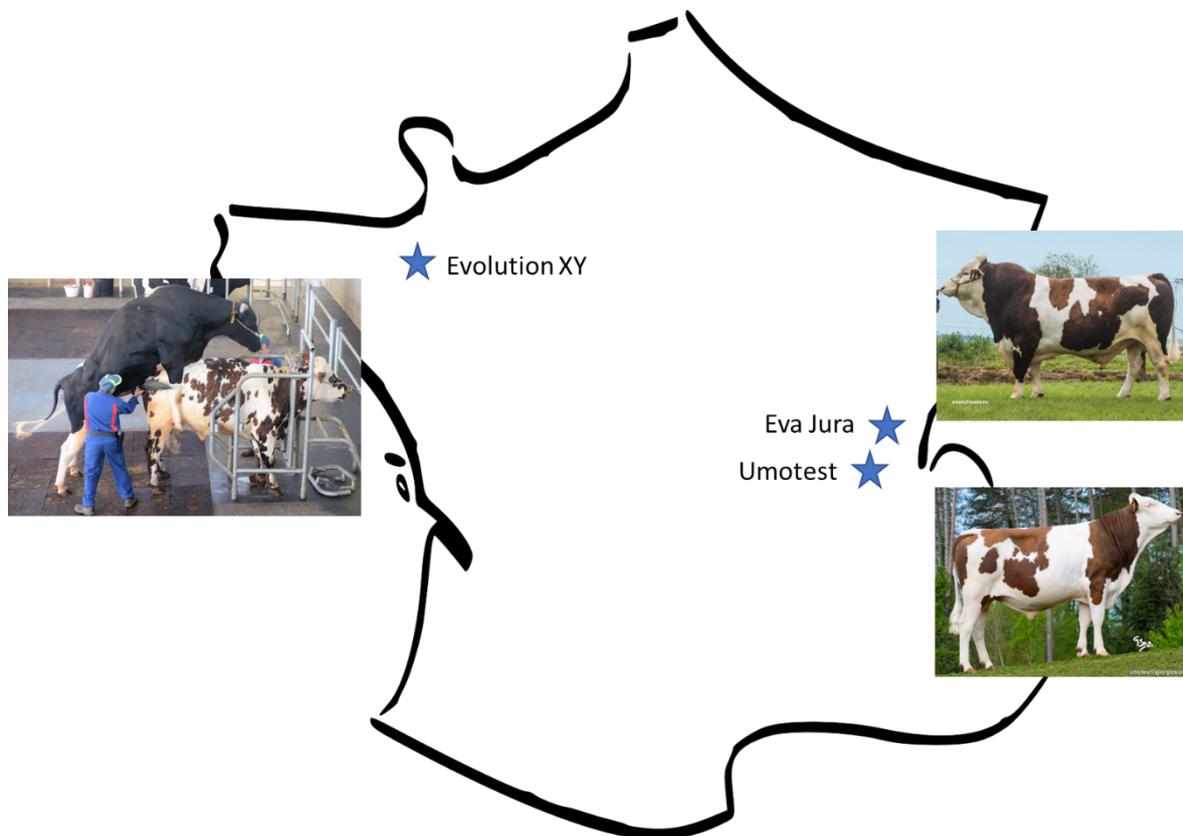


Figure 35 : Localisation des centres de production de semence.

154 éjaculats de taureaux Holstein (N=54) et Montbéliard (N=100) ont été utilisés pour la recherche de biomarqueur de la fertilité mâle. Un seul centre (Évolution XY) a fourni la semence de taureaux Holstein. En revanche, pour celle des taureaux Montbéliards deux centres de production situés relativement proche l'un de l'autre, ont été sollicités (Eva Jura et Umotest). ©Alicce ©Umotest ©Veeteeltvlees pour EVA Jura

Chaque taureau est représenté par un pool de 2 à 5 éjaculats produits entre 17 et 19 mois d'âge. Les taureaux ont été choisis suivant leur indicateur de fertilité basé sur le TNR56 corrigé (effet taureau X campagne issu de l'indexation fertilité femelle à 56 jours), et regroupés en mâles « supfertiles » si leur indicateur est supérieur à +1 écart-type de la moyenne de la

population globale étudiée (centrée réduite) ou en mâles « subfertiles » si leur indicateur est inférieur à -1 écart-type. Chez les taureaux Holstein, 21 sont subfertiles et 33 supfertiles. Chez les taureaux Montbéliards, 43 sont subfertiles et 57 supfertiles.

## 2.2) Analyse des paramètres fonctionnels de la semence

Les principaux paramètres fonctionnels post-décongélation de la semence ont été analysés : viabilité (% de spermatozoïdes ayant une membrane intacte), développement énergétique (% de spermatozoïdes ayant une mitochondrie active (HMMP) et niveau de production énergétique (MYH)), oxydation cellulaire (% de spermatozoïdes vivants oxydés et niveau d'oxydation) et motilité (% de spermatozoïdes mobiles, progressifs et paramètres liés à la vitesse ( $\mu\text{m/s}$ ) et au type ( $\mu\text{m}$  ou %) de déplacement : VAP, VSL, ALH, LIN...). Le système CASA « IVOS II » (Hamilton Thorne) a été utilisé pour l'analyse de la motilité et le cytomètre en flux « Easyocyte 8HT » couplé aux kits « Easykits » (IMV Technologies) pour les autres tests. Les principes et les modes opératoires des différents kits sont décrits dans Sellem et al.,2015<sup>64</sup>.

## 2.3) Extraction d'ARN total et séquençage des sncRNA.

Le protocole d'extraction d'ARN total, basé sur une extraction au Trizol-chloroforme modifiée et optimisée, est identique à celui utilisé précédemment pour l'établissement des catalogues. Le contrôle qualité a comporté un dosage en fluorimétrie (Qubit® RNA HS Assay Kits, Q32852 Life Technologies) comme réalisé précédemment pour l'établissement des catalogues<sup>243</sup>.

Le séquençage NGS ILLUMINA a été confié en prestation de service à Qiagen. Les librairies ont été réalisées à partir de 10ng d'ARN extrait, en utilisant le kit QIAseq miRNA Library Kit® et l'indexation UMI (*Unique Molecular Indexes*). Les UMI représentent un système de code-barre utilisé pour supprimer les biais techniques d'amplifications lors de la création des banques. L'approche est donc identique à celle mise en œuvre pour l'établissement de la dynamique d'expression et l'origine tissulaire des sncRNA. Le séquençage haut débit a été réalisé à l'aide

d'un séquenceur Illumina HiSeq2000, avec un objectif de 20 millions de séquences (75 paires de bases en *single-end*) par échantillon.

Le pipeline bio-informatique est également identique à celui utilisé précédemment : les séquences résultantes de biais d'amplification ont été éliminées, tout comme les séquences inférieures à 17 nucléotides ou présentant un score de qualité faible (Phred score < 25) ; miRDeep2 a ensuite été utilisé, avec le génome bovin (BosTau9) comme référence, pour quantifier et identifier les miRNA connus (miRBase v22) et prédire de nouveaux miRNA ; les séquences ont été annotées à l'aide de plusieurs bases de données publiques, dont Ensembl release 94 ([www.ensembl.org](http://www.ensembl.org)), RFAM release 14 (<http://rfam.wustl.edu>), les ARN non codants bovins de l'ENA (<https://www.ebi.ac.uk/ena>), les ARNt bovins de GtRNadb ([www.gtrnadb.ucsc.edu](http://www.gtrnadb.ucsc.edu)) et les piARN humains, murins et bovins de l'ENA.

#### 2.4) Les différentes approches statistiques

Un filtre sur le niveau d'expression a été appliqué, afin d'éliminer des analyses toutes les séquences dont la somme des comptages, toutes banques confondues, ne dépassait pas 100 copies. Deux approches statistiques ont été utilisées pour mettre en évidence d'éventuels biomarqueurs : une analyse statistique descriptive multidimensionnelle pour traiter un grand nombre de variable et s'extraire de tout modèle sous-jacent (Between Class Analysis, BCA ; package R *ade4* avec la fonction *dudi.bca*) et une analyse différentielle pour identifier les sncRNA présentant des différences d'expression statistiquement significatives (package R *DeSeq2*, p-value ajustée (Benjamini-Hochberg)  $\leq 5\%$ ).

La mise au point d'une équation de prédiction en race Montbéliarde, exploitant un sous-ensemble pertinent de biomarqueurs, a reposé sur une approche « d'ensemble learning » combinant deux algorithmes de « machine learning » (collaboration avec Datastat, IDELE). Pour ce faire, le dispositif de taureaux a été divisé en deux sous-ensembles : un sous-ensemble « d'apprentissage » (70% de la population étudiée) et un sous-ensemble de « test » (30% restant). Afin de réduire le nombre de variables à traiter, la matrice des sncRNA a été réduite aux séquences différentiellement exprimées identifiées par *DeSeq2*. Les variables les plus informatives ont été sélectionnées par la méthode VSURF (package R *vsurf*). La prédiction de fertilité est ensuite réalisée par un méta-modèle, qui combine les prédictions obtenues à

partir des méthodes de « Séparateurs à vaste marge » (SVM) et de « Forêts aléatoires » (package R « caretEnsemble »). L'algorithme des forêts aléatoires génère plusieurs « arbres de décision ». Chaque arbre est entraîné sur un sous-ensemble aléatoire d'observations avec un sous-ensemble aléatoire de variables. Les prédictions sont ensuite compilées (moyenne pour une donnée quantitative et un vote pour une qualitative) afin de rendre un statut à chaque taureau. La méthode SVM consiste à séparer les données en classes à l'aide d'une frontière simple, de telle façon que la distance entre les différents groupes de données et la frontière qui les sépare soit maximale. L'utilisation d'un méta-classifieur (« ensemble learning »), combinant ces deux approches, permet non seulement d'améliorer le pouvoir prédictif du modèle, mais également de remédier au phénomène de surapprentissage. Le prédicteur ainsi obtenu sur l'ensemble d'apprentissage est ensuite appliqué sur l'échantillon de test. La précision, la sensibilité et la spécificité, métriques données par la table de confusion, sont utilisées pour juger de l'efficacité du prédicteur.

Le package Hmisc a été utilisé pour créer les matrices de corrélations entre les phénotypes. Les corrélations de Pearson sont considérées comme significatives lorsque les p-values ajustées (Benjamini-Hochberg) sont inférieures ou égale à 5%. Seules les corrélations dont la valeur absolue est supérieure à 0.7 sont analysées.

### 3) Résultats

#### 3.1) Fertilité et paramètres spermatiques des éjaculats utilisés

Comme illustré par la figure 36, les deux groupes de fertilité intra race présentent une différence significative ( $p < 0.05$ ) de fertilité, mesurée par le TNR56 corrigé :  $[-0.8 \pm 1.3 ; +2.6 \pm 0.9]$  et  $[-1.3 \pm 1.1 ; +3.6 \pm 1.0]$  en moyenne plus ou moins l'écart-type pour respectivement les taureaux Holstein et Montbéliard subfertiles et supfertiles.

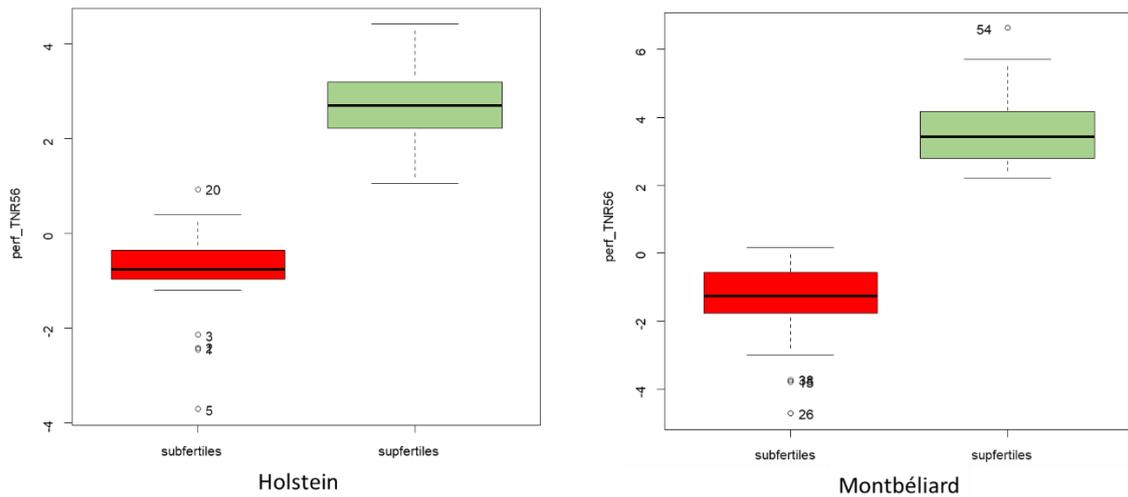


Figure 36 : Groupes de fertilité taureaux Holstein et Montbéliard.

Pour les deux races de l'étude, les deux groupes de taureaux choisis présentent bien une différence importante de fertilité.

En dépit de ce différentiel important de fertilité, la qualité moyenne post-décongélation des éjaculats utilisés ne diffère pas significativement entre les groupes de taureaux, comme illustré sur la figure 37.

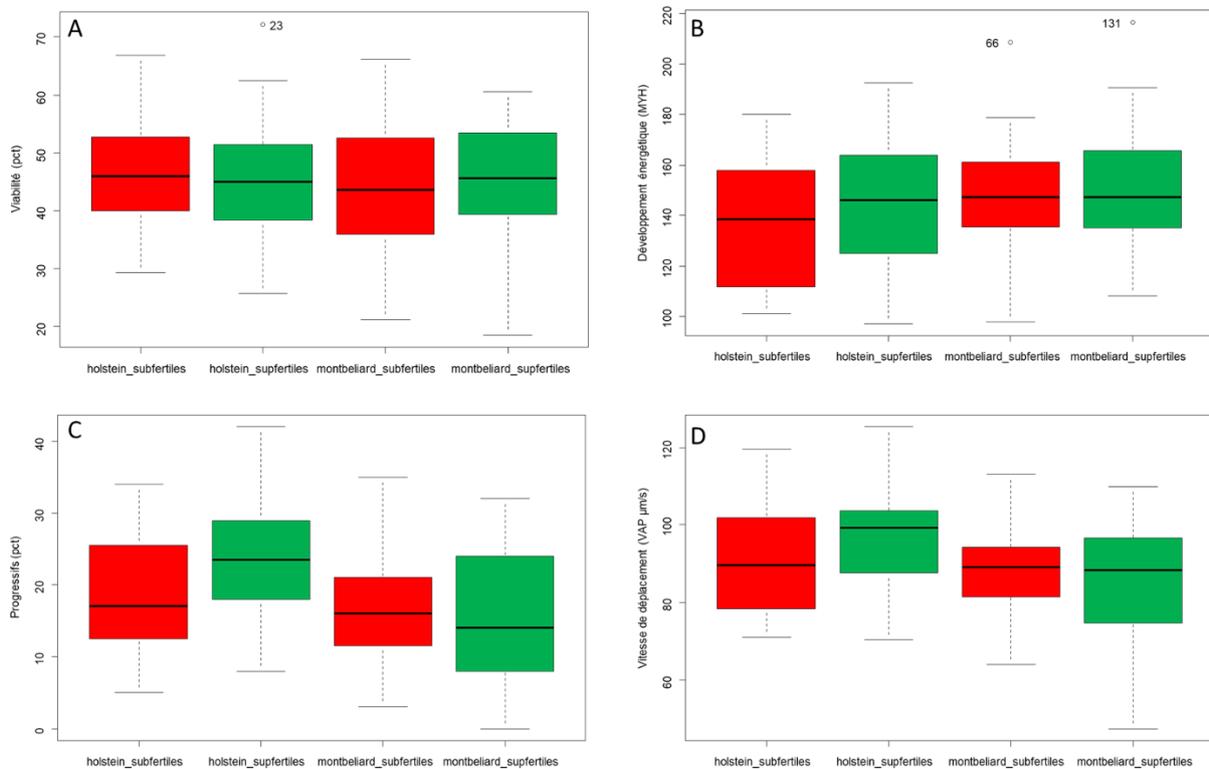


Figure 37 : Fonctionnalité des éjaculats utilisés.

Des mesures de qualité de la semence ont été réalisées à l'aide du CASA (motilité) et de la cytométrie en flux (viabilité et développement énergétique). Ces graphiques illustrent la variabilité d'un phénotype au sein du

groupe de fertilité d'une race en particulier. (A) Pourcentage de spermatozoïdes viables, (B) moyenne d'intensité de fluorescence représentant le niveau énergétique de la cellule, (C) pourcentage de spermatozoïdes progressifs et (D) vitesse de déplacement en  $\mu\text{m/s}$ .

### 3.2) Données obtenues par séquençage

Globalement, 413 952 sncRNA ont été identifiés, présentant une grande gamme de niveaux d'expression moyens (de <1 à 289 374 copies). Environ 94% de ces sncRNA ont une moyenne d'expression tous taureaux confondus inférieure à 10 copies, 5% entre 10 à 100 copies et moins de 1% plus de 100 copies. Seules 42 séquences présentent une expression supérieure à 10 000 copies et uniquement 4 avec une expression supérieure à 100 000 copies.

Les miRNA ne représentent que 3% des séquences identifiées, mais 41% de l'expression totale normalisée (figure 38). Les piRNA majoritairement présents en nombre de séquence (40%) ne représentent que 11% de l'expression globale.

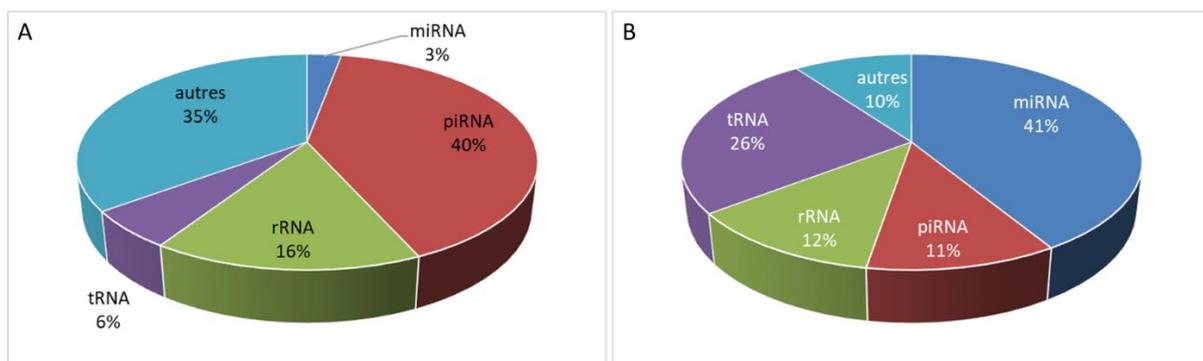


Figure 38 : Diversité des séquences contenues dans le spermatozoïde éjaculé.

La diversité des sncRNA est représentée en pourcentage du nombre de séquences présentes (A) ou en pourcentage d'expression (B). Il est intéressant de noter que les miRNA ne représentent que 3% des séquences contenues dans le spermatozoïde éjaculé, mais expriment 41% de l'expression totale. Les piRNA majoritairement présents en nombre de séquence ne représentent qu'un dixième de l'expression globale.

Les tsRNA et rsRNA contribuent à hauteur de 26% et 12% à l'expression globale, respectivement. Les autres classes de sncRNA (circRNA, snoRNA), les fragments d'ARN messenger et les séquences non identifiées représentent 10% de l'expression globale). Les 20 séquences les plus exprimées en moyenne représentent 33% de l'expression totale (240 677 903 copies). Elles comportent 14 miRNA (12 connus et 2 prédits), représentant 68.4% de l'expression du « Top20 » et 6 tsRNA (31.6% de l'expression du total du « Top20 » ; figure 39). Deux des miRNA du « Top20 » sont représentés par deux isomiR (bta-miR-191 et bta-miR-21-

5p). Le bta-miR-16 est quant à lui représenté par trois isomiR. Parmi les 6 tsRNA présents dans le « Top20 », 3 isoaccepteurs sont représentés : Gln, Gly et Ser.



Figure 39 : Diversité des sncRNA selon leur niveau d'expression.

33% de l'expression globale des sncRNA sont captés par les 20 séquences les plus exprimées. Seulement deux familles de sncRNA composent ce Top20 (miRNA et tRNA) et majoritairement par des miRNA (68%). Le reste de l'expression est constitué majoritairement par des miRNA (28%), tRNA (22%), rRNA (18%) et piRNA (17%).

### 3.3) sncRNA biomarqueurs de fertilité mâle

L'analyse descriptive par BCA consiste, comme l'ACP, à déterminer des combinaisons de variables décorréliées les unes des autres, formant des axes principaux indépendants expliquant au mieux la variance des données. Dans le cas de la BCA, les axes sont calculés en fonction des groupes d'échantillons, ce qui permet d'identifier des axes et un espace des individus qui maximise les différences entre les groupes. La figure 40, illustre graphiquement la séparation de ces groupes selon les deux premiers axes. La part de variabilité imputée aux différents groupes de fertilité est de l'ordre de 3%. Visiblement, la race de taureau est un facteur expliquant un pourcentage bien plus élevé de la variabilité globale obtenue entre tous les taureaux de l'étude. Bien que la fertilité représente un facteur peu explicatif de la variabilité globale, la séparation entre les groupes de fertilité est claire et ce résultat souligne le fait que certains sncRNA sont vraisemblablement significativement différentiellement exprimés entre les groupes de fertilité.

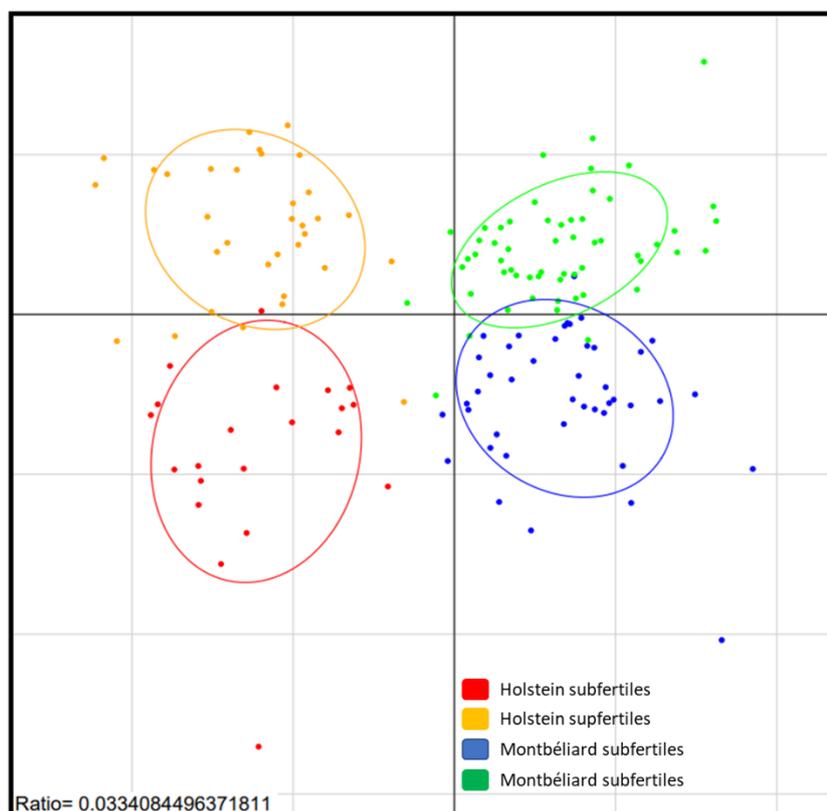


Figure 40 : BCA des groupes de fertilité des taureaux Holstein et Montbéliard.

Environ 3% de la variabilité totale observée peut être expliquée par les groupes de fertilité. La race en revanche semble expliquer nettement plus les différences entre ces taureaux.

L'analyse différentielle a été réalisée avec DeSeq2 pour chaque race séparément au vu des résultats de la BCA. Afin de réduire le nombre de variables et dans un souci pratique (i.e identifier des biomarqueurs suffisamment exprimés pour permettre une exploitation en routine), seules 30 260 sncRNA ont été analysés, qui présentent une moyenne d'expression brute intra race supérieure à 10 copies. Tous les sncRNA ayant une *p-value* ajustée inférieure à 0.05 ont été considérés comme différentiellement exprimés de façon significative. Toutefois, pour cette première approche d'identification de biomarqueurs de la fertilité, ceux présentant une *p-value* ajustée entre 0.05 et 0.1 ont également été sélectionnés et considéré comme différentiellement exprimés afin de ne pas écarter trop vite des sncRNA potentiellement intéressants. Le tableau suivant (Table 1) liste en fonction des races et des familles, le nombre de sncRNA significatifs et les 20 sncRNA les plus exprimées sont illustrés sur la figure 41.

<i>Holstein</i>				<i>Montbéliarde</i>		
Classe	Nbr	Up	Down	Nbr	Up	Down
rsRNA	20	4	16	400	149	251
tsRNA	4	2	2	21	9	12
miRNA	22	0	22	7	6	1
piRNA	13	2	11	82	36	46
Autres	17	9	8	42	18	24
<b>Total</b>	<b>76</b>	<b>17</b>	<b>59</b>	<b>552</b>	<b>218</b>	<b>334</b>

Table 1 : sncRNA différemment exprimés selon le niveau de fertilité.

Sont indiqués pour chaque classe de sncRNA le nombre total de sncRNA (Nbr) ainsi que le nombre de sncRNA surexprimés (up) ou sous-exprimés (down) chez les supfertiles pour les taureaux Holstein et Montbéliard.

read	Fold_change	Adj_pVal	race	Famille	expression_subfertiles (Moy)	expression_supfertiles (Moy)
TGAGAACTGAATTCATAGGTT	1.833	0.056	Holstein	miRNA	25307.1	13803.5
ACAGCTGACTTCCAATCAGCTAGTTTCGGTCT	0.549	0.045	Montbéliard	tRNA	593.3	1080.8
GTACGAAAGGACAAGAGAATAAGGCC	0.637	0.056	Montbéliard	rRNA	523.4	822.5
CGACTGGCGATGCGGCGG	1.355	0.091	Montbéliard	rRNA	762.0	562.5
CCGGGGAGGTAGTGACGAAAAATAACAATACAGGA	1.524	0.063	Montbéliard	rRNA	663.2	435.5
AGTAAGGTCAGCTAATTAAGCTATCGGGCCCATACCCCGA	2.146	0.002	Montbéliard	tRNA	718.7	334.9
CCGGGGAGGTAGTGACGAAAAATAACAATACAGGAC	1.508	0.083	Montbéliard	rRNA	624.5	414.2
TTGCCAAGAATGTTTTCAATCAAGAACGAAAGTCGG	1.570	0.070	Montbéliard	rRNA	627.3	399.6
CCGGGGAGGTAGTGACGAAAAATAACAATACAGGAC	1.578	0.053	Montbéliard	rRNA	621.6	394.0
CCAACAACAAATGATTAACCTAGCCCAATACTGGA	0.603	0.082	Montbéliard	rRNA	362.5	601.3
CCGGGGAGGTAGTGACGAAAAATAACAATACAGGA	1.568	0.053	Montbéliard	rRNA	549.2	350.6
CCAACAACAAATGATTAACCTAGCCCAATACTGGA	0.605	0.093	Montbéliard	rRNA	335.8	555.6
TGAGAACTGAATTCATAGGTTTT	2.051	0.056	Holstein	miRNA	590.6	287.8
GATGGATATGCTTCAAGGAC	1.570	0.084	Montbéliard	non-identifiée	409.9	338.0
ATGGAGTGGTTAGTGGGGTGGTA	4.238	0.048	Montbéliard	piRNA	579.8	136.8
CAGCTGACTTCCAATCAGCTAGTTTCGGTCT	0.627	0.080	Montbéliard	tRNA	271.7	433.9
GCGGGACCGGGGGGTTCCCGGGG	0.243	0.009	Montbéliard	piRNA	137.4	565.0
AGGTGAAATCTTGGACGGG	1.413	0.077	Montbéliard	rRNA	407.8	288.7
CAACAACAAATGATTAACCTAGCCCAATACTGGA	0.596	0.061	Montbéliard	rRNA	250.9	421.6
GTTAGGTCAAGGTGAACCTATGAAATGGG	0.649	0.048	Montbéliard	rRNA	261.8	403.8

Figure 41 : Séquences différemment les plus exprimées entre les groupes de fertilité.

Les analyses réalisées à l'aide du script DeSeq2 ont mis en évidence 628 séquences d'intérêt différemment exprimées entre les groupes de fertilité, dont 196 avec une p-value ajustée inférieure à 5%. Parmi ces séquences, 81 présentent un nombre de copies moyen par banque supérieur à 100. Les 20 plus exprimées sont illustrées dans ce tableau.

Le détail des types de tsRNA et des iso-accepteurs correspondants est donné dans le tableau suivant (Table 2).

	<i>Holstein</i>	<i>Montbéliarde</i>
<b>tRF5</b>		1
<b>5'-tRH</b>		10
<b>i-tRF</b>	3	9
<b>3'-tRH</b>	1	1
<b>Ala</b>		1
<b>Asp</b>	1	
<b>Gln</b>		1
<b>Gly</b>		6
<b>Met</b>	1	11
<b>Ser</b>	2	1
<b>Val</b>		1
<b>12S</b>		44
<b>16S</b>	2	104
<b>18S</b>	2	179
<b>28S</b>	13	70
<b>5.8S</b>	1	3
<b>5S</b>	2	

Table 2 : Détail des tsRNA et rsRNA différenciellement exprimés selon le niveau de fertilité.

Est indiqué le nombre de biomarqueurs de chaque type pour les taureaux Holstein et Montbéliard. tRF5 : fragment 5' de ARNt ; 5'-tRH et 3'-tRH : moitié 5' ou 3' d'un ARNt ; i-tRF : fragment interne d'un ARNt.

Pour les taureaux Holstein, 76 sncRNA présentent une expression différentielle entre les deux groupes de fertilité ( $p$ -value ajustée  $\leq 0.1$ ), dont 45 avec une  $p$ -value ajustée significative ( $\leq 0.05$ ). Les miRNA (22), rsRNA (20) et dans une moindre mesure les piRNA (13) sont représentés par un nombre de sncRNA équivalent. Les tsRNA sont en revanche moins nombreux (4). Parmi les 76 biomarqueurs chez les Holstein, 17 sont surexprimées chez les taureaux supfertiles et 59 le sont chez les taureaux subfertiles. En termes d'expression, les miRNA représentent 93% de l'expression totale des biomarqueurs (groupes de fertilité confondus). Seuls sept biomarqueurs (4 miRNA, 2 rsRNA et 1 piRNA) sont exprimés en moyenne à plus de 100 copies. Les 22 séquences de miRNA correspondent à 2 miRNA connus : bta-miR-146a,b et bta-miR-192. Étonnamment, 21 correspondent à des isomiR de bta-miR-146a,b, sous-exprimées

chez les taureaux supfertiles (moyenne des « *fold change* » : 2) et représentant plus de 98% de l'expression de cette classe (et 92% de l'expression totale des sncRNA d'intérêt). Une analyse d'enrichissement en termes GO (*Gene Ontologie*) à partir des 257 gènes cibles de bta-miR-146a,b met en évidence quelques processus biologiques enrichis (figure 42), dont plusieurs en lien avec le développement cellulaire et tissulaire (développement des cellules mésenchymateuses, morphogenèse cellulaire ou tissulaire, différenciation épithéliale). Les tsRNA correspondent principalement à des fragments internes d'ARNt (i-tRF). Les fragments d'ARNr 28S représentent 68% des rsRNA.

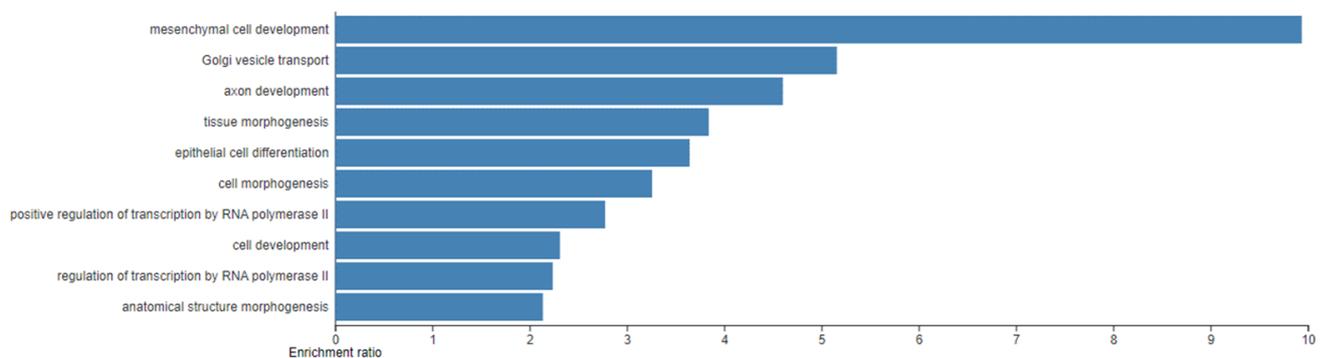


Figure 42 : Processus biologiques significativement enrichis par les gènes ciblés du bta-miR-146. Le bta-miR-146 cible *in silico* 257 gènes (Targetscan). Différents processus biologiques sont significativement enrichis (Gene Ontology ; Webgestalt).

Concernant les taureaux Montbéliards, 552 sncRNA sont identifiées avec une *p-value* ajustée  $\leq 0.1$ , dont 151 significativement (*p-value* ajustée  $\leq 5\%$ ). Contrairement aux Holstein où les proportions des différentes familles étaient plutôt équilibrées, les rsRNA représentent ici une importante majorité (72%) des biomarqueurs. La seconde famille la plus représentée est celle des piRNA (15%). Les autres familles représentent chacune moins de 10% du nombre de sncRNA d'intérêt. Parmi ces 552 biomarqueurs, 218 sont surexprimées chez les taureaux supfertiles et 334 le sont chez les taureaux subfertiles. Seules trois de ces sncRNA sont communs aux deux races (2 rsRNA et 1 piRNA). Les rsRNA représentent 75% de l'expression totale, les piRNA et tsRNA comptant pour 10% chacun et les miRNA pour moins de 1%. Au total, 75 biomarqueurs (57 rsRNA, 9 piRNA, 8 tsRNA et 1 autre) sont exprimés en moyenne à plus de 100 copies. Les 7 séquences de miRNA correspondent à 4 miRNA connus (bta-miR-6526, bta-miR-2483-3p, bta-miR-2284x, bta-miR-30d) et deux prédits. Les tsRNA

correspondent principalement à des 5'-tRH (48%) et des i-tRF (42%) d'isoaccepteurs Glycine et Méthionine. Contrairement à la race Holstein, les fragments d'ARNr 28S ne représentent que 18% des rsRNA, les fragments d'ARNr 18S étant majoritaires (45%).

Parmi les 628 biomarqueurs, 67% (51 sur 76) en race Holstein et 78% (427 sur 552) en race Montbéliarde correspondent à des sncRNA acquis au cours du transit épидидymaire ou post-éjaculation (table 3).

Race	Groupe	miRNA	piRNA	rRNA	tRNA	nbr_sncRNA
Holstein	Testiculaire		4	3		7
Holstein	Epididymaire	19	2	8	3	32
Holstein	Ejaculé	3	6	9	1	19
Montbéliard	Testiculaire		11	49	3	63
Montbéliard	Epididymaire	2	17	76	13	108
Montbéliard	Ejaculé	5	54	275	5	319

Table 3 : Origine tissulaire des sncRNA différemment exprimés selon le niveau de fertilité.

Les groupes (« testiculaire, épидидymaire et éjaculé ») ont été formés en fonction des clusters Kmeans identifiés dans les travaux « origine ».

Plus particulièrement, chez les taureaux Holstein, 82% des biomarqueurs miRNA sont acquis au niveau de la queue de l'épididyme. Plus de 50% des piRNA le sont au niveau prostatique (alors que 31% présentent une expression testiculaire). L'origine des rsRNA est plus variée : 15% d'origine testiculaire, 15% et 20% issus de la tête et du corps de l'épididyme et 45% des rsRNA d'origine prostatique. La majorité (50%) des tsRNA sont acquis au niveau du corps de l'épididyme.

De même en race Montbéliarde, 80% des biomarqueurs sont acquis au niveau de la queue de l'épididyme (30% des miRNA et tsRNA, 20% des piRNA et 13% des rsRNA) ou au niveau prostatique (71% des miRNA, 65% des piRNA et rsRNA, 20% des tsRNA). Les tsRNA sont acquis principalement (35%) au niveau du corps de l'épididyme.

La recherche des cibles potentielles dans l'ovocyte des miRNA différemment exprimés selon le groupe de fertilité a été effectuée. Le génome embryonnaire étant en grande partie silencieux jusqu'au stade 8 cellules chez le bovin<sup>244</sup>, l'impact des miRNA apportés par le spermatozoïde sur le pool ovocytaire d'ARN messagers pourrait être important. L'ovocyte bovin comptabilise plus de 10 000 d'ARNm différents<sup>245</sup>. Parmi les 1000 ARNm les plus

exprimés dans l'ovocyte, 441 sont ciblés par les 6 miRNA connus identifiés par DeSeq2, tels que ceux des gènes (*STAT3*, *SMARCA5*, *BMPR1A*, *BAZ2A*, *SMAD1*). Afin d'explorer plus avant l'adéquation entre ARN messagers ciblés par ces quelques miRNA du spermatozoïde et leur présence dans l'embryon, l'analyse des gènes exprimés dans l'ICM bovin a été effectuée. Ces cellules qui donneront l'épiblaste puis 3 lignées embryonnaires différentes (mésoderme, ectoderme et endoderme), doivent conserver leur pluripotence afin de dériver vers d'autres lignées cellulaires spécifiques de nos organismes. L'ICM exprime plus de 11 000 gènes chez le bovin et parmi eux, 159 y sont différentiellement surexprimés par rapport aux cellules du trophoctoderme<sup>246</sup>. Les miRNA spermatiques différentiellement exprimés selon la fertilité ciblent 45 de ces 159 ARN messagers. Sans surprise, la voie biologique la plus enrichie par ces 45 gènes est le développement de l'embryon. 20% de ces gènes, soit 9 sur 45, sont impliqués dans cette voie (*FN1*, *HNF4A*, *KDM2B*, *LFNG*, *NANOG*, *OTX2*, *PRDM1*, *SMAD6* et *TANC2*).

#### 3.4) Prédiction de fertilité en race Montbéliarde

Parmi les 552 sncRNA identifiés par DeSeq2 chez les taureaux Montbéliards, l'algorithme VSURF a considéré 183 sncRNA comme les plus informatives. Ce jeu de sncRNA est en tout point semblable au panel complet : pourcentage similaire de sncRNA avec expression différentielle significative, même proportion relative de classes de sncRNA et de sncRNA annotés comme surexprimés dans le spermatozoïde éjaculé.

Le méta-classifieur basé sur la combinaison de deux approches de Forêts aléatoire et SVM a été évalué sur le dispositif de « test » comportant 13 mâles subfertiles et de 17 supfertiles. Le prédicteur a identifié 11 taureaux subfertiles sur les 13 et 16 supfertiles sur les 17. La précision du modèle atteint 90%, la sensibilité 85% et la spécificité 94%. Les tables de confusions pour les dispositifs d'apprentissage et de test sont présentées dans la figure 43.

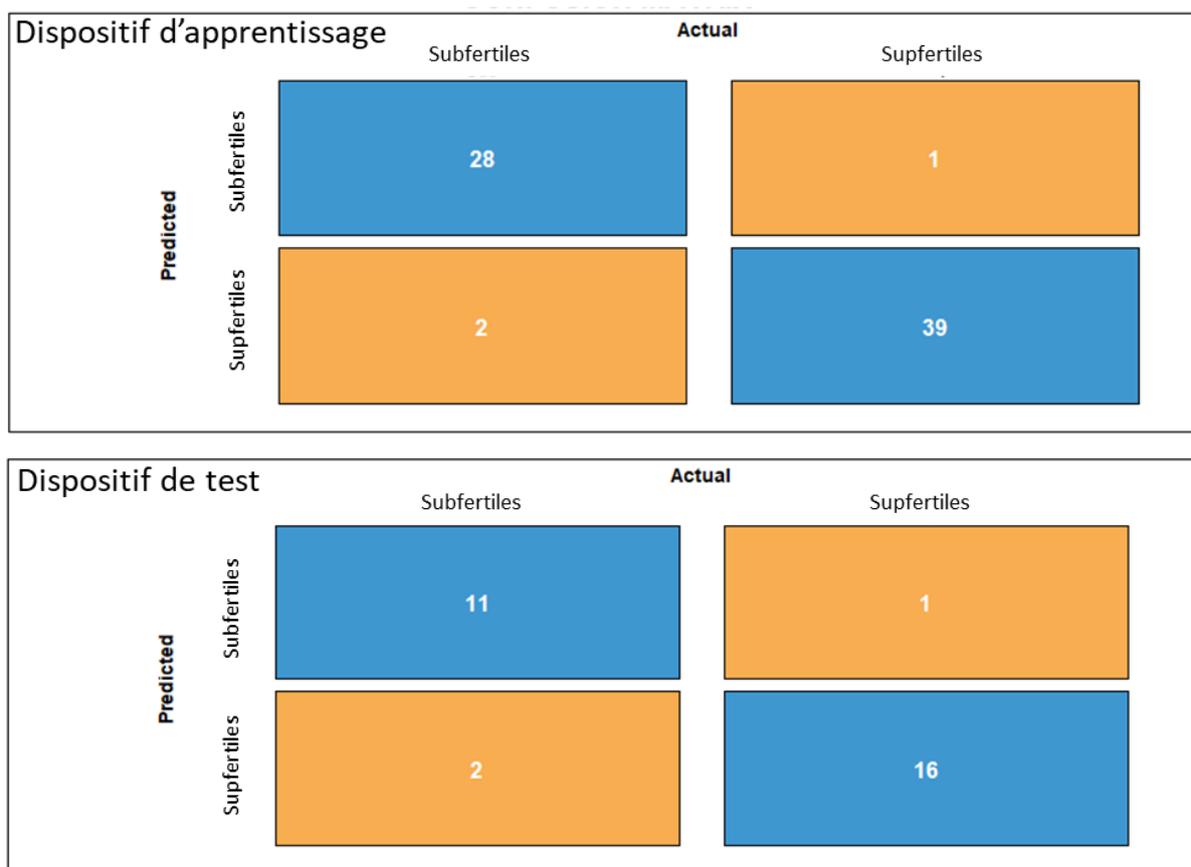


Figure 43 : Matrices de confusions du méta-classifieur.

Les deux approches SVM et Random Forest ont été combinées au sein d'un « meta-classifieur » afin d'améliorer la force de prédiction de fertilité. La matrice de confusion du dispositif d'apprentissage permet de se rendre compte de l'adéquation du modèle au jeu de données (70% de la population des taureaux Montbéliard). La matrice de confusion du dispositif Test, permet d'apprécier l'efficacité de prédiction du modèle (30% de la population des taureaux Montbéliard). La précision du modèle atteint 90%, sa sensibilité 85% et sa spécificité 94%.

### 3.5) Relation entre sncRNA et la qualité de la semence

La corrélation entre niveau d'expression de chaque sncRNA et valeur de chaque paramètre phénotypique a été calculée, produisant 220 corrélations significatives ( $r \geq |0.7|$ ), exclusivement en race Holstein, dont 157 positives et 63 négatives. À l'exception d'une corrélation entre la viabilité et un rsRNA, toutes ces corrélations concernent un paramètre en lien avec le niveau de production énergétique de la cellule (MYH, moyenne d'intensité de fluorescence des spermatozoïdes ayant des mitochondries actives).

Les miRNA et piRNA sont majoritairement représentés parmi les séquences statistiquement corrélées à ce phénotype (31% et 29% respectivement, les autres sncRNA avoisinant les 15%

chacun). Plus de 80% des isomiR corrélés au MYH sont acquis post-éjaculation. Près de la moitié des piRNA statistiquement liés à MYH ont une expression spécifiquement testiculaire (44%). Un tiers est acquis post-éjaculation et environ 25% présentent une expression épидидymaire. La moitié des rsRNA (49%) est également acquis post-éjaculation. Une part importante d'entre eux (35%) sont acquis au cours du transit épидидymaire. Les rsRNA sont principalement représentés par des fragments de 16S et 18S (figure 44). En revanche, les tsRNA présentent plutôt un profil d'acquisition en cours de transit (46% des sncRNA) et sont principalement représentés par les 5'-tRH et i-tRF. Il est intéressant de noter que les tRF représentant l'autre moitié du tRNA fragmenté (3'-tRF) sont quasiment absents des sncRNA liés à MYH. Enfin 4 iso accepteurs sont représentés de manière équivalente parmi ces sncRNA.

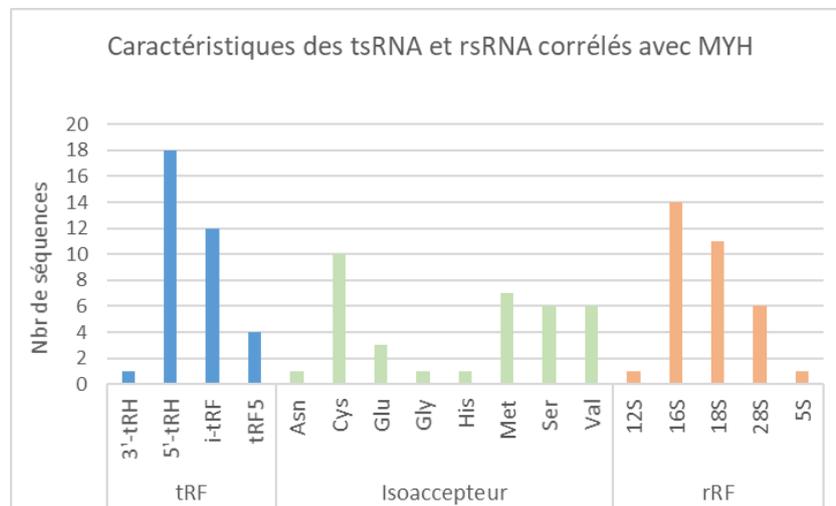


Figure 44 : Caractéristiques des tsRNA et rsRNA corrélés avec MYH.

Parmi les tsRNA corrélés à MYH, les sncRNA les plus nombreux sont les 5'-tRH et i-tRF. Quatre iso-accepteurs sont représentés de manière équivalente (Cys, Met, Ser et Val). Concernant les rsRNA, les sncRNA les plus fréquents sont ceux dérivant du 16S et 18S.

De par les cibles de cinq miRNA principalement présents (miR-93, miR-100, miR-135, miR-191 et miR-425), le processus biologique KEGG « mTOR » est statistiquement enrichi chez le bovin et chez l'humain. Les gènes ciblés de cette voie biologique sont *PDPK1*, *PIK3R2*, *MAPK1*, *STK11*, *ULK1*, *AKT1*, *MTOR*, *BRAF*, *RPS6KA3* et *EIF4EBP1* (figure 45).

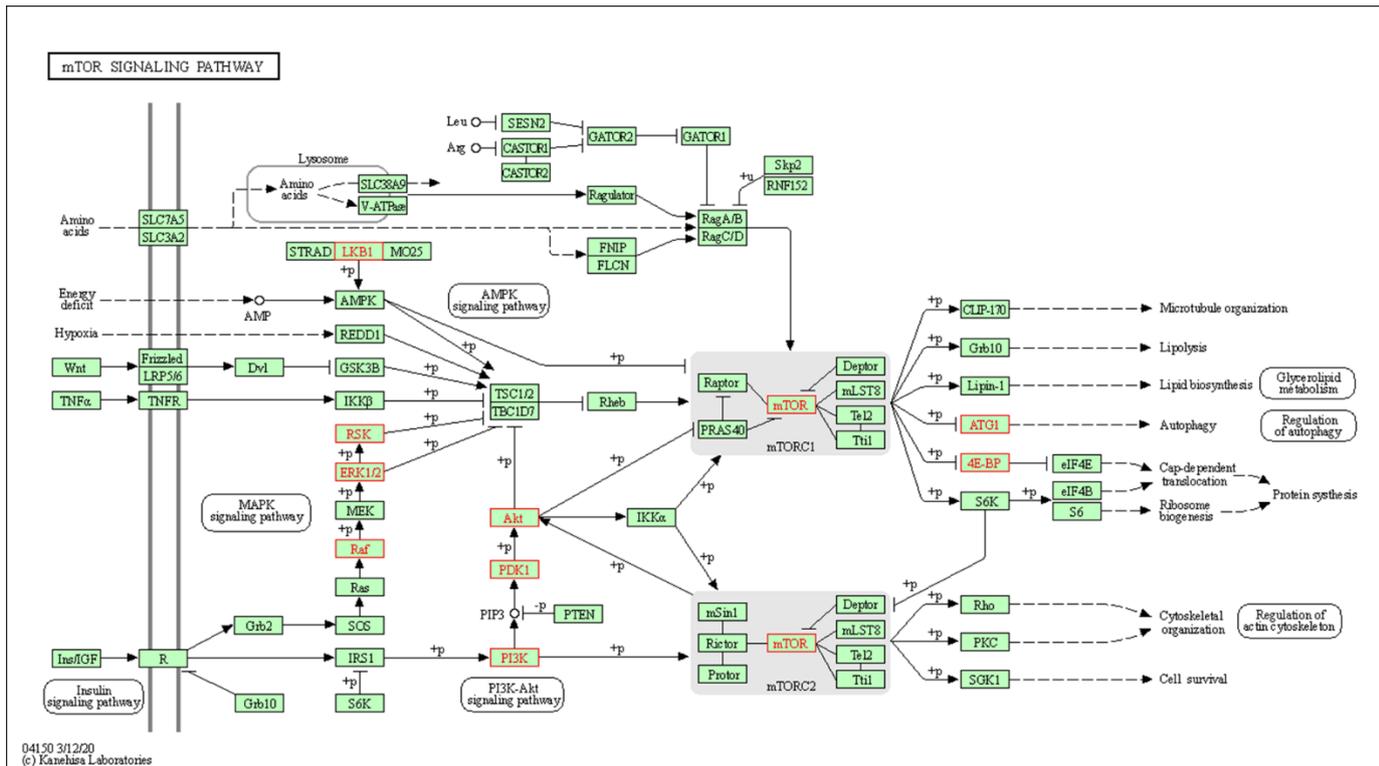


Figure 45 : Voie de signalisation du processus biologique mTOR. Les gènes identifiés en rouge sont ciblés par les miRNA les plus représentés parmi les séquences statistiquement corrélées à la production énergétique des mitochondries (MYH).

## 1) Discussion

### 4.1) La construction du dispositif expérimental

Le dispositif de taureaux a été construit afin d'éviter les principales sources de biais ou d'erreurs identifiées *a priori*. L'indicateur de fertilité TNR56 corrigé est issu de l'évaluation femelle et prend en compte toutes les inséminations réalisées ainsi que les effets année, inséminateur et niveau génétique de la femelle. La plage d'années de naissance des taureaux a été limitée à la période 2011 et 2015, afin d'éviter de trop grandes variations de niveaux génétiques entre les taureaux en raison du progrès génétique annuel induit par la sélection, ainsi que d'éventuelles modifications de profils épigénétiques liées à des modifications environnementales. L'âge de production des éjaculats utilisés a également été contraint entre 17 et 19 mois, afin de limiter l'impact d'éventuels changements de profils épigénétiques avec

l'âge. Sur ce dernier point, nous avons obtenu en 2019 des résultats indiquant qu'il n'y avait pas de variation significative sur une période allant de 13 à 19 mois d'âge (données non encore publiées). Résultats confortés également par des travaux récents portant sur les profils en miRNA spermatiques de 4 taureaux Holstein prélevés à 10, 12 et 16 mois<sup>247</sup>.

Le choix d'établir les groupes sur la base du TNR56 avait pour objectif de cibler les effets en lien avec le pouvoir fécondant et le développement de l'embryon, sachant que le TNR28 est entaché d'erreurs et manque de précision. Le TNR56 est également pertinent, car il est mesuré après la fin de l'organogenèse chez le bovin (42 jours post-fécondation). Toutefois, dans l'hypothèse où les sncRNA spermatiques peuvent impacter le développement post-implantatoire de l'embryon et du fœtus, il pourrait être intéressant d'analyser également des groupes constitués sur la base des TNR90 et TNR282 (mise bas).

#### 4.2) Les sncRNA biomarqueurs de qualité de semence

Quelques centaines de sncRNA d'intérêt ( $p$ -value ajustée  $\leq 0.1$ ) ont été mis en évidence, dont 196 significativement différentiellement exprimées entre les groupes de fertilité ( $p$ -value ajustée  $\leq 0.05$ ). Ces potentiels biomarqueurs semblent spécifiques de la race, ce qui peut témoigner d'une différence génétique (mutation dans la race impactant directement ou indirectement le niveau d'expression de ces sncRNA) ou d'un effet épigénétique lié à l'environnement (les taureaux sont hébergés dans des centres et des régions différentes). De façon surprenante, alors que l'écart de performance de fertilité entre les taureaux superfertiles et subfertiles est similaire pour les deux races, l'analyse différentielle identifie près de 7 fois plus de biomarqueurs en race Montbéliarde. La baisse de puissance statistique liée aux effectifs plus faibles et déséquilibrés en race Holstein contribue peut-être à ce résultat. Toutefois, des différences biologiques peuvent également être à l'origine de cette observation. En effet, la race Holstein présente un niveau de fertilité globalement plus faible que la race Montbéliarde (près de 15 points de TNR d'écart) en lien avec un niveau génétique plus faible des animaux. Il est ainsi possible que les groupes de fertilité traduisent principalement une différence génétique en race Holstein, alors que la part d'épigénétique pourrait être plus importante en race Montbéliarde. Une telle hypothèse pourrait aussi

expliquer le peu de recouvrement des jeux de biomarqueurs entre les deux races. De même, les différences observées en matière de classes (prépondérance des trsRNA en Montbéliarde *versus* proportion équilibrée de rsRNA, miRNA et piRNA en Holstein), de type de tsRNA (Gly et Met 5'-tRH et i-tRF en Montbéliarde *versus* Ser i-tRF en Holstein), de rsRNA (28S en Montbéliarde *versus* 18S en Holstein) et d'origine (15% *versus* 41% acquis au niveau de la queue de l'épididyme et 65% *versus* 33% acquis au niveau prostatique en race Montbéliarde *versus* Holstein) suggèrent également des mécanismes différents.

De multiples travaux ont mis en évidence l'intérêt des sncRNA, principalement les miRNA, comme biomarqueur pour l'évaluation de la qualité du sperme chez l'homme, le porc ou le bovin<sup>237,248-251</sup>. Nos résultats concernant les miRNA en lien avec la fertilité et le niveau de production énergétique sont globalement cohérents avec les données publiées. En effet, la plupart ont été identifiés comme différentiellement exprimés chez des patients infertiles, présentant ou non des altérations des spermatozoïdes. Par exemple, une surexpression de miR-192 et de miR-146a, les 2 miRNA sous-exprimés chez les taureaux Holstein supfertiles, a été observée chez des patients atteints d'azoospermie sécrétoire (NOA) et asthenozoospermiques<sup>252</sup> respectivement. De même, miR-30d différentiellement exprimé chez les taureaux supfertiles Montbéliards est sur-exprimé chez des patients normozoospermique infertiles<sup>253</sup> et asthenozoospermiques<sup>250</sup>. La surexpression de miR-2284 observée chez les taureaux Montbéliard supfertiles a également été observée chez des taureaux allaitants nord américains<sup>254</sup> et une surexpression de miR-2483 a été associée à la fertilité (TNR56) de la semence sexée<sup>255</sup> alors que l'expression de bta-miR-6526 a été associé à la motilité des spermatozoïdes<sup>237</sup>. Concernant les 18 miRNA connus corrélés avec le niveau de production énergétique, 17 ont été associés à des altérations de paramètres spermatiques ou de fertilité : miR-30d, miR-93, miR-132-5p sont surexprimés chez des patients normozoospermique infertiles<sup>253</sup> ; miR-30a-5p, miR-30d, miR-30e-5p sont surexprimés chez des patients asthenozoospermiques<sup>250</sup> ; let-7b, miR-30d, miR-30e-5p, miR-375 sont surexprimés alors que miR-34c-5p et miR-425 sont sous-exprimés chez des patients oligoasthenozoospermiques<sup>250</sup> ; let-7b et miR-375 sont surexprimés chez des patients oligozoospermiques alors que miR-34b et miR-122 sont sous-exprimés<sup>256</sup> ; miR-135a est sous-exprimé chez des patients infertiles<sup>257</sup> ; miR-34c est surexprimé chez les taureaux

subfertiles<sup>251</sup> et lié à la fertilité de la semence sexée<sup>255</sup>; bta-miR-93 et bta-miR-106b ont été associés à une bonne motilité des spermatozoïdes, alors que bta-miR-100 et bta-miR-2285n sont exprimés dans les spermatozoïdes peu motiles<sup>237</sup>; miR-132-3p, miR-191-5p et miR-191-3p sont surexprimés alors que miR-101-3p est sous-exprimé dans des semences produisant un faible taux d'embryon de bonne qualité par FIV<sup>248,258</sup>.

Concernant les autres classes de sncRNA, une étude récente a mis en évidence des tsRNA et des rsRNA spermatiques exprimés différemment entre deux groupes de semence produisant un fort ou un faible taux d'embryon FIV de bonne qualité<sup>248</sup>. En particulier, 10 tsRNA ont été identifiés correspondant à des isoaccepteurs Gly, Thr et Glu (sous-exprimés) et Pro, Asn et Arg (surexprimés), principalement des tRF5 (40%), des 5'-tRH (30%) et des 3'-tRH (20%). Les tsRNA associés à la fertilité des taureaux Holstein ne présentent pas des caractéristiques similaires puisque composés principalement d'i-tRF et d'isoaccepteurs Ser et Asp. Les tsRNA corrélés au niveau de production énergétique ne présentent pas non plus des caractéristiques similaires, hormis une proportion élevée de 5'-tRH. Les caractéristiques des tsRNA en race Montbéliarde sont plus proches, puisque comportant une grande proportion de 5'-tRH et d'isoaccepteurs Gly. Les différences observées peuvent résulter de différences d'espèces, mais aussi des différences de phénotype, la fertilité ne se résumant pas à la seule qualité des embryons...

De même 7 rsRNA différemment exprimés ont été identifiés, dérivant d'ARNr 28S (85%) ou ARNr 18S (15%), ce qui est cohérent avec nos résultats en race Holstein, mais pas en race Montbéliarde où les fragments d'ARNr 18S sont majoritaires (45%). Les rsRNA corrélés au niveau de production énergétique ne présentent pas les mêmes caractéristiques, avec une majorité de fragments d'ARNr 16S et 18S.

Les biomarqueurs de fertilité identifiés peuvent témoigner d'une altération de la spermatogenèse, d'un défaut de maturation affectant la fonctionnalité de la semence et sa fécondance ou le développement ultérieur de l'embryon. Or, l'essentiel des biomarqueurs identifiés est acquis par le spermatozoïde dans la queue de l'épididyme ou ultérieurement. Seuls 12% des biomarqueurs identifiés sont d'origine testiculaire, représentant 13 piRNA, 3 tsRNA et 46 rsRNA dont les fonctions au cours de la spermatogenèse ne sont pas connues. Cela suggère que, contrairement aux situations pathologiques d'infertilité décrites chez l'Homme, la subfertilité des taureaux du dispositif ne résulterait que peu de défaut de

spermatogenèse, ce qui est également cohérent avec la sélection opérée par les centres, l'absence d'altération des paramètres spermatiques et la quasi-absence de corrélation entre sncRNA et paramètres spermatiques dans notre cohorte. Les seules corrélations observées l'ont été avec le niveau de production énergétique MYH, 75% des biomarqueurs corrélés étant acquis à partir de la queue de l'épididyme. Parmi les 25 miRNA corrélés, 10 sont associés avec le fonctionnement de la mitochondrie : pre-let-7b, miR-34, miR-93, miR-106, miR-132<sup>259</sup>. De même, une expression différentielle (Fold Change < -3) de miR-30a, miR-30d, miR-30e, miR-101, miR-191 a été observée dans les mitochondries cardiaques de patients atteints de diabète de type II<sup>260</sup>. De plus, une analyse des pathways KEGG à partir des cibles de ces miRNA met en lumière la voie de signalisation mTOR, dont l'implication dans la régulation du fonctionnement de la mitochondrie a été montré<sup>261</sup>, en lien avec la glycolyse, la synthèse des lipides et le processus d'autophagie. En particulier, le complexe mTORC1 est inhibé dans les spermatozoïdes motiles de donneurs âgés, ce qui conduirait à l'activation de la cascade conduisant à l'autophagie<sup>262</sup>. La destruction des protéines ou des organelles endommagées telles que les mitochondries protégerait la cellule de dommages plus importants induits par ces défauts (accumulation de protéines oxydées, source de génération de ROS, conduisant à la dégradation de structures cellulaires). D'autres travaux publiés sur l'expression de mTOR dans la semence de patients infertiles corroborent ces éléments<sup>263</sup>. Ainsi, les miRNA acquis au cours de la maturation des spermatozoïdes pourraient contribuer à réguler l'activité énergétique du spermatozoïde. Il est intéressant de constater que parmi les miRNA associés à la fertilité, miR-146a est également connu pour jouer un rôle dans le métabolisme énergétique en favorisant le métabolisme oxydatif des lipides, en augmentant le nombre de mitochondries et en favorisant la respiration mitochondriale dans les hépatocytes<sup>264,265</sup>. Il intervient également dans la modulation de l'apoptose et la régulation de l'autophagie<sup>266</sup>. La production énergétique du spermatozoïde est généralement associée au fonctionnement du flagelle et à la motilité. Ainsi, l'augmentation de l'expression de mTOR est inversement corrélée au pourcentage de spermatozoïdes progressifs. De même, l'activation de l'autophagie entraîne une augmentation de la motilité des spermatozoïdes<sup>267</sup>. Or, nous ne mettons pas en évidence de corrélation avec la motilité. Dans ce cadre, le lien avec la fertilité pourrait plutôt résulter de l'effet anti-apoptotique de miR-146a et, via son action sur l'autophagie, d'un mécanisme pro-survie permettant une meilleure conservation de la semence en attente d'éjaculation, à l'instar de ce qui a été observé en cours de stockage des

spermatozoïdes chez l'étalon<sup>268</sup>. Une autre hypothèse serait que la régulation mise en place n'aurait pas pour objectif d'assurer la fonctionnalité immédiate du spermatozoïde, mais de le préparer aux étapes ultérieures, par exemple la capacitation et l'hyperactivation ayant lieu dans les voies femelles.

L'expression de miR-146a dans le spermatozoïde est le reflet de sa concentration dans le fluide épидидymaire et le plasma séminal. Son action sur la fertilité pourrait ne pas être en lien avec la physiologie spermatique, mais résulter d'une action positive sur l'implantation de l'embryon. En effet, l'implication de miR-146a a été proposée dans la modulation des macrophages et de l'activité des cellules dendritiques du tractus génital féminin. Cela pourrait contribuer à induire une immunotolérance et permettre l'implantation de l'embryon, comme suggéré par les dérégulations de ce miARN observées dans les fausses couches et la pré-éclampsie<sup>269</sup>.

Les biomarqueurs de fertilité peuvent aussi impacter le développement de l'embryon. Par exemple, un rôle de tRF-Gly-GCC a été montré dans la répression du rétrotransposon MERVL au stade zygote, impactant le développement ultérieur de l'embryon<sup>203</sup>. De même, les miRNA maternels étant des régulateurs essentiels de la transition zygotique et des premières phases du développement embryonnaire<sup>270,271</sup>, l'apport de miRNA spermatiques peut interférer avec leur action. Parmi les biomarqueurs identifiés, miR-6526 est présent dans l'ovocyte bovin et son niveau d'expression triple dans l'embryon 8 cellules<sup>272</sup>. De même miR-30c et miR-146a sont exprimés dans l'embryon 2 cellules et leur expression augmente dans l'embryon 4 cellules<sup>273</sup>. Par ailleurs, miR-192, miR-146a et miR-30d sont produits par l'embryon bovin 8 cellules et excrétés dans le milieu de culture<sup>274</sup> et une surexpression de miR-2284 a été détectée dans le milieu de culture des embryons de meilleure qualité<sup>275</sup>. Une modification de l'expression de miR-30d a été observée dans des embryons de Zebra Fish, dont le développement s'est arrêté<sup>276</sup>. Le bta-miR-146 semble avoir une place particulière. Une partie des gènes ciblés par ce miRNA (*STAT3*, *SMARCA5*, *BMPR1A*, *BAZ2A*, *SMAD1*, *NANOG*, *PRDM1*...) sont retrouvés dans des processus biologiques liés au développement de l'organisme (morphogénèse cellulaire ou tissulaire, différenciation épithéliale et développement des cellules mésenchymateuses). Cette dernière voie biologique prend place lors de la gastrulation. Cette étape du développement embryonnaire (18 jours post-fécondation chez le bovin) voit la formation des trois feuilletts cellulaires (ectoderme,

endoderme et mésoderme), dérivant de l'épiblaste et initialement des cellules de la masse interne (ICM ; *inner cell mass*)<sup>277</sup>. Ces trois différentes lignées donneront, en se différenciant, toutes les structures de l'organisme (peau, organes...). Au vu des processus biologiques ciblés par ce miRNA, une perturbation de la régulation de ces gènes pourrait engendrer des problèmes de structuration et/ou de fonctionnement des organes. Cette hypothèse est vérifiée par le rôle du gène *BMPRI1A* dans la formation du cœur. Ciblé par le miR-146 et faisant partie de la voie de développement des cellules mésenchymateuses, sa délétion chez la souris entraîne un arrêt du développement embryonnaire entre les stades E11.5 et E12. Son absence empêche la transition des cellules du mésenchyme vers des cellules endothéliales de l'endocarde<sup>278</sup>. L'organogenèse prenant fin au 42<sup>e</sup> jour de la gestation chez les bovins, le TNR56 choisi pour la création des groupes de fertilité pourrait donc être en mesure de capter une partie de ces arrêts de gestation (contrairement au TNR28).

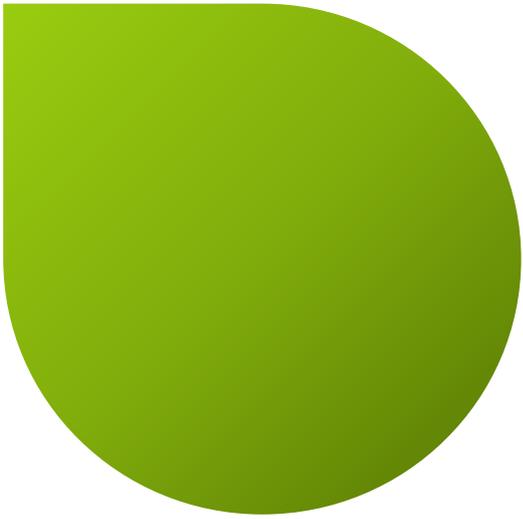
#### 4.3) La prédiction de la fertilité mâle par l'expression des sncRNA spermatique ?

Les approches de « machine learning » permettent de combiner efficacement différents phénotypes au sein d'un régresseur. En revanche, leur puissance de calcul mène parfois à des problèmes de surapprentissage du modèle : le régresseur prédit très bien le dispositif d'apprentissage, mais aucunement celui de test.

Dû au déséquilibre de population trop important entre les taureaux Holstein subfertiles et superfertiles, ils n'ont pas été intégrés dans cette approche. Le cumul de nouveaux taureaux dans le dispositif permettra de pallier ce manque. C'est pourquoi seuls les taureaux Montbéliards ont été utilisés.

Le nombre de variables a été réduit en sélectionnant uniquement les plus informatives parmi celles détectées par DeSeq2 comme différentiellement exprimées selon la fertilité. Cette réduction de variables bien que nécessaire, limite la découverte de nouveaux biomarqueurs potentiels. Il serait intéressant de refaire ces analyses avec un autre jeu de petits ARN non codants, choisis cette fois-ci parmi les plus exprimés, les plus variables ou par famille d'ARN. La combinaison des séquences retenues par VSURF pourrait constituer une table plus informative.

La combinaison des 183 petits ANR non codants sélectionnés permet de prédire le groupe de fertilité des 30 taureaux Montbéliard du dispositif de test de manière satisfaisante. En effet, le méta-classifieur atteint un score de précision de 89% lorsqu'il prédit un taureau comme « supfertile » (16 bonnes prédictions sur 18) et il permet d'identifier 94% des taureaux réellement supfertiles du dispositif (16 sur 17). Concernant la prédiction des taureaux subfertiles, la précision atteinte par le modèle est de 92% (11 bonnes prédictions sur 12) et 85% de l'ensemble des taureaux subfertiles du dispositif sont identifiés (11 sur 13). Ces résultats très prometteurs doivent être confirmés sur un dispositif plus conséquent. La question de l'identification des taureaux à fertilité « normal » devra également être investiguée. Le prédicteur obtenu sur un dispositif de taureaux extrêmes en fertilité pourrait éventuellement ne pas convenir face à un dispositif représentatif de la population annuelle des taureaux.



## Discussion globale et perspectives

## 1) Les sncRNA sont-ils de bons biomarqueurs de la fertilité mâle ?

L'utilisation de taureaux ou d'éjaculats subfertiles diminue l'efficacité globale des schémas de sélection et engendre des pertes économiques pour la filière. Depuis plusieurs décennies, la recherche d'un prédicteur de la fertilité mâle occupe différentes équipes de recherches académiques ou privées à travers le monde. Basés initialement sur le dogme de la qualité « morpho-fonctionnelle » de la semence, ces travaux n'ont pas permis d'expliquer suffisamment les variations de fertilité observées. Le LabCom SeQuaMol, fondé par INRAE/BREED/MECP<sup>2</sup> et Alice, a ainsi ambition d'explorer d'autres approches et paramètres, en particulier l'épigénome du spermatozoïde. Le programme de recherche novateur et ambitieux élaboré en son sein vise à étudier les différentes marques épigénétiques du gamète mâle : la méthylation de l'ADN, les modifications post-traductionnelles des histones et protamines et le profil de petits ARN non codants (sncRNA).

Les travaux développés dans cette thèse se focalisent sur les sncRNA du spermatozoïde éjaculé, qui sont délivrés à l'ovocyte au moment de la fécondation. Longtemps considérés comme des reliquats de la spermatogenèse, où ils jouent un rôle essentiel, ces éléments de régulation de l'expression génique peuvent participer à la construction du phénotype de l'embryon, du fœtus et *in fine* de la descendance. Des variations d'expression ou de diversité en sncRNA spermatiques peuvent donc potentiellement avoir un retentissement sur une variété de caractères d'intérêt, dont la fertilité, la santé, la production...

La question se pose donc de savoir en quelle mesure les sncRNA spermatiques peuvent être de bons biomarqueurs de la fertilité mâle, qu'ils soient un vestige du passé témoignant de la qualité du processus de spermatogenèse, ou des effecteurs contribuant un bon développement de la descendance. Une stratégie en entonnoir, comportant 3 actions a été établie afin de répondre à cette question.

La première étape du projet fut d'établir le catalogue le plus exhaustif possible des sncRNA du spermatozoïde bovin mature éjaculé. Les résultats obtenus ont mis en lumière les classes de sncRNA contenues dans le spermatozoïde (principalement miRNA, piRNA, tsRNA et rsRNA), les spécificités d'expression et l'importance du mécanisme d'édition contribuant à générer une grande diversité de séquences d'isomiR, d'isopiR et autres « isoARN ». Ces résultats,

prérequis pour une recherche de biomarqueurs, ont également mis en évidence des spécificités, comme la prépondérance des rsRNA 18S/28S et des fragments 5'-tRHs issus d'isoaccepteurs Gly/Glu. Aux antipodes d'une vision de « dégradome », qui a conduit de nombreuses équipes à éliminer ses séquences de leurs analyses, nos résultats suggèrent un rôle fonctionnel pour ces fragments, peut-être par une activité « miRNA-like » via leur interaction avec AGO2<sup>279</sup>. Nous avons également mis en évidence des biais d'édition qui mériteraient d'être explorés plus avant afin d'élucider les mécanismes sous-jacents, comme le biais d'édition des piRNA en faveur d'ajouts/délétions de nucléotide en 3'. De même, la fréquence élevée d'édition des miRNA au niveau du 1<sup>er</sup> nucléotide de la région « seed » interroge à la fois sur la définition actuelle de cette région, censée être conservée, et sur les approches bioinformatiques considérant comme isomiR des séquences ciblant potentiellement un répertoire de gènes différents.

La seconde étape a permis d'identifier l'origine des sncRNA portés par le spermatozoïde éjaculé. L'objectif était de savoir si les sncRNA du spermatozoïde éjaculé représentaient des vestiges de la spermatogenèse ou étaient acquis après leur maturation testiculaire et potentiellement impliqués dans l'acquisition de la fonctionnalité des spermatozoïdes au cours de la maturation ou dans le développement embryonnaire. Le dispositif expérimental a permis de répondre à la question : L'analyse de la diversité et des profils d'expression spatio-temporels des sncRNA du spermatozoïde testiculaire, épидидymaire (Tête, Corps et Queue) et éjaculé a révélé des tendances générales spécifiques de chaque classe, mais également des profils spécifiques de certains sncRNA, allant parfois à l'encontre de la tendance générale, comme certains piRNA dont l'expression post-épидидymaire augmente, alors que la tendance générale est à l'extinction de l'expression dès l'entrée du spermatozoïde dans la tête de l'épидидyme, avec une réduction de plus de 90% de leur expression, au profit des miRNA, tsRNA et rsRNA. Le rôle de ces sncRNA, acquis durant le transit épидидymaire ou ultérieurement, est encore mystérieux et la compréhension des mécanismes impliqués nécessite encore une recherche intense. En effet, les spermatozoïdes étant considérés comme silencieux au plan transcriptionnel et traductionnel, les possibilités de régulation de la maturation post-testiculaire par les sncRNA semblent réduites. Et pourtant, le KO conditionnel de DICER dans les cellules de l'épидидyme induit une dérégulation de la maturation et impacte

la fertilité, sans que l'on puisse actuellement distinguer les effets liés à la perturbation du fonctionnement de l'épididyme lui-même et de ses sécrétions protéiques de ceux liés à la perturbation du contenu en sncRNA des épидидymosomes et du spermatozoïde. Une action pourrait toutefois s'envisager sur le pool de transcrits hérités de la spermatogenèse, à supposer qu'il existe une activité traductionnelle dans le spermatozoïde post-testiculaire, similaire par exemple à celle existant au cours de la phase de capacitation et qui s'appuie sur la machinerie de traduction mitochondriale<sup>280</sup>. Dans ce contexte, il est intéressant de noter que nos travaux montrent un enrichissement de voies liées à la régulation de la fonction mitochondriale et du métabolisme énergétique, associées aux cibles des sncRNA dont l'expression culmine dans le spermatozoïde épидидymaire.

Un rôle dans la maturation semblant marginal, voire peu probable, il est tentant de spéculer sur un rôle dans la transition zygotique ou le développement embryonnaire. Dans cet ordre d'idée, les travaux de Conine et al., publiés en 2018<sup>227</sup>, démontrent clairement la nécessité du legs spermatique non codant pour le développement embryonnaire murin normal. En effet, sans l'apport de sncRNA acquis en fin de transit épидидymaire, le développement de l'embryon s'arrête entre le stade blastocyste et l'implantation, montrant des effets induits à moyen/long terme. Des travaux ultérieurs seront nécessaires pour tester cette hypothèse et identifier les classes de sncRNA responsables de cet effet. Quoi qu'il en soit, toutes les séquences identifiées acquises en fin de transit épидидymaire par le spermatozoïde représentent des biomarqueurs candidats prometteurs, potentiellement en lien avec la fécondance et le développement embryonnaire.

Les derniers travaux de la thèse se sont appuyés sur un dispositif conséquent de taureaux de fertilité contrastée. L'analyse de leur profil en sncRNA a révélé des corrélations entre l'expression de plusieurs séquences et le métabolisme énergétique de la mitochondrie, en particulier le processus d'autophagie. La machinerie de l'autophagie est présente dans les spermatozoïdes et est essentielle non seulement au cours de la spermiogenèse, mais également au cours de la maturation, ou une forme particulière d'autophagie appelée mitophagie régulerait la motilité et la viabilité des spermatozoïdes, en coopération avec les mécanismes d'apoptose<sup>281</sup>. Ces résultats font ainsi potentiellement écho aux observations et hypothèses exposées précédemment. D'autres hypothèses peuvent être avancées

concernant ce lien, car le pouvoir fécondant du sperme est directement lié à l'activité mitochondriale. En effet, l'incubation du sperme de patients infertiles avec de l'ATP améliore le taux de fécondation *in vitro*<sup>282</sup>, vraisemblablement en lien avec une meilleure capacitation et une activité flagellaire accrue. Toutefois, nous n'avons pas mis en évidence de corrélation avec les paramètres de mobilité. Sachant que les spermatozoïdes stockés dans les trompes de Fallope connaissent une phase de capacitation et d'hyperactivation à l'approche de l'ovulation, ce qui leur permet de se diriger vers l'ampoule tubaire, il est tentant de faire l'hypothèse que les modifications de métabolisme induites par les sncRNA ne se traduisent pas complètement au moment de l'éjaculation, mais permettent de potentialiser l'effet de signaux ultérieurs émis par la femelle, par exemple au moment de l'ovulation. Mais au-delà de ces relations, plusieurs centaines de sncRNA sont différenciellement exprimés selon le groupe de fertilité (dont la moitié avaient déjà été identifiés comme candidats prometteurs par les travaux sur l'origine des ARN), mettant en lumière l'impact des sncRNA sur la qualité de la semence et leur intérêt comme biomarqueurs de contrôle qualité. Dans cet ordre d'idée, il est à noter que la dilution et la cryoconservation de la semence altèrent le contenu en sncRNA (et ARNm) du spermatozoïde de taureau<sup>283</sup>. À la lueur de nos résultats, des travaux de recherche semblent nécessaires pour évaluer en quelle mesure et par quels mécanismes cette altération pourrait expliquer la réduction de fertilité observée en IA par rapport à la monte naturelle. Par ailleurs, le profil en sncRNA pourrait constituer un marqueur supplémentaire à prendre en compte pour améliorer les technologies de cryoconservation de la semence.

L'association de ces biomarqueurs par un algorithme « d'ensemble learning » a permis de mettre en lumière un prédicteur efficace de la fertilité des taureaux Montbéliards. Bien que le dispositif de taureaux utilisé soit conséquent, il sera nécessaire de confirmer ces résultats de prédiction sur un nombre de mâles plus important, présentant des caractéristiques de fertilité moins contrastées. Il pourrait également être intéressant d'étudier le lien entre anomalies spermatiques et contenu en sncRNA. Cette preuve de concept doit également être apportée pour les autres races du projet.

Qu'ils soient des vestiges du passé ou une promesse pour l'avenir, les sncRNA du spermatozoïde bovin représentent des biomarqueurs intéressants pour la prédiction de la fertilité mâle. En revanche, la caractérisation des taureaux par les professionnels nécessitera

le développement d'un outil « terrain » rapide, simple et peu coûteux si nous souhaitons valoriser largement ces résultats. Au-delà de la prédiction de fertilité, les connaissances et les technologies acquises au cours de cette thèse ouvrent de nouveaux champs d'investigations quant à l'analyse de l'impact environnemental sur l'épigénome du gamète mâle, sur la régulation de l'expression génique de l'embryon et finalement sur la construction du phénotype de la génération suivante.

## 2) Perspectives

### 2.1) Impact sur le développement embryonnaire

Comme discuté précédemment, l'hypothèse du rôle des sncRNA spermatiques dans le développement embryonnaire constitue une question scientifique particulièrement intrigante et passionnante. Une des perspectives de ce travail de thèse est donc naturellement de se pencher sur l'impact que pourraient avoir certains sncRNA acquis en fin de transit épидидymaire, dont les biomarqueurs en lien avec la fertilité. Pour cela, nous envisageons d'utiliser comme modèle l'embryon bovin produit par fécondation *in vitro* et de modifier expérimentalement le contenu en sncRNA au moment de la fécondation. Une telle étude nécessite une réflexion approfondie concernant divers points méthodologiques, notamment la technologie de transfert des séquences. La micro-injection dans le zygote est une technique maîtrisée au sein d'ALLICE qui peut répondre à cet objectif. Toutefois, nous souhaitons développer une alternative plus efficiente, afin d'éviter tous les effets négatifs de cette technologie sur le développement de l'embryon (technologie invasive, provoquant des arrêts de développement, nécessitant un savoir-faire particulier et modifiant l'expression de certains gènes<sup>284</sup>). D'autres approches peuvent être envisagées, telles que l'utilisation de vésicules à doubles couches lipidiques renfermant les séquences à transférer, ou par « balnéation » de ces séquences dans le milieu contenant le zygote. Cette dernière approche, très simple d'utilisation, a été utilisée pour analyser l'impact du miR-10b sur le développement embryonnaire bovin<sup>285</sup>. Les miRNA synthétiques ont été ajoutés au milieu de culture 21h post-fécondation. Bien que les mécanismes de pénétration des miRNA, passifs ou actifs, n'aient pas

été déterminés, les auteurs démontrent un effet de la séquence sur le développement de l'embryon. Peu invasive, cette technique permettrait une limitation des altérations du développement embryonnaire. Elle permettrait aussi d'envisager de mettre en présence les sncRNA synthétiques de manière concomitante à celle des spermatozoïdes dès l'étape de fécondation *in vitro*, afin d'avoir une action le plus tôt possible.

Le second point méthodologique à aborder concerne la famille de sncRNA à incorporer. Les miRNA sont historiquement les plus étudiés, leur mécanisme d'action est connu et leurs cibles peuvent être prédites, ce qui en fait de bons candidats. Mais nos données ont également illustré l'acquisition d'autres sncRNA, comme les tsRNA, rsRNA ou les piRNA, en fin de transit épидидymaire. De plus, la littérature récente illustre (i) que le complexe RISC est capable de charger diverses séquences autres que des miRNA ; (ii) une implication des tsRNA spermatiques dans la régulation des premiers clivages (tRF- Gln-TTG) et la modulation du phénotype de la descendance de pères stressés (tRF-Gly ou Glu) et iii) un impact des rsRNA spermatiques dans la qualité des embryons obtenus par FIV. L'analyse de ces classes, dont certains représentants sont associés à la fertilité dans notre étude semble de ce fait totalement justifiée.

Par ailleurs, nos données (couplées aux travaux déjà publiés) démontrent également qu'il existe une diversité importante de séquences issues d'édition des ARN (isomiR, isopiR...). Nous avons aussi montré que dans le cas des isomiR, ces variations pouvaient en particulier modifier la région « Seed » (par édition du 1<sup>er</sup> nucléotide de la Seed ou décalage de la séquence du fait d'une addition/délétion en 5') et donc potentiellement changer le répertoire de cibles. Il serait donc intéressant de vérifier si les isomiR ont un rôle fonctionnel différent de la séquence canonique du miR<sup>117</sup>.

Par ailleurs, il nous semble important de prévoir le transfert non seulement des séquences « mimics » des sncRNA, mais également les séquences antisens, de façon à pouvoir mesurer à la fois l'impact d'une sur- et d'une sous-expression.

Dans ces conditions, il serait plus intéressant de tester non seulement les sncRNA individuellement, mais aussi en « cocktail » d'iso-sncRNA représentant la diversité de séquences du sncRNA canonique d'intérêt. Différents bouquets pourraient être testés contenant des familles de sncRNA différentes.

Enfin, un protocole de phénotypage spécifique de l'embryon doit être élaboré. En effet, quelques publications montrent un effet dans les étapes de développement précoce jusqu'au stade blastocystes, mais les travaux le plus récents font état de perturbations phénotypiques apparaissant de façon plus tardive, impactant l'implantation. Les seules analyses de morphocinétiques pourraient donc être insuffisantes pour conclure. Le transfert d'embryon et le suivi de gestation étant limitants, l'analyse du transcriptome des embryons pourrait permettre de mettre en évidence des profils transcriptionnels différents, témoignant de l'impact du sncRNA.

Tout au long des travaux menés dans le cadre de la thèse, un premier candidat semble émerger naturellement des résultats obtenus. Il s'agit du bta-miR-100. Ce miRNA fait partie des séquences les plus exprimées dans le spermatozoïde bovin. C'est également l'un de ceux qui présentent le plus d'isomiR, dont l'expression est statistiquement enrichie en fin de transit épидидymaire. De plus cette séquence a déjà été décrite dans la littérature comme étant impliquée, via son impact sur ces gènes cibles et notamment *SMARCA5*, dans la régulation de la prolifération des cellules souches spermatogoniales et dans le développement normal de l'embryon murin<sup>286</sup>. La protéine *SMARCA5* est un membre de la famille SWI/SNF. Elles ont la capacité de modifier l'état de condensation de la chromatine pour rendre accessibles les gènes à la transcription. Un premier bouquet d'isomiR-100 pourrait être testé en parallèle d'un anti-miR-100 (et si possible général aux différents isomiR testés) et d'une séquence « scramble ». Afin de ne pas garder ces tâches uniquement en « perspective », un projet ANR (PRCE axe 3.3), appelé PaterLeg, a été déposé pour aborder ces questions.

## 2.2) La prédiction de fertilité : Quel outil terrain ?

Au-delà de l'acquisition de connaissances académiques, l'objectif finalisé de ce travail est le développement d'un outil terrain permettant la prédiction efficace de la fertilité mâle. Les contraintes du cahier des charges de cet outil devront permettre de trouver un équilibre entre pouvoir prédictif, facilité de mise en œuvre et coût d'utilisation.

Deux utilisations sont envisageables : soit pour le contrôle en routine des éjaculats (si l'expression et/ou la diversité des sncRNA varient sur de courtes périodes dans la vie productive de l'animal), soit pour établir le statut des taureaux, par exemple au démarrage de la fonction sexuelle (si pas ou peu de variation). Ces deux applications n'induisent pas les mêmes contraintes en termes de rapidité d'exécution et surtout de coût, qui conditionnent le nombre et la nature des marqueurs utilisés ainsi que les options technologiques possibles. En effet, des mesures répétées sur différents éjaculats d'un même mâle supposent un outil peu coûteux, mis en œuvre plutôt au niveau des centres de production pour limiter les contraintes logistiques et réduire les délais de réponse. À l'inverse, s'il suffit d'une mesure pour caractériser un taureau, le coût unitaire peut être plus élevé et l'analyse peut être confiée à un prestataire, car le délai de réponse n'est pas un facteur limitant.

Ainsi, l'outil « éjaculat » devra nécessairement faire appel à une technologie facilement maîtrisable par les laborantins des centres et ne devra être basé que sur à un set minimal de marqueurs nécessaires et suffisants pour bien discriminer la qualité de la semence. L'outil « taureau » pourrait intégrer la mesure d'un panel de biomarqueurs plus complet et faire appel à des technologies plus complexes. Une partie du dispositif animal analysé au cours de cette thèse a permis de répondre à cette question : Le profil spermatique de petits ARN non-codants d'éjaculats de taureaux Holstein produits entre 13 et 19 mois d'âge ne varie pas, sauf probablement événements sanitaires ou climatiques, points qui restent à étudier. Il s'agit de la majeure partie de la vie productive d'un taureau Holstein ou Montbéliard. Une seule analyse pourrait donc suffire pour catégoriser le taureau.

Différentes technologies peuvent être envisagées, allant de la puce à ARN au séquençage haut débit, en passant par les technologies exploitant la RTqPCR. Bien que la puce à ARN soit une technologie éprouvée<sup>287,288</sup>, elle présente un inconvénient majeur pour notre application : toutes les séquences utilisées doivent être de même taille, ce qui bloque les possibilités d'exploration pour des sncRNA de familles différentes avec le même outil. Les technologies basées sur l'exploitation de la RTqPCR sont également bien éprouvées. Il est possible de mettre en place un système de plaques (96 ou 386 puits) permettant une amplification spécifique des séquences choisies (couple d'amorces) et un suivi en temps réel de la néosynthèse. Il est aussi envisageable de multiplexer les amplifications au sein d'un même puits démultipliant ainsi le nombre de sncRNA pouvant être évalués. Couplée à des robots de

prélèvements et à des automates de passage de plaques, un grand nombre d'animaux pourrait être testé en continu. À titre d'information, la réalisation de ces analyses par une société de service coûte pour 500 taureaux à raison de 96 séquences à analyser, 83€ HT par échantillon. Le développement de la PCR digitale (ddPCR) devrait permettre de réduire ces coûts tout en simplifiant la préparation des échantillons.

Le séquençage haut débit n'est pas à première vue une technologie utilisable en routine, de par son coût (350€ par échantillon, de l'extraction d'ARN total au séquençage tel que réalisé dans cette étude). Mais son usage pourrait s'avérer compétitif à terme, à la faveur de nouvelles approches, encore en cours de développement pour des petites séquences telles que les miRNA ou autres petits ARN. Par exemple, la technologie de séquençage dirigé « UPX » développé par Qiagen permet l'analyse de 100 à 2000 séquences cibles sur une seule ligne de séquençage, augmentant le débit et réduisant le coût à l'échantillon.

Les travaux conduits dans cette thèse n'abordent que les sncRNA comme potentiels biomarqueurs de la fertilité mâle. Mais le LabCom SeQuaMol analyse en parallèle les autres marques épigénétiques portées par le spermatozoïde comme la méthylation de l'ADN et les modifications post-traductionnelles des histones et protamines. Les compositions en protéines, lipides et glycanes du spermatozoïde sont également analysées. En effet, comme les variations de fertilité entre individus étant multifactorielles, nous anticipons que la prise en compte d'un seul phénotype moléculaire ne permette pas d'atteindre une fiabilité de prédiction suffisante sur de grandes cohortes de taureaux moins contrastés que ceux utilisés pour la preuve de concept. C'est pourquoi le prédicteur de la fertilité devra très probablement combiner des types de biomarqueurs différents pour atteindre l'efficacité recherchée. Des travaux d'intégration de toutes ces différentes marques en un seul prédicteur sont en cours au sein du LabCom (Thèse CIFRE Aliche). Mais cela ajoute un niveau de complexité supplémentaire : Plus le nombre de biomarqueurs nécessaires à la prédiction augmente, plus le prédicteur est complexe à produire et plus son coût augmente. Le LabCom SeQuaMol a atteint ses objectifs « scientifiques », il faut maintenant pouvoir transformer ces informations en outil utilisable par les professionnels.

Pour la réalisation technique des analyses, trois possibilités sont envisageables. La première nécessite le recours à une société de service de type Qiagen. La seconde s'oriente plutôt vers

un transfert de savoir-faire vers une structure dont l'analyse d'échantillons biologiques est le métier (Labogena (Jouy-en-Josas) ou le LNCR ; Laboratoire National des Contrôles des Reproducteurs (Maisons-Alfort)). La troisième consiste à assurer le service au sein du LabCom SeQuaMol. Les trois solutions présentent des avantages et des inconvénients. Le nombre d'échantillons annuel et les coûts d'acquisition et de fonctionnement des technologies nécessaires détermineront le choix de la plateforme d'analyse.

La prédiction portant sur le taureau, plus cette prédiction est précoce, plus l'information est utile, évitant par exemple de recruter en centre de futurs reproducteurs subfertiles. Or, le travail réalisé ne permet qu'une prédiction tardive sur la semence, nécessitant un animal pubère. Une extension du travail au profil de sncRNA sanguins pourrait permettre de lever cette limite. Pour explorer cette voie, un nouveau dispositif expérimental est en cours de constitution, permettant le prélèvement d'échantillons de sang sur une centaine de taurillons à leur entrée en centre de production (entre 1 et 3 mois d'âge), puis à 17 mois. Le sérum contenant les sncRNA « libres » ou contenus dans des vésicules extracellulaires, ainsi que les globules blancs sont conservés pour étudier l'épigénome des cellules blanches et rechercher des biomarqueurs épigénétiques circulants, en lien avec l'épigénome spermatique et/ou la fertilité des taureaux.

### 2.3) Exploitation des vésicules extracellulaires ?

Les vésicules extracellulaires, qu'elles viennent de l'épididyme ou de la prostate, sont essentielles aux spermatozoïdes. Une extension du travail de thèse pourrait être envisagée, afin de caractériser le contenu des épидидyosomes et des prostasomes et ainsi obtenir une vision plus fine de la dynamique d'acquisition des sncRNA dans le spermatozoïde, voire d'aborder la question d'une possible activité de transcription et/ou de traduction dans le spermatozoïde.

Mais ces vecteurs d'information « somatique vers somatique » ou « somatique vers germinale » peuvent également être des outils intéressants pour délivrer un contenu spécifique aux cellules d'intérêt. Deux champs d'applications pourraient s'envisager visant l'amélioration

de la qualité de la semence congelée et l'amélioration des résultats de procréations assistées. La semence bovine est quasiment exclusivement utilisée congelée pour des raisons de logistique. Pour cela, la semence fraîchement récoltée est diluée de 10 à 30 fois (100 millions de spermatozoïdes/ml) dans un milieu de cryoconservation, ce qui a pour effet de diluer le plasma séminal.

Par ailleurs, comme évoqué précédemment, la cryopréservation de la semence altère le contenu en sncRNA du spermatozoïde. Si les hypothèses émises quant au rôle des sncRNA dans la capacitation du spermatozoïde, la modulation de l'immunité de la femelle, la fécondation et le développement de l'embryon s'avèrent exactes, on peut imaginer qu'un apport exogène de sncRNA dans les milieux de dilution pourrait permettre d'améliorer la fertilité post-décongélation. Pour cela, une production industrielle de vésicules extracellulaires contenant les sncRNA présents habituellement dans les épидидymosomes et prostasomes devrait être envisagée.

L'utilisation liée à ces travaux permettrait d'améliorer la procréation assistée. Une grande partie des infertilités masculines restent encore aujourd'hui idiopathiques. Certaines d'entre elles pourraient être liées à un défaut de diversité ou d'expression des sncRNA spermatiques. La détermination du profil spermatique de ces séquences et leur comparaison à un profil de donneur sain permettraient d'identifier certaines variations de contenu, engendrant potentiellement des perturbations du développement embryonnaire. Une adaptation spécifique des conditions de FIV ou d'ICSI par l'ajout de vésicules au contenu souhaité ou éventuellement couplée à la « balnéation » pour l'incorporation de séquences dans l'ovocyte, permettrait potentiellement de pallier ces défauts.

#### 2.4) Le spermatozoïde peut-il être une sentinelle ?

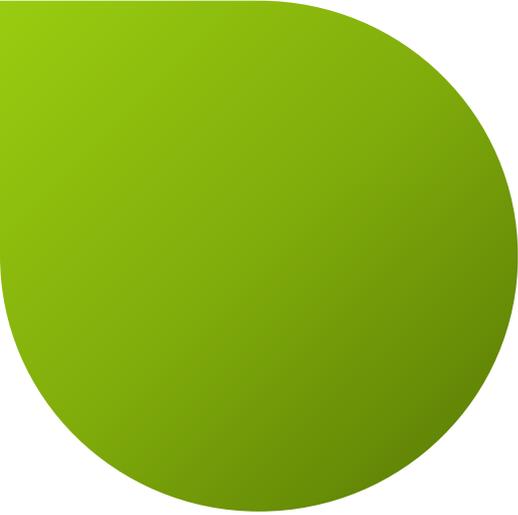
Le contenu en sncRNA spermatiques est donc représentatif non seulement des événements de néosynthèse/dégradation survenus pendant la spermatogenèse, mais est également remodelé rapidement au cours du transit épидидymaire par les conditions environnementales, en particulier le régime alimentaire<sup>289</sup>. Différentes équipes ont en effet démontré qu'un stress alimentaire (hypoprotéique ou hyperlipidique) induit des perturbations du profil en sncRNA

spermatique chez le rongeur<sup>234,235</sup>. Dans certains cas, la descendance de ces mâles présente des altérations métaboliques associées à des modifications de l'expression des sncRNA hépatiques, suggérant un passage intergénérationnel de l'information. D'autres stress environnementaux, tels que le stress thermique (vagues de chaleur couplées au taux d'humidité<sup>290</sup>), chimique (perturbateur de type Bisphénol)<sup>291,292</sup> ou électromagnétique (ondes de basses fréquences<sup>293</sup>) impactent la production, la qualité et les profils en scnRNA spermatiques. Chez l'Homme, différents facteurs extrinsèques (style de vie) et intrinsèques (obésité pathologique) ont aussi été décrits comme ayant des effets sur les profils d'expression des sncRNA dans le sperme<sup>294,295</sup>. Chez l'Homme, l'analyse de l'épigénome spermatique est parfois réalisée dans le cas d'infertilité, mais très peu dans le cadre d'un suivi post-traumatique (stress, pathologies autres que l'infertilité).

Ainsi analyser le profil de sncRNA spermatiques pourrait être une démarche systématique de suivi des effets de l'environnement sur les individus. Ce diagnostic permettrait d'établir les risques liés à différentes expositions pour l'individu lui-même et pour sa descendance. Il serait ainsi possible d'établir des recommandations de remédiation (modifications du style de vie, nutrition, limitation des expositions en intensité et/ou en durée) et déterminer le laps de temps nécessaire à un retour « à la normale » afin de limiter les préjudices pour la descendance. Mettre en place ce suivi chez le taureau permettrait donc au-delà de l'estimation de la fertilité de prendre en considération les effets environnementaux et d'y remédier par une conduite individuelle appropriée. Il est aussi envisageable d'utiliser ce suivi comme un système d'alerte des diverses expositions environnementales et d'en tirer parti pour la santé humaine : le suivi épigénétique en routine des taureaux ou des petits ruminants pourrait contribuer à alimenter un observatoire visant à détecter des variations anormales d'épigénome spermatique, qui pourraient témoigner d'une exposition des perturbations environnementales potentiellement néfastes aussi pour l'Homme, faisant des ruminants des sentinelles précoces (puberté à 12 mois) des facteurs de perturbations épigénétiques.

Depuis de nombreuses années, les technologies de reproduction assistée sont de plus en plus utilisées et innovantes, repoussant les limites de la biologie avec un objectif de projet parental chez l'Homme ou de production plus efficace chez l'animal de rente. Les gamètes mâles (mais aussi les gamètes femelles) sont manipulés, soumis à des conditions environnementales stressantes (contact important avec l'oxygène, milieu de culture, dilueur

de cryo-congélation, congélation, sexage...) pouvant modifier non seulement leur intégrité mais également leur épigénome<sup>296,297</sup>. Chez le bovin, il serait souhaitable d'évaluer tout progrès technologique à l'aulne de la qualité épigénétique de la semence (test de nouveaux cryo conservateurs, sexage de la semence...). La création d'un observatoire regroupant sur plusieurs générations informations phénotypique et épigénétique pourrait permettre de rechercher d'éventuelles conséquences sur les performances à court (chez l'individu) et à long terme (chez la descendance) de perturbations environnementales variées. Au-delà des connaissances académiques apportées par nos données et d'applications dans le développement d'outil diagnostique, le modèle animal bovin offre la possibilité d'explorer plus avant les corrélations entre modifications de l'expression des sncRNA spermatiques et altérations d'expression de gènes ou de réseaux de gènes au cours du développement embryonnaire précoce. Ainsi des études fonctionnelles de surexpression ou d'inactivation des sncRNA peuvent être envisagées, avec une observation des conséquences transcriptionnelles ou de régulations génomiques en vue d'établir les relations de cause à effet. Des études à plus long terme de la descendance et de ses performances phénotypiques sont aussi possibles offrant la possibilité d'établir les modalités de transmission intergénérationnelle d'informations non génétiques.



## Références bibliographiques

1. Spallanzani, L. & Bonnet, C. *Dissertations relative to the natural history of animals and vegetables*, (J. Murray, 1784).
2. Smith, C. & Mackay, S. Morphological development and fate of the mouse mesonephros. *J Anat* **174**, 171-84 (1991).
3. Luo, X., Ikeda, Y. & Parker, K.L. A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell* **77**, 481-90 (1994).
4. Birk, O.S. *et al.* The LIM homeobox gene Lhx9 is essential for mouse gonad formation. *Nature* **403**, 909-13 (2000).
5. Wrobel, K.H. & Suss, F. Identification and temporospatial distribution of bovine primordial germ cells prior to gonadal sexual differentiation. *Anat Embryol (Berl)* **197**, 451-67 (1998).
6. Tanaka, S.S. *et al.* The mouse homolog of Drosophila Vasa is required for the development of male germ cells. *Genes Dev* **14**, 841-53 (2000).
7. Chen, H.H. *et al.* DAZL limits pluripotency, differentiation, and apoptosis in developing primordial germ cells. *Stem Cell Reports* **3**, 892-904 (2014).
8. Gill, M.E., Hu, Y.C., Lin, Y. & Page, D.C. Licensing of gametogenesis, dependent on RNA binding protein DAZL, as a gateway to sexual differentiation of fetal germ cells. *Proc Natl Acad Sci U S A* **108**, 7443-8 (2011).
9. Archambeault, D.R. & Yao, H.H. Activin A, a product of fetal Leydig cells, is a unique paracrine regulator of Sertoli cell proliferation and fetal testis cord expansion. *Proc Natl Acad Sci U S A* **107**, 10526-31 (2010).
10. Wrobel, K.H. Prespermatogenesis and spermatogoniogenesis in the bovine testis. *Anatomy and Embryology* **202**, 209-222 (2000).
11. Barrios, F. *et al.* Opposing effects of retinoic acid and FGF9 on Nanos2 expression and meiotic entry of mouse germ cells. *J Cell Sci* **123**, 871-80 (2010).
12. Liron, J.P. *et al.* Association between GNRHR, LHR and IGF1 polymorphisms and timing of puberty in male Angus cattle. *BMC Genet* **13**, 26 (2012).
13. Curtis, S.K. & Amann, R.P. Testicular development and establishment of spermatogenesis in Holstein bulls. *J Anim Sci* **53**, 1645-57 (1981).
14. Griswold, M.D. The central role of Sertoli cells in spermatogenesis. *Semin Cell Dev Biol* **9**, 411-6 (1998).
15. Staub, C. & Johnson, L. Review: Spermatogenesis in the bull. *Animal* **12**, s27-s35 (2018).
16. de Mateo, S. & Sassone-Corsi, P. Regulation of spermatogenesis by small non-coding RNAs: role of the germ granule. *Semin Cell Dev Biol* **29**, 84-92 (2014).
17. Kotaja, N. MicroRNAs and spermatogenesis. *Fertil Steril* **101**, 1552-62 (2014).
18. Leblond, C.P. & Clermont, Y. Definition of the stages of the cycle of the seminiferous epithelium in the rat. *Ann N Y Acad Sci* **55**, 548-73 (1952).
19. Barth, A.D. & Oko, R. Abnormal morphology of bovine spermatozoa. (1989).
20. Krawetz, S.A., Connor, W. & Dixon, G.H. Cloning of bovine P1 protamine cDNA and the evolution of vertebrate P1 protamines. *DNA* **6**, 47-57 (1987).
21. Dacheux, J.L. *et al.* The contribution of proteomics to understanding epididymal maturation of mammalian spermatozoa. *Syst Biol Reprod Med* **58**, 197-210 (2012).
22. Almquist, J.O., Branas, R.J. & Barber, K.A. Postpuberal changes in semen production of Charolais bulls ejaculated at high frequency and the relation between testicular measurements and sperm output. *J Anim Sci* **42**, 670-6 (1976).
23. Lunstra, D.D. & Echternkamp, S.E. Puberty in beef bulls: acrosome morphology and semen quality in bulls of different breeds. *J Anim Sci* **55**, 638-48 (1982).
24. Cosentino, M.J. & Cockett, A.T. Structure and function of the epididymis. *Urol Res* **14**, 229-40 (1986).
25. Goyal, H.O. Morphology of the bovine epididymis. *Am J Anat* **172**, 155-72 (1985).
26. Vernocchi, V. *et al.* Sperm ubiquitination in epididymal feline semen. *Theriogenology* **82**, 636-42 (2014).

27. Sutovsky, P. *et al.* A putative, ubiquitin-dependent mechanism for the recognition and elimination of defective spermatozoa in the mammalian epididymis. *J Cell Sci* **114**, 1665-75 (2001).
28. Sutovsky, P., Neuber, E. & Schatten, G. Ubiquitin-dependent sperm quality control mechanism recognizes spermatozoa with DNA defects as revealed by dual ubiquitin-TUNEL assay. *Mol Reprod Dev* **61**, 406-13 (2002).
29. Chabory, E. *et al.* Mammalian glutathione peroxidases control acquisition and maintenance of spermatozoa integrity. *J Anim Sci* **88**, 1321-31 (2010).
30. Zhou, W., De Iuliis, G.N., Dun, M.D. & Nixon, B. Characteristics of the Epididymal Luminal Environment Responsible for Sperm Maturation and Storage. *Front Endocrinol (Lausanne)* **9**, 59 (2018).
31. Tajik, P., Arman, A. & Taktaz, T. Bovine epididymal sperm morphology obtained from caput, corpus and cauda epididymides. *Pak J Biol Sci* **10**, 3936-9 (2007).
32. Amann, R.P., Seidel, G.E., Jr. & Mortimer, R.G. Fertilizing potential in vitro of semen from young beef bulls containing a high or low percentage of sperm with a proximal droplet. *Theriogenology* **54**, 1499-515 (2000).
33. Guyonnet, B., Dacheux, F., Dacheux, J.L. & Gatti, J.L. The epididymal transcriptome and proteome provide some insights into new epididymal regulations. *J Androl* **32**, 651-64 (2011).
34. Thery, C. *et al.* Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles* **7**, 1535750 (2018).
35. Frenette, G. *et al.* Characterization of two distinct populations of epididymosomes collected in the intraluminal compartment of the bovine cauda epididymis. *Biol Reprod* **83**, 473-80 (2010).
36. Rejraji, H. *et al.* Lipid remodeling of murine epididymosomes and spermatozoa during epididymal maturation. *Biol Reprod* **74**, 1104-13 (2006).
37. Sullivan, R. Epididymosomes: a heterogeneous population of microvesicles with multiple functions in sperm maturation and storage. *Asian J Androl* **17**, 726-9 (2015).
38. Rowlison, T., Ottinger, M.A. & Comizzoli, P. Key factors enhancing sperm fertilizing ability are transferred from the epididymis to the spermatozoa via epididymosomes in the domestic cat model. *J Assist Reprod Genet* **35**, 221-228 (2018).
39. Viana, A.G.A. *et al.* Proteomic landscape of seminal plasma associated with dairy bull fertility. *Scientific Reports* **8**, 16323 (2018).
40. Juyena, N.S. & Stelletta, C. Seminal plasma: an essential attribute to spermatozoa. *J Androl* **33**, 536-51 (2012).
41. Villemure, M., Lazure, C. & Manjunath, P. Isolation and characterization of gelatin-binding proteins from goat seminal plasma. *Reprod Biol Endocrinol* **1**, 39 (2003).
42. Robertson, S.A. Seminal plasma and male factor signalling in the female reproductive tract. *Cell Tissue Res* **322**, 43-52 (2005).
43. Kasimanickam, R.K., Kasimanickam, V.R., Arangasamy, A. & Kastelic, J.P. Sperm and seminal plasma proteomics of high- versus low-fertility Holstein bulls. *Theriogenology* **126**, 41-48 (2018).
44. Vojtech, L. *et al.* Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. *Nucleic Acids Res* **42**, 7290-304 (2014).
45. Candenas, L. & Chianese, R. Exosome Composition and Seminal Plasma Proteome: A Promising Source of Biomarkers of Male Infertility. *Int J Mol Sci* **21**(2020).
46. Miller, D. The epic journey of sperm through the female reproductive tract. *animal*, 1-11 (2018).
47. Wilmut, I. & Hunter, R.H. Sperm transport into the oviducts of heifers mated early in oestrus. *Reprod Nutr Dev* **24**, 461-8 (1984).

48. Nishimura, H., Kim, E., Nakanishi, T. & Baba, T. Possible function of the ADAM1a/ADAM2 Fertilin complex in the appearance of ADAM3 on the sperm surface. *J Biol Chem* **279**, 34957-62 (2004).
49. Teijeiro, J.M., Dapino, D.G. & Marini, P.E. Porcine oviduct sperm binding glycoprotein and its deleterious effect on sperm: a mechanism for negative selection of sperm? *Biol Res* **44**, 329-37 (2011).
50. Hung, P.H. & Suarez, S.S. Regulation of sperm storage and movement in the ruminant oviduct. *Soc Reprod Fertil Suppl* **67**, 257-66 (2010).
51. Purdy, P.H. & Graham, J.K. Effect of adding cholesterol to bull sperm membranes on sperm capacitation, the acrosome reaction, and fertility. *Biol Reprod* **71**, 522-7 (2004).
52. Aitken, R.J. & Drevet, J.R. The Importance of Oxidative Stress in Determining the Functionality of Mammalian Spermatozoa: A Two-Edged Sword. *Antioxidants (Basel)* **9**(2020).
53. Sutovsky, P. Review: Sperm-oocyte interactions and their implications for bull fertility, with emphasis on the ubiquitin-proteasome system. *Animal* **12**, s121-s132 (2018).
54. Holm, P. *et al.* Developmental kinetics of the first cell cycles of bovine in vitro produced embryos in relation to their in vitro viability and sex. *Theriogenology* **50**, 1285-99 (1998).
55. Schefers, J.M. & Weigel, K.A. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. *Animal Frontiers* **2**, 4-9 (2012).
56. Fraser, L., Strzerek, J. & Kordan, W. Effect of freezing on sperm nuclear DNA. *Reprod Domest Anim* **46 Suppl 2**, 14-7 (2011).
57. MacKenna, A., Crosby, J., Huidobro, C., Correa, E. & Duque, G. Semen quality before cryopreservation and after thawing in 543 patients with testicular cancer. *JBRA Assist Reprod* **21**, 31-34 (2017).
58. Milani, C. *et al.* Effect of post-thaw dilution with caffeine, pentoxifylline, 2'-deoxyadenosine and prostatic fluid on motility of frozen-thawed dog semen. *Theriogenology* **74**, 153-64 (2010).
59. Janett, F. *et al.* Influence of Eqvalan (ivermectin) on quality and freezability of stallion semen. *Theriogenology* **55**, 785-92 (2001).
60. Takahashi, T., Itoh, R., Nishinomiya, H., Katoh, M. & Manabe, N. Effect of linoleic acid albumin in a dilution solution and long-term equilibration for freezing of bovine spermatozoa with poor freezability. *Reprod Domest Anim* **47**, 92-7 (2012).
61. Estrada, E., Rodriguez-Gil, J.E., Rivera Del Alamo, M.M., Pena, A. & Yeste, M. Effects of reduced glutathione on acrosin activity in frozen-thawed boar spermatozoa. *Reprod Fertil Dev* (2015).
62. Monteiro, J.C. *et al.* Influence of ascorbic acid and glutathione antioxidants on frozen-thawed canine semen. *Reprod Domest Anim* **44 Suppl 2**, 359-62 (2009).
63. Wu, Z. *et al.* Cryopreservation of stallion spermatozoa using different cryoprotectants and combinations of cryoprotectants. *Anim Reprod Sci* **163**, 75-81 (2015).
64. Sellem, E. *et al.* Use of combinations of in vitro quality assessments to predict fertility of bovine semen. *Theriogenology* (2015).
65. Sutovsky, P., Aarabi, M., Miranda-Vizueté, A. & Oko, R. Negative biomarker based male fertility evaluation: Sperm phenotypes associated with molecular-level anomalies. *Asian J Androl* **17**, 554-60 (2015).
66. Dogan, S. *et al.* Sperm protamine-status correlates to the fertility of breeding bulls. *Biol Reprod* **92**, 92 (2015).
67. Kennedy, C.E. *et al.* Protein expression pattern of PAWP in bull spermatozoa is associated with sperm quality and fertility following artificial insemination. *Mol Reprod Dev* **81**, 436-49 (2014).
68. Gliozzi, T.M., Turri, F., Manes, S., Cassinelli, C. & Pizzi, F. The combination of kinetic and flow cytometric semen parameters as a tool to predict fertility in cryopreserved bull semen. *Animal*, 1-8 (2017).

69. Taylor, J.F., Schnabel, R.D. & Sutovsky, P. Review: Genomics of bull fertility. *Animal*, 1-12 (2018).
70. Fritz, S. *et al.* Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. *PLoS One* **8**, e65550 (2013).
71. Boissonnas, C.C., Jouannet, P. & Jammes, H. Epigenetic disorders and male subfertility. *Fertil Steril* **99**, 624-31 (2013).
72. Budworth, P.R., Amann, R.P. & Chapman, P.L. Relationships between computerized measurements of motion of frozen-thawed bull spermatozoa and fertility. *J Androl* **9**, 41-54 (1988).
73. Perticarari, S. *et al.* A new multiparameter flow cytometric method for human semen analysis. *Hum Reprod* **22**, 485-94 (2007).
74. Nagy, S., Jansen, J., Topper, E.K. & Gadella, B.M. A triple-stain flow cytometric method to assess plasma- and acrosome-membrane integrity of cryopreserved bovine sperm immediately after thawing in presence of egg-yolk particles. *Biol Reprod* **68**, 1828-35 (2003).
75. Perrier, J.P. *et al.* A multi-scale analysis of bull sperm methylome revealed both species peculiarities and conserved tissue-specific features. *BMC Genomics* **19**, 404 (2018).
76. Comazzetto, S. *et al.* Oligoasthenoteratozoospermia and infertility in mice deficient for miR-34b/c and miR-449 loci. *PLoS Genet* **10**, e1004597 (2014).
77. Rahman, M.S., Lee, J.S., Kwon, W.S. & Pang, M.G. Sperm proteomics: road to male fertility and contraception. *Int J Endocrinol* **2013**, 360986 (2013).
78. Waddington, C.H. The Epigenotype. *International Journal of Epidemiology* **41**, 10-13 (2012).
79. Sadakierska-Chudy, A. & Filip, M. A comprehensive view of the epigenetic landscape. Part II: Histone post-translational modification, nucleosome level, and chromatin regulation by ncRNAs. *Neurotox Res* **27**, 172-97 (2015).
80. Beedle, M.T., Topping, T., Hogarth, C. & Griswold, M. Differential localization of histone variant TH2B during the first round compared with subsequent rounds of spermatogenesis. *Dev Dyn* **248**, 488-500 (2019).
81. Chamroux, A., Cocquet, J., Henry-Berger, J., Drevet, J.R. & Kocer, A. A Decade of Exploring the Mammalian Sperm Epigenome: Paternal Epigenetic and Transgenerational Inheritance. *Front Cell Dev Biol* **6**, 50 (2018).
82. Corzett, M., Mazrimas, J. & Balhorn, R. Protamine 1: protamine 2 stoichiometry in the sperm of eutherian mammals. *Mol Reprod Dev* **61**, 519-27 (2002).
83. Hamilton, T.R.S. *et al.* Detection of protamine 2 in bovine spermatozoa and testicles. *Andrology* **7**, 373-381 (2019).
84. Sillaste, G. *et al.* A novel hypothesis for histone-to-protamine transition in *Bos taurus* spermatozoa. *Reproduction* **153**, 241-251 (2017).
85. Shinagawa, T. *et al.* Disruption of Th2a and Th2b genes causes defects in spermatogenesis. *Development* **142**, 1287-92 (2015).
86. Fenic, I., Sonnack, V., Failing, K., Bergmann, M. & Steger, K. In vivo effects of histone-deacetylase inhibitor trichostatin-A on murine spermatogenesis. *J Androl* **25**, 811-8 (2004).
87. Glaser, S. *et al.* The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics Chromatin* **2**, 5 (2009).
88. Shirley, C.R., Hayashi, S., Mounsey, S., Yanagimachi, R. & Meistrich, M.L. Abnormalities and reduced reproductive potential of sperm from Tnp1- and Tnp2-null double mutant mice. *Biol Reprod* **71**, 1220-9 (2004).
89. Aoki, V.W. *et al.* Sperm protamine 1/protamine 2 ratios are related to in vitro fertilization pregnancy rates and predictive of fertilization ability. *Fertil Steril* **86**, 1408-15 (2006).
90. Grandi, F.C. *et al.* Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites. *Genome Res* **25**, 1135-46 (2015).

91. Gal-Yam, E.N. *et al.* Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc Natl Acad Sci U S A* **105**, 12979-84 (2008).
92. Tang, Q. *et al.* Idiopathic male infertility and polymorphisms in the DNA methyltransferase genes involved in epigenetic marking. *Sci Rep* **7**, 11219 (2017).
93. Marques, C.J. *et al.* Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. *Mol Hum Reprod* **14**, 67-74 (2008).
94. Marques, C.J., Carvalho, F., Sousa, M. & Barros, A. Genomic imprinting in disruptive spermatogenesis. *Lancet* **363**, 1700-2 (2004).
95. Boissonnas, C.C. *et al.* Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eur J Hum Genet* **18**, 73-80 (2010).
96. Nanassy, L. & Carrell, D.T. Abnormal methylation of the promoter of CREM is broadly associated with male factor infertility and poor sperm quality but is improved in sperm selected by density gradient centrifugation. *Fertil Steril* **95**, 2310-4 (2011).
97. Li, B. *et al.* Altered DNA methylation patterns of the H19 differentially methylated region and the DAZL gene promoter are associated with defective human sperm. *PLoS One* **8**, e71215 (2013).
98. Poplinski, A., Tuttelmann, F., Kanber, D., Horsthemke, B. & Gromoll, J. Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. *Int J Androl* **33**, 642-9 (2010).
99. Urduingio, R.G. *et al.* Aberrant DNA methylation patterns of spermatozoa in men with unexplained infertility. *Hum Reprod* **30**, 1014-28 (2015).
100. Hammond, S.M. An overview of microRNAs. *Adv Drug Deliv Rev* **87**, 3-14 (2015).
101. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855-62 (1993).
102. Griffiths-Jones, S., Saini, H.K., van Dongen, S. & Enright, A.J. miRBase: tools for microRNA genomics. *Nucleic Acids Research* **36**, D154-D158 (2008).
103. Bourdon, C. *et al.* RumimiR: a detailed microRNA database focused on ruminant species. *Database (Oxford)* **2019**(2019).
104. Schuster, A. *et al.* SpermBase: A Database for Sperm-Borne RNA Contents. *Biol Reprod* **95**, 99 (2016).
105. Hata, A. & Kashima, R. Dysregulation of microRNA biogenesis machinery in cancer. *Crit Rev Biochem Mol Biol* **51**, 121-34 (2016).
106. Zhao, H., Zhao, H., Zhang, Y. & Zhou, Y. MicroRNA199b promotes cell proliferation and invasion in Wilms' tumour by directly targeting Runtrelated transcription factor 3. *Mol Med Rep* **18**, 1812-1819 (2018).
107. Andersen, G.B., Knudsen, A., Hager, H., Hansen, L.L. & Tost, J. miRNA profiling identifies deregulated miRNAs associated with osteosarcoma development and time to metastasis in two large cohorts. *Mol Oncol* **12**, 114-131 (2018).
108. Siracusa, J. *et al.* Phenotype-Specific Response of Circulating miRNAs Provides New Biomarkers of Slow or Fast Muscle Damage. *Front Physiol* **9**, 684 (2018).
109. Ferraris, O. *et al.* Circulating microRNA profile in a mouse model of Crimean-Congo haemorrhagic fever. *Virus Res* **263**, 16-20 (2019).
110. Camps, C. *et al.* Integrated analysis of microRNA and mRNA expression and association with HIF binding reveals the complexity of microRNA expression regulation under hypoxia. *Mol Cancer* **13**, 28 (2014).
111. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-9 (2003).
112. Hutvagner, G. *et al.* A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the *let-7* Small Temporal RNA. *Science* **293**, 834-838 (2001).
113. Li, L. *et al.* The landscape of miRNA editing in animals and its impact on miRNA biogenesis and targeting. *Genome Res* **28**, 132-143 (2018).

114. Katoh, T. *et al.* Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes Dev* **23**, 433-8 (2009).
115. Lee, D., Park, D., Park, J.H., Kim, J.H. & Shin, C. Poly(A)-specific ribonuclease sculpts the 3' ends of microRNAs. *RNA* **25**, 388-405 (2019).
116. Tan, G.C. *et al.* 5' isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res* **42**, 9424-35 (2014).
117. Tan, G.C. & Dibb, N. IsomiRs have functional importance. *Malays J Pathol* **37**, 73-81 (2015).
118. Fernandez-Valverde, S.L., Taft, R.J. & Mattick, J.S. Dynamic isomiR regulation in *Drosophila* development. *RNA* **16**, 1881-8 (2010).
119. Wyman, S.K. *et al.* Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res* **21**, 1450-61 (2011).
120. Burroughs, A.M. *et al.* A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome Res* **20**, 1398-410 (2010).
121. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-33 (2009).
122. Garcia-Lopez, J. *et al.* Diversity and functional convergence of small noncoding RNAs in male germ cell differentiation and fertilization. *RNA* **21**, 946-62 (2015).
123. Donkin, I. & Barres, R. Sperm epigenetics and influence of environmental factors. *Mol Metab* (2018).
124. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**, 91-105 (2007).
125. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437-41 (2004).
126. Newman, M.A., Thomson, J.M. & Hammond, S.M. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* **14**, 1539-49 (2008).
127. Heo, I. *et al.* TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**, 696-708 (2009).
128. Friedman, R.C., Farh, K.K., Burge, C.B. & Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92-105 (2009).
129. Bernstein, E. *et al.* Dicer is essential for mouse development. *Nat Genet* **35**, 215-7 (2003).
130. Krill, K.T., Gurdziel, K., Heaton, J.H., Simon, D.P. & Hammer, G.D. Dicer deficiency reveals microRNAs predicted to control gene expression in the developing adrenal cortex. *Molecular endocrinology (Baltimore, Md.)* **27**, 754-768 (2013).
131. Andl, T. *et al.* The miRNA-processing enzyme dicer is essential for the morphogenesis and maintenance of hair follicles. *Curr Biol* **16**, 1041-9 (2006).
132. Kanellopoulou, C. *et al.* Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes & development* **19**, 489-501 (2005).
133. Weick, E.M. & Miska, E.A. piRNAs: from biogenesis to function. *Development* **141**, 3458-71 (2014).
134. Russell, S.J. *et al.* Identification of PIWIL1 Isoforms and Their Expression in Bovine Testes, Oocytes, and Early Embryos. *Biol Reprod* **94**, 75 (2016).
135. Gou, L.T. *et al.* Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res* **24**, 680-700 (2014).
136. Czech, B. *et al.* piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet* **52**, 131-157 (2018).
137. Aravin, A.A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**, 785-99 (2008).
138. Czech, B. & Hannon, G.J. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem Sci* **41**, 324-337 (2016).
139. Wu, X. *et al.* The Biogenesis and Functions of piRNAs in Human Diseases. *Mol Ther Nucleic Acids* **21**, 108-120 (2020).

140. Pillai, R.S. & Chuma, S. piRNAs and their involvement in male germline development in mice. *Dev Growth Differ* **54**, 78-92 (2012).
141. Maruyama, S., Sugahara, J., Kanai, A. & Nozaki, H. Permuted tRNA genes in the nuclear and nucleomorph genomes of photosynthetic eukaryotes. *Mol Biol Evol* **27**, 1070-6 (2010).
142. Sellem, E. *et al.* A comprehensive overview of bull sperm-borne small non-coding RNAs and their diversity across breeds. *Epigenetics & Chromatin* **13**(2020).
143. Garcia-Silva, M.R., Cabrera-Cabrera, F., Guida, M.C. & Cayota, A. Hints of tRNA-Derived Small RNAs Role in RNA Silencing Mechanisms. *Genes (Basel)* **3**, 603-14 (2012).
144. Zhang, X., Cozen, A.E., Liu, Y., Chen, Q. & Lowe, T.M. Small RNA Modifications: Integral to Function and Disease. *Trends Mol Med* **22**, 1025-1034 (2016).
145. Keam, S.P. & Hutvagner, G. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. *Life (Basel)* **5**, 1638-51 (2015).
146. Raina, M. & Ibba, M. tRNAs as regulators of biological processes. *Front Genet* **5**, 171 (2014).
147. Schorn, A.J., Gutbrod, M.J., LeBlanc, C. & Martienssen, R. LTR-Retrotransposon Control by tRNA-Derived Small RNAs. *Cell* **170**, 61-71.e11 (2017).
148. Yan, Q., Zhu, C., Guang, S. & Feng, X. The Functions of Non-coding RNAs in rRNA Regulation. *Front Genet* **10**, 290 (2019).
149. Tomecki, R., Sikorski, P.J. & Zakrzewska-Placzek, M. Comparison of preribosomal RNA processing pathways in yeast, plant and human cells - focus on coordinated action of endo- and exoribonucleases. *FEBS Lett* **591**, 1801-1850 (2017).
150. Wei, H. *et al.* Profiling and identification of small rDNA-derived RNAs and their potential biological functions. *PLoS One* **8**, e56842 (2013).
151. Chen, Z. *et al.* Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs. *PLoS one* **12**, e0176458-e0176458 (2017).
152. Lambert, M., Benmoussa, A. & Provost, P. Small Non-Coding RNAs Derived From Eukaryotic Ribosomal RNA. *Non-coding RNA* **5**, 16 (2019).
153. Chu, C. *et al.* A sequence of 28S rRNA-derived small RNAs is enriched in mature sperm and various somatic tissues and possibly associates with inflammation. *J Mol Cell Biol* **9**, 256-259 (2017).
154. Zhang, Y. *et al.* Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nature Cell Biology* **20**, 535-540 (2018).
155. Locati, M.D. *et al.* Identifying small RNAs derived from maternal- and somatic-type rRNAs in zebrafish development. *Genome* **61**, 371-378 (2018).
156. Locati, M.D. *et al.* Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development. *RNA (New York, N.Y.)* **23**, 1188-1199 (2017).
157. Guan, L. & Grigoriev, A. Age-Related Argonaute Loading of Ribosomal RNA Fragments. *Microna* **9**, 142-152 (2020).
158. Li, X., Ao, J. & Wu, J. Systematic identification and comparison of expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in mouse germline stem cells. *Oncotarget* **8**, 26573-26590 (2017).
159. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333-8 (2013).
160. Hansen, T.B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384-8 (2013).
161. Ouyang, H. *et al.* Circular RNA circSVIL Promotes Myoblast Proliferation and Differentiation by Sponging miR-203 in Chicken. *Front Genet* **9**, 172 (2018).
162. Zhang, H. *et al.* Exosomal circRNA derived from gastric tumor promotes white adipose browning by targeting the miR-133/PRDM16 pathway. *Int J Cancer* (2018).
163. Guo, L., Zheng, L., Zhao, Y. & Wang, Q. Profiling and Bioinformatic Analyses Indicate Differential circRNA and miRNA/isomiR Expression and Interactions. *Biomed Res Int* **2018**, 8518563 (2018).

164. Franco-Zorrilla, J.M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet* **39**, 1033-7 (2007).
165. Laiho, A., Kotaja, N., Gyenesi, A. & Sironen, A. Transcriptome profiling of the murine testis during the first wave of spermatogenesis. *PLoS One* **8**, e61558 (2013).
166. Chalmel, F. *et al.* The conserved transcriptome in human and rodent male gametogenesis. *Proc Natl Acad Sci U S A* **104**, 8346-51 (2007).
167. Chen, Q., Yan, W. & Duan, E. Epigenetic inheritance of acquired traits through sperm RNAs and sperm RNA modifications. *Nat Rev Genet* **17**, 733-743 (2016).
168. Korhonen, H.M. *et al.* Dicer is required for haploid male germ cell differentiation in mice. *PLoS One* **6**, e24821 (2011).
169. Zimmermann, C. *et al.* Germ cell-specific targeting of DICER or DGCR8 reveals a novel role for endo-siRNAs in the progression of mammalian spermatogenesis and male fertility. *PLoS One* **9**, e107023 (2014).
170. Papaioannou, M.D. & Nef, S. microRNAs in the testis: building up male fertility. *J Androl* **31**, 26-33 (2010).
171. Kotaja, N. *et al.* The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proc Natl Acad Sci U S A* **103**, 2647-52 (2006).
172. Ro, S., Park, C., Sanders, K.M., McCarrey, J.R. & Yan, W. Cloning and expression profiling of testis-expressed microRNAs. *Dev Biol* **311**, 592-602 (2007).
173. Niu, Z. *et al.* MicroRNA-21 regulates the self-renewal of mouse spermatogonial stem cells. *Proc Natl Acad Sci U S A* **108**, 12740-5 (2011).
174. He, Z. *et al.* MiRNA-20 and mirna-106a regulate spermatogonial stem cell renewal at the post-transcriptional level via targeting STAT3 and Ccnd1. *STEM CELLS* **31**, 2205-2217 (2013).
175. Huang, Y.L. *et al.* miR-100 promotes the proliferation of spermatogonial stem cells via regulating Stat3. *Mol Reprod Dev* **84**, 693-701 (2017).
176. Huszar, J.M. & Payne, C.J. MicroRNA 146 (Mir146) modulates spermatogonial differentiation by retinoic acid in mice. *Biol Reprod* **88**, 15 (2013).
177. Yang, Q.-E., Racicot, K.E., Kaucher, A.V., Oatley, M.J. & Oatley, J.M. MicroRNAs 221 and 222 regulate the undifferentiated state in mammalian male germ cells. *Development* **140**, 280 (2013).
178. Huang, Z. *et al.* miR-34c disrupts spermatogonial stem cell homeostasis in cryptorchid testes by targeting Nanos2. *Reproductive Biology and Endocrinology* **16**, 97 (2018).
179. Yu, M. *et al.* miR-34c enhances mouse spermatogonial stem cells differentiation by targeting Nanos2. *J Cell Biochem* **115**, 232-42 (2014).
180. Bouhallier, F. *et al.* Role of miR-34c microRNA in the late steps of spermatogenesis. *RNA* **16**, 720-31 (2010).
181. Bouhallier, F. *et al.* Role of miR-34c microRNA in the late steps of spermatogenesis. *RNA (New York, N.Y.)* **16**, 720-731 (2010).
182. Yu, M. *et al.* miR-34c Enhances Mouse Spermatogonial Stem Cells Differentiation by Targeting Nanos2. *Journal of Cellular Biochemistry* **115**, 232-242 (2014).
183. Damestoy, A., Perrard, M.-H., Vigier, M., Sabido, O. & Durand, P. Transforming growth factor beta-1 decreases the yield of the second meiotic division of rat pachytene spermatocytes in vitro. *Reproductive biology and endocrinology : RB&E* **3**, 22-22 (2005).
184. Kostereva, N. & Hofmann, M.C. Regulation of the Spermatogonial Stem Cell Niche. *Reproduction in Domestic Animals* **43**, 386-392 (2008).
185. Hilz, S., Fogarty, E.A., Modzelewski, A.J., Cohen, P.E. & Grimson, A. Transcriptome profiling of the developing male germ line identifies the miR-29 family as a global regulator during meiosis. *RNA Biol* **14**, 219-235 (2017).
186. Liu, T. *et al.* MicroRNA-122 influences the development of sperm abnormalities from human induced pluripotent stem cells by regulating TNP2 expression. *Stem Cells Dev* **22**, 1839-50 (2013).

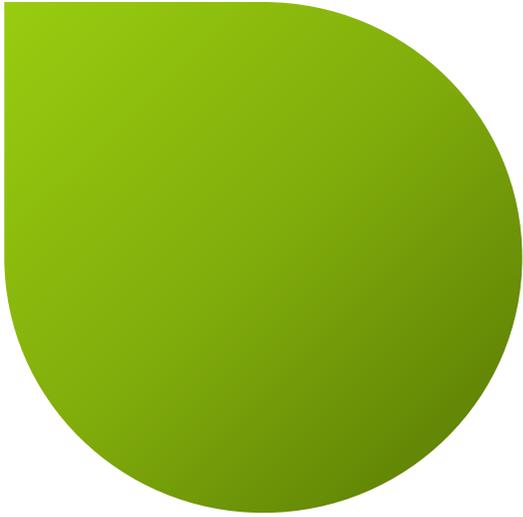
187. Dai, L. *et al.* Testis-specific miRNA-469 Up-regulated in Gonadotropin-regulated Testicular RNA Helicase (GRTH/DDX25)-null Mice Silences Transition Protein 2 and Protamine 2 Messages at Sites within Coding Region: IMPLICATIONS OF ITS ROLE IN GERM CELL DEVELOPMENT. *Journal of Biological Chemistry* **286**, 44306-44318 (2011).
188. Tariq, K., Metzendorf, C., Peng, W., Sohail, S. & Zhang, H. miR-8-3p regulates mitoferrin in the testes of *Bactrocera dorsalis* to ensure normal spermatogenesis. *Sci Rep* **6**, 22565 (2016).
189. Chuma, S. & Nakano, T. piRNA and spermatogenesis in mice. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **368**, 20110338-20110338 (2013).
190. Zoch, A. *et al.* SPOCD1 is an essential executor of piRNA-directed de novo DNA methylation. *Nature* **584**, 635-639 (2020).
191. Newkirk, S.J. *et al.* Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proc Natl Acad Sci U S A* **114**, E5635-e5644 (2017).
192. Yabuta, Y. *et al.* TDRD5 is required for retrotransposon silencing, chromatoid body assembly, and spermiogenesis in mice. *J Cell Biol* **192**, 781-95 (2011).
193. Yan, W.E.-I.-C.B.O.R. piRNA-independent PIWI function in spermatogenesis and male fertility. *Biol Reprod* **96**, 1121-1123 (2017).
194. Chu, C. *et al.* Epididymal small non-coding RNA studies: progress over the past decade. *Andrology* **7**, 681-689 (2019).
195. Peng, H. *et al.* A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. *Cell Res* **22**, 1609-12 (2012).
196. Nixon, B. *et al.* The MicroRNA Signature of Mouse Spermatozoa Is Substantially Modified During Epididymal Maturation. *Biology of Reproduction* **93**, Article 91, 1-20 (2015).
197. Sharma, U. *et al.* Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm. *Dev Cell* **46**, 481-494 e6 (2018).
198. Hutcheon, K. *et al.* Analysis of the small non-protein-coding RNA profile of mouse spermatozoa reveals specific enrichment of piRNAs within mature spermatozoa. *RNA Biology* **14**, 1776-1790 (2017).
199. Belleannée, C., Legare, C., Calvo, E., Thimon, V. & Sullivan, R. microRNA signature is altered in both human epididymis and seminal microvesicles following vasectomy. *Hum Reprod* **28**, 1455-67 (2013).
200. Nixon, B. *et al.* Profiling of epididymal small non-protein-coding RNAs. *Andrology* **7**, 669-680 (2019).
201. Belleannée, C., Calvo, É., Caballero, J. & Sullivan, R. Epididymosomes Convey Different Repertoires of MicroRNAs Throughout the Bovine Epididymis1. *Biology of Reproduction* **89**(2013).
202. Sullivan, R. & Saez, F. Epididymosomes, prostasomes, and liposomes: their roles in mammalian male reproductive physiology. *Reproduction* **146**, R21-35 (2013).
203. Sharma, U. *et al.* Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391-396 (2016).
204. Baskaran, S., Panner Selvam, M.K. & Agarwal, A. Exosomes of male reproduction. *Adv Clin Chem* **95**, 149-163 (2020).
205. Reilly, J.N. *et al.* Characterisation of mouse epididymosomes reveals a complex profile of microRNAs and a potential mechanism for modification of the sperm epigenome. *Sci Rep* **6**, 31794 (2016).
206. Chen, C. *et al.* Comparative profiling of small RNAs of pig seminal plasma and ejaculated and epididymal sperm. *Reproduction* **153**, 785-796 (2017).
207. Sharma, U. *et al.* Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391-6 (2016).
208. Sharma, U. *et al.* Small RNAs Are Trafficked from the Epididymis to Developing Mammalian Sperm. *Dev Cell* **46**, 481-494.e6 (2018).
209. Nixon, B. *et al.* Next Generation Sequencing Analysis Reveals Segmental Patterns of microRNA Expression in Mouse Epididymal Epithelial Cells. *PLoS One* **10**, e0135605 (2015).

210. Barcelo, M., Mata, A., Bassas, L. & Larriba, S. Exosomal microRNAs in seminal plasma are markers of the origin of azoospermia and can predict the presence of sperm in testicular tissue. *Hum Reprod* **33**, 1087-1098 (2018).
211. Abu-Halima, M. *et al.* Altered micro-ribonucleic acid expression profiles of extracellular microvesicles in the seminal plasma of patients with oligoasthenozoospermia. *Fertil Steril* **106**, 1061-1069 e3 (2016).
212. Sullivan, R., Frenette, G. & Girouard, J. Epididymosomes are involved in the acquisition of new sperm proteins during epididymal transit. *Asian J Androl* **9**, 483-91 (2007).
213. Baker, M.A. The 'omics revolution and our understanding of sperm cell biology. *Asian J Androl* **13**, 6-10 (2011).
214. Gur, Y. & Breitbart, H. Mammalian sperm translate nuclear-encoded proteins by mitochondrial-type ribosomes. *Genes & development* **20**, 411-416 (2006).
215. Gur, Y. & Breitbart, H. Protein translation in mammalian sperm. *Soc Reprod Fertil Suppl* **65**, 391-7 (2007).
216. Gapp, K. *et al.* Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. *Nat Neurosci* **17**, 667-9 (2014).
217. Daxinger, L. & Whitelaw, E. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* **13**, 153-62 (2012).
218. Rassoulzadegan, M. *et al.* RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469-74 (2006).
219. Wagner, K.D. *et al.* RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. *Dev Cell* **14**, 962-9 (2008).
220. Liu, W.M. *et al.* Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proc Natl Acad Sci U S A* **109**, 490-4 (2012).
221. Chen, Q. *et al.* Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* **351**, 397-400 (2016).
222. Sharma, U. Paternal Contributions to Offspring Health: Role of Sperm Small RNAs in Intergenerational Transmission of Epigenetic Information. *Front Cell Dev Biol* **7**, 215 (2019).
223. Zhang, Y., Shi, J., Rassoulzadegan, M., Tuorto, F. & Chen, Q. Sperm RNA code programmes the metabolic health of offspring. *Nat Rev Endocrinol* **15**, 489-498 (2019).
224. Chan, J.C. *et al.* Reproductive tract extracellular vesicles are sufficient to transmit intergenerational stress and program neurodevelopment. *Nature Communications* **11**, 1499 (2020).
225. Yuan, S. *et al.* Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. *Development* **143**, 635-47 (2016).
226. Yang, W.J. *et al.* Dicer is required for embryonic angiogenesis during mouse development. *J Biol Chem* **280**, 9330-5 (2005).
227. Conine, C.C., Sun, F., Song, L., Rivera-Perez, J.A. & Rando, O.J. Small RNAs Gained during Epididymal Transit of Sperm Are Essential for Embryonic Development in Mice. *Dev Cell* **46**, 470-480 e3 (2018).
228. Lv, C. *et al.* MicroRNA-21 in extracellular vesicles contributes to the growth of fertilized eggs and embryo development in mice. *Biosci Rep* (2018).
229. Zhang, Z. *et al.* MicroRNA-29b regulates DNA methylation by targeting Dnmt3a/3b and Tet1/2/3 in porcine early embryo development. *Dev Growth Differ* **60**, 197-204 (2018).
230. Alves, M.B.R. *et al.* Sperm-borne miR-216b modulates cell proliferation during early embryo development via K-RAS. *Sci Rep* **9**, 10358 (2019).
231. Chen, X. *et al.* Early cleavage of preimplantation embryos is regulated by tRNAGln-TTG-derived small RNAs present in mature spermatozoa. *Journal of Biological Chemistry* **295**, 10885-10900 (2020).
232. Salilew-Wondim, D. *et al.* The Role of MicroRNAs in Mammalian Fertility: From Gametogenesis to Embryo Implantation. *Int J Mol Sci* **21**(2020).

233. Carone, B.R. *et al.* Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084-96 (2010).
234. de Castro Barbosa, T. *et al.* High-fat diet reprograms the epigenome of rat spermatozoa and transgenerationally affects metabolism of the offspring. *Mol Metab* **5**, 184-197 (2016).
235. Grandjean, V. *et al.* RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Sci Rep* **5**, 18193 (2015).
236. Rodgers, A.B., Morgan, C.P., Leu, N.A. & Bale, T.L. Transgenerational epigenetic programming via sperm microRNA recapitulates effects of paternal stress. *Proc Natl Acad Sci U S A* **112**, 13699-704 (2015).
237. Capra, E. *et al.* Small RNA sequencing of cryopreserved semen from single bull revealed altered miRNAs and piRNAs expression between High- and Low-motile sperm populations. *BMC Genomics* **18**, 14 (2017).
238. Stowe, H.M. *et al.* The bull sperm microRNAome and the effect of fescue toxicosis on sperm microRNA expression. *PLoS One* **9**, e113163 (2014).
239. Abu-Halima, M. *et al.* Altered microRNA expression profiles of human spermatozoa in patients with different spermatogenic impairments. *Fertil Steril* **99**, 1249-1255 e16 (2013).
240. Liu, T., Cheng, W., Gao, Y., Wang, H. & Liu, Z. Microarray analysis of microRNA expression patterns in the semen of infertile men with semen abnormalities. *Mol Med Rep* **6**, 535-42 (2012).
241. Fagerlind, M., Stalhammar, H., Olsson, B. & Klinga-Levan, K. Expression of miRNAs in Bull Spermatozoa Correlates with Fertility Rates. *Reprod Domest Anim* (2015).
242. Govindaraju, A. *et al.* Dynamics of microRNAs in bull spermatozoa. *Reprod Biol Endocrinol* **10**, 82 (2012).
243. Sellem, E. *et al.* A comprehensive overview of bull sperm-borne small non-coding RNAs and their diversity across breeds. *Epigenetics Chromatin* **13**, 19 (2020).
244. Cuthbert, J.M., Russell, S.J., White, K.L. & Benninghoff, A.D. The maternal-to-zygotic transition in bovine in vitro-fertilized embryos is associated with marked changes in small non-coding RNAs. *Biol Reprod* **100**, 331-350 (2019).
245. Graf, A. *et al.* Fine mapping of genome activation in bovine embryos by RNA sequencing. *Proc Natl Acad Sci U S A* **111**, 4139-44 (2014).
246. Zhao, X.M. *et al.* Transcriptome analyses of inner cell mass and trophectoderm cells isolated by magnetic-activated cell sorting from bovine blastocysts using single cell RNA-seq. *Reprod Domest Anim* **51**, 726-35 (2016).
247. Wu, C., Blondin, P., Vigneault, C., Labrecque, R. & Sirard, M.A. Sperm miRNAs- potential mediators of bull age and early embryo development. *BMC Genomics* **21**, 798 (2020).
248. Hua, M. *et al.* Identification of small non-coding RNAs as sperm quality biomarkers for in vitro fertilization. *Cell Discovery* **5**, 20 (2019).
249. Alves, M.B.R., Celeghini, E.C.C. & Belleannée, C. From Sperm Motility to Sperm-Borne microRNA Signatures: New Approaches to Predict Male Fertility Potential. *Front Cell Dev Biol* **8**, 791 (2020).
250. Abu-Halima, M. *et al.* Altered microRNA expression profiles of human spermatozoa in patients with different spermatogenic impairments. *Fertil Steril* **99**, 1249-1255.e16 (2013).
251. Fagerlind, M., Stålhammar, H., Olsson, B. & Klinga-Levan, K. Expression of miRNAs in bull spermatozoa correlates with fertility rates. *Reprod Domest Anim* **50**(2015).
252. Wang, C. *et al.* Altered profile of seminal plasma microRNAs in the molecular diagnosis of male infertility. *Clin Chem* **57**, 1722-31 (2011).
253. Salas-Huetos, A. *et al.* Spermatozoa from normozoospermic fertile and infertile individuals convey a distinct miRNA cargo. *Andrology* **4**, 1028-1036 (2016).
254. PERKINS, S. *et al.* Influence of microRNAs from semen on bovine fertility. in *Society for the Study of Reproduction Annual Meetin* (2019).
255. Keles, E. *et al.* The micro-RNA content of unsorted cryopreserved bovine sperm and its relation to the fertility of sperm after sex-sorting. *BMC Genomics* **22**, 30 (2021).

256. Muñoz, X., Mata, A., Bassas, L. & Larriba, S. Altered miRNA Signature of Developing Germ-cells in Infertile Patients Relates to the Severity of Spermatogenic Failure and Persists in Spermatozoa. *Scientific reports* **5**, 17991-17991 (2015).
257. Tian, H. *et al.* Semen-specific miRNAs: Suitable for the distinction of infertile semen in the body fluid identification? *Forensic Sci Int Genet* **33**, 161-167 (2018).
258. Xu, H. *et al.* MicroRNA expression profile analysis in sperm reveals hsa-mir-191 as an auspicious omen of in vitro fertilization. *BMC Genomics* **21**, 165 (2020).
259. Barrey, E. *et al.* Pre-microRNA and mature microRNA in human mitochondria. *PLoS one* **6**, e20220-e20220 (2011).
260. Pinti, M.V. *et al.* microRNA Changes in Diabetic Cardiac Mitochondria: What are they doing there? *The FASEB Journal* **33**, 713.3-713.3 (2019).
261. Morita, M. *et al.* mTOR coordinates protein synthesis, mitochondrial activity and proliferation. *Cell Cycle* **14**, 473-80 (2015).
262. Silva, J.V. *et al.* mTOR Signaling Pathway Regulates Sperm Quality in Older Men. *Cells* **8**(2019).
263. Mahran, A.M. *et al.* The correlation between mammalian target of rapamycin (mTOR) gene expression and sperm DNA damage among infertile patients with and without varicocele. *Andrologia* **51**, e13341 (2019).
264. Li, K. *et al.* miR-146a improves hepatic lipid and glucose metabolism by targeting MED1. *Int J Mol Med* **45**, 543-555 (2020).
265. Heggermont, W.A. *et al.* Inhibition of MicroRNA-146a and Overexpression of Its Target Dihydrolipoyl Succinyltransferase Protect Against Pressure Overload-Induced Cardiac Hypertrophy and Dysfunction. *Circulation* **136**, 747-761 (2017).
266. Pan, J.-A. *et al.* miR-146a attenuates apoptosis and modulates autophagy by targeting TAF9b/P53 pathway in doxorubicin-induced cardiotoxicity. *Cell Death & Disease* **10**, 668 (2019).
267. Aparicio, I.M. *et al.* Autophagy-related proteins are functionally active in human spermatozoa and may be involved in the regulation of cell survival and motility. *Sci Rep* **6**, 33647 (2016).
268. Gallardo Bolaños, J.M. *et al.* Autophagy and Apoptosis Have a Role in the Survival or Death of Stallion Spermatozoa during Conservation in Refrigeration. *PLOS ONE* **7**, e30688 (2012).
269. Kamity, R., Sharma, S. & Hanna, N. MicroRNA-Mediated Control of Inflammation and Tolerance in Pregnancy. *Frontiers in immunology* **10**, 718-718 (2019).
270. Svoboda, P. & Flemr, M. The role of miRNAs and endogenous siRNAs in maternal-to-zygotic reprogramming and the establishment of pluripotency. *EMBO reports* **11**, 590-597 (2010).
271. Salilew-Wondim, D. *et al.* The Role of MicroRNAs in Mammalian Fertility: From Gametogenesis to Embryo Implantation. *International journal of molecular sciences* **21**, 585 (2020).
272. Cuthbert, J.M., Russell, S.J., White, K.L. & Benninghoff, A.D. The maternal-to-zygotic transition in bovine in vitro-fertilized embryos is associated with marked changes in small non-coding RNAs†. *Biology of Reproduction* **100**, 331-350 (2019).
273. Tang, F. *et al.* Maternal microRNAs are essential for mouse zygotic development. *Genes & Development* **21**, 644-648 (2007).
274. Melo-Baez, B. *et al.* MicroRNAs from Extracellular Vesicles Secreted by Bovine Embryos as Early Biomarkers of Developmental Competence. *Int J Mol Sci* **21**(2020).
275. Mellisho, E.A. *et al.* Extracellular vesicles secreted during blastulation show viability of bovine embryos. *Reproduction* **158**, 477-492 (2019).
276. Soares, A.R. *et al.* Ethanol exposure induces upregulation of specific microRNAs in zebrafish embryos. *Toxicol Sci* **127**, 18-28 (2012).
277. Sandra, O., Charpigny, G., Galio, L. & Hue, I. Preattachment Embryos of Domestic Animals: Insights into Development and Paracrine Secretions. *Annu Rev Anim Biosci* **5**, 205-228 (2017).
278. Kaneko, K. *et al.* Endothelial expression of bone morphogenetic protein receptor type 1a is required for atrioventricular valve formation. *Ann Thorac Surg* **85**, 2090-8 (2008).

279. Luo, S. *et al.* Drosophila tsRNAs preferentially suppress general translation machinery via antisense pairing and participate in cellular starvation response. *Nucleic Acids Res* **46**, 5250-5268 (2018).
280. Gur, Y. & Breitbart, H. Protein synthesis in sperm: dialog between mitochondria and cytoplasm. *Mol Cell Endocrinol* **282**, 45-55 (2008).
281. Aparicio, I.M. *et al.* Autophagy-related proteins are functionally active in human spermatozoa and may be involved in the regulation of cell survival and motility. *Scientific reports* **6**, 33647-33647 (2016).
282. Rossato, M. *et al.* Sperm treatment with extracellular ATP increases fertilization rates in in-vitro fertilization for male factor infertility. *Hum Reprod* **14**, 694-7 (1999).
283. Shangguan, A. *et al.* Cryopreservation Induces Alterations of miRNA and mRNA Fragment Profiles of Bull Sperm. *Frontiers in Genetics* **11**(2020).
284. Tan, M., van Tol, H.T.A., Mokry, M., Stout, T.A.E. & Roelen, B.A.J. Microinjection induces changes in the transcriptome of bovine oocytes. *Sci Rep* **10**, 11211 (2020).
285. Lin, X. *et al.* Bta-miR-10b Secreted by Bovine Embryos Negatively Impacts Preimplantation Embryo Quality. *Front Genet* **10**, 757 (2019).
286. Stopka, T. & Skoultchi, A.I. The ISWI ATPase Snf2h is required for early mouse development. *Proc Natl Acad Sci U S A* **100**, 14097-102 (2003).
287. Abu-Halima, M. *et al.* MicroRNA expression profiles in human testicular tissues of infertile men with different histopathologic patterns. *Fertil Steril* **101**, 78-86 e2 (2014).
288. Permeth-Wey, J. *et al.* A genome-wide investigation of microRNA expression identifies biologically-meaningful microRNAs that distinguish between high-risk and low-risk intraductal papillary mucinous neoplasms of the pancreas. *PLoS One* **10**, e0116869 (2015).
289. Nätt, D. *et al.* Human sperm displays rapid responses to diet. *PLOS Biology* **17**, e3000559 (2019).
290. Seifi-Jamadi, A. *et al.* Influence of seasonal differences on semen quality and subsequent embryo development of Belgian Blue bulls. *Theriogenology* **158**, 8-17 (2020).
291. Radwan, M. *et al.* Urinary Bisphenol A Levels and Male Fertility. *Am J Mens Health* **12**, 2144-2151 (2018).
292. Stermer, A.R., Reyes, G., Hall, S.J. & Boekelheide, K. Small RNAs in Rat Sperm Are a Predictive and Sensitive Biomarker of Exposure to the Testicular Toxicant Ethylene Glycol Monomethyl Ether. *Toxicol Sci* **169**, 399-408 (2019).
293. Liu, Y. *et al.* Extremely Low-Frequency Electromagnetic Fields Affect the miRNA-Mediated Regulation of Signaling Pathways in the GC-2 Cell Line. *PLoS One* **10**, e0139949 (2015).
294. Donkin, I. *et al.* Obesity and Bariatric Surgery Drive Epigenetic Variation of Spermatozoa in Humans. *Cell Metab* **23**, 369-78 (2016).
295. Ingerslev, L.R. *et al.* Endurance training remodels sperm-borne small RNA expression and methylation at neurological gene hotspots. *Clin Epigenetics* **10**, 12 (2018).
296. Zhou, H. *et al.* Differences in small noncoding RNAs profile between bull X and Y sperm. *PeerJ* **8**, e9822 (2020).
297. Shangguan, A. *et al.* Cryopreservation Induces Alterations of miRNA and mRNA Fragment Profiles of Bull Sperm. *Front Genet* **11**, 419 (2020).



# Annexes

1) Figures supplémentaires du premier article

**Additional Fig S1. RNA quality control.**

**Additional Fig S2. NGS data quality controls.**

**Additional Fig S3. Reads annotated as mRNA fragments.**

**Additional Fig S4. Reads annotated as rRNA.**

**Additional Fig S5. Biogenesis of tRFs and classification of fragments.**

**Additional Fig S6. Frequency of tRFs isotypes.**

**Additional Fig S7. Distribution of cleavage sites along the anticodon-Loop.**

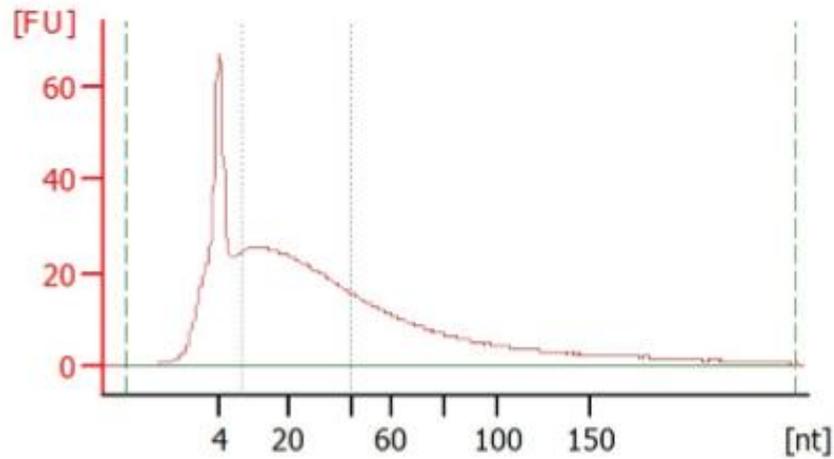
**Additional Fig S8. Distribution of Pearson correlation for all and clustered miRNAs.**

**Additional Fig S9. Five putative functional regulation networks.**

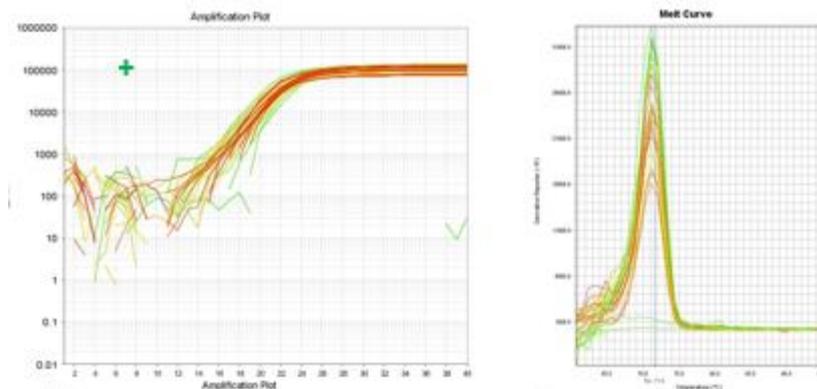
**Additional Fig S10. Classification and examples of isomiRs for bta-chr29\_51574.**

**Additional Fig S11. Frequency of substitution and nucleotide usage in miRNA sequences.**

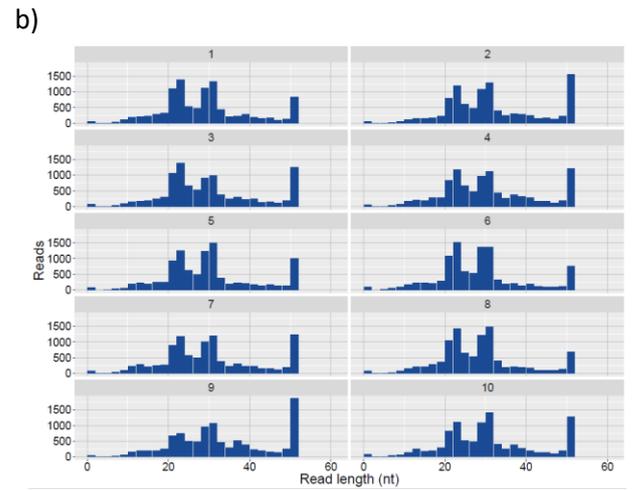
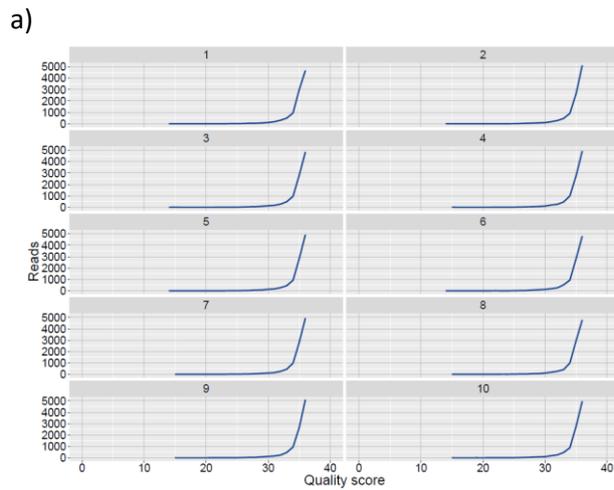
a)



b)



Additional Fig S1: RNA quality control. Bioanalyzer profil and RTqPCR with miRNA-125b-5p primers were performed to validate RNA size, quality and concentrations, starting from 5 ng of total RNA, as estimated by Qubit® RNA HS Assay Kit. **(a)** Typical electrophoretic profiles were obtained, with the expected dome around 20 nucleotides and no evidence of 18S and 28S rRNAs. **(b)** Consistent amplification results were obtained (Ct in the range 20-21), indicating that the Qubit –estimated concentration was reliable and that no inhibitors remain in the RNA preparation. Single peak melting curves were also obtained, indicating that a single, specific product has been produced.

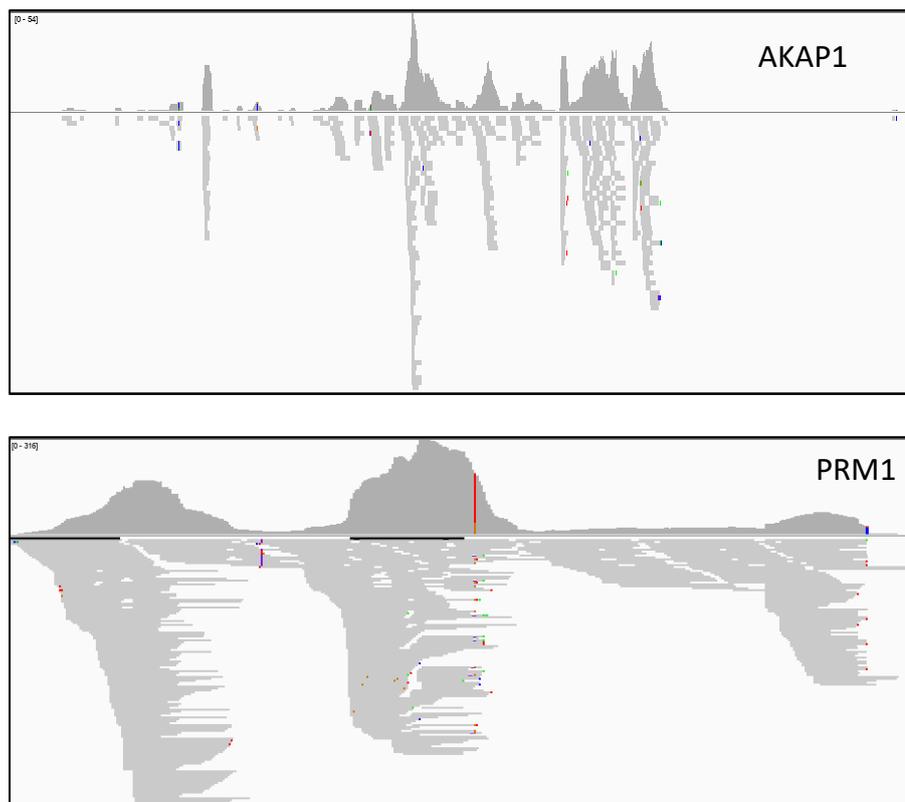


Additional Fig S2: NGS data quality controls. **(a)** Phred Quality score distribution over all sequences for the 10 first samples, showing Q score above 30 for more than 97% of reads. **(b)** Read length distribution for the 10 first samples, showing two main peaks at 18-26nt (microRNA or siRNA) and 28-32nt (piRNA or tsRNAs).

a)

Size range	Nbr	%
<20	939	3.4
20-25	4378	15.8
26-30	4798	17.3
31-35	3578	12.9
36-40	2349	8.5
41-45	1192	4.3
>46	10421	37.7

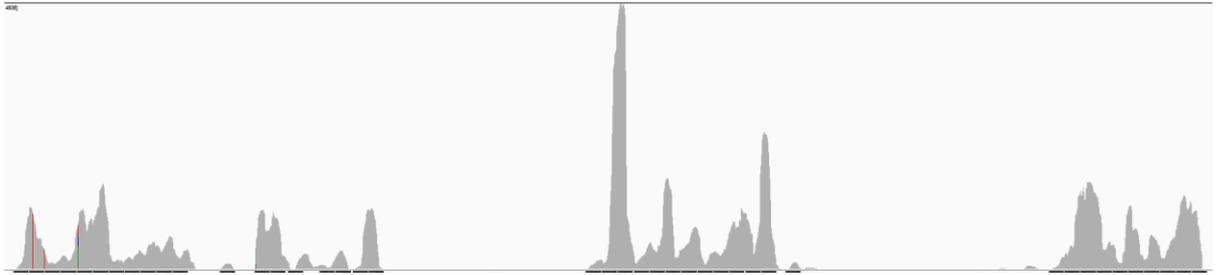
b)



Additional Fig S3: reads annotated as mRNA. About 27655 reads have been identified as mRNA fragments. **(a)** The majority of these reads were 50nt in length. **(b)** IGV profiles of two genes (AKAP1 and PRM1) covered by well distributed unique reads having high count levels.

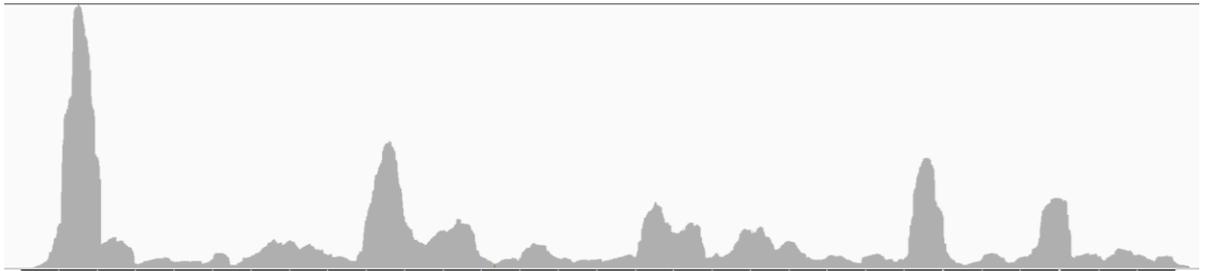
a)

18S rRNA



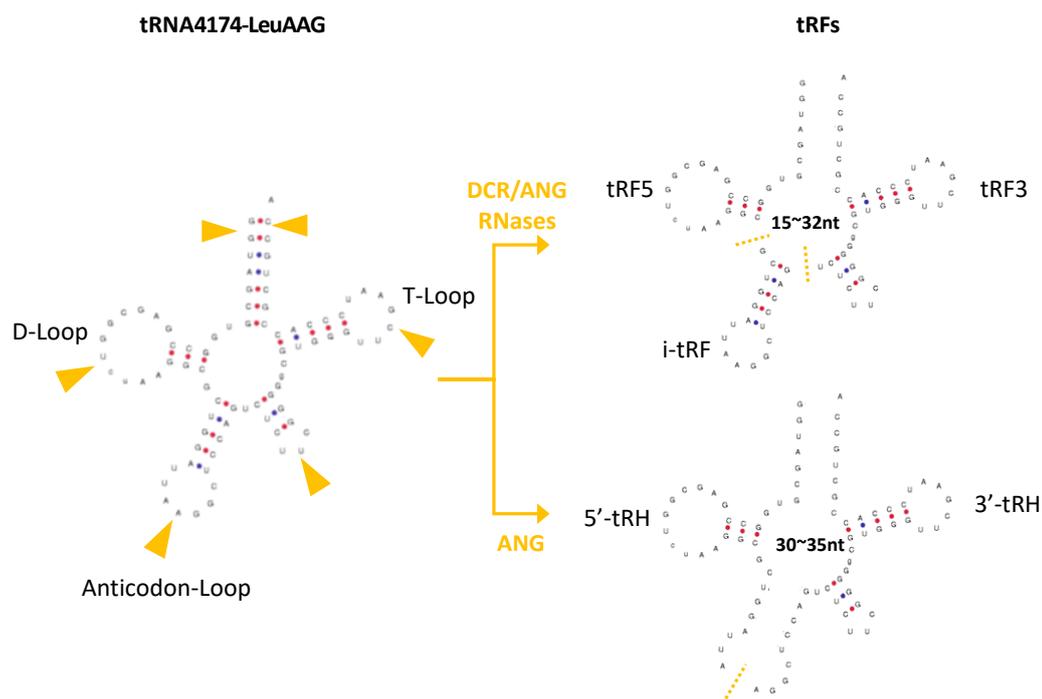
b)

28S rRNA

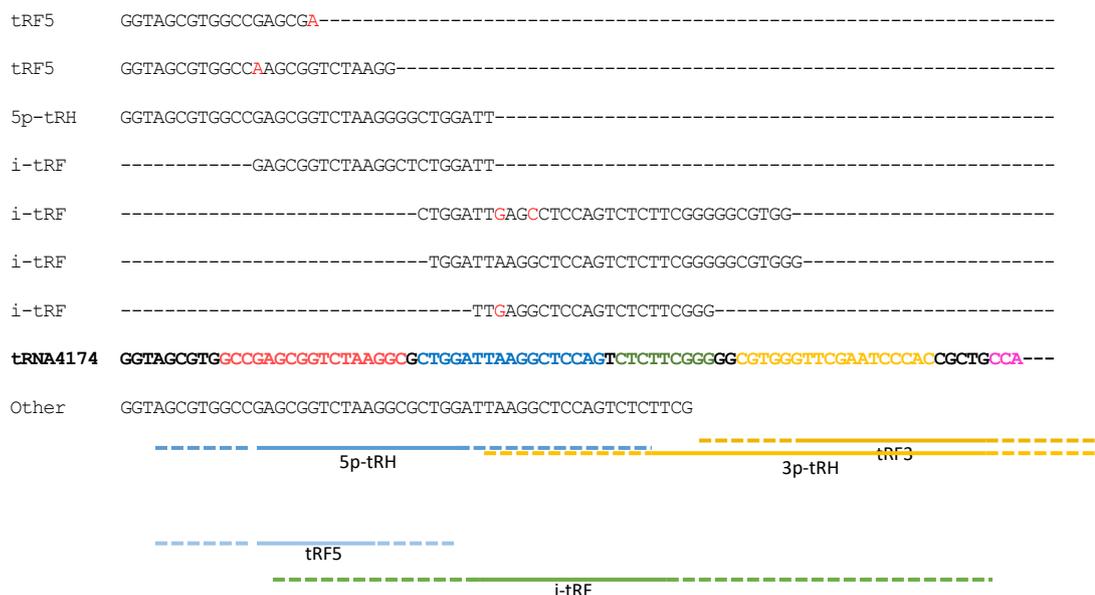


Additional Fig S4. The distribution of reads across 18S (**a**) and 28S (**b**) rRNAs shows a particular pattern made of several peaks and read-poor sub regions, suggesting that rRNAs are fragmented by selective RNA cleavage.

a)

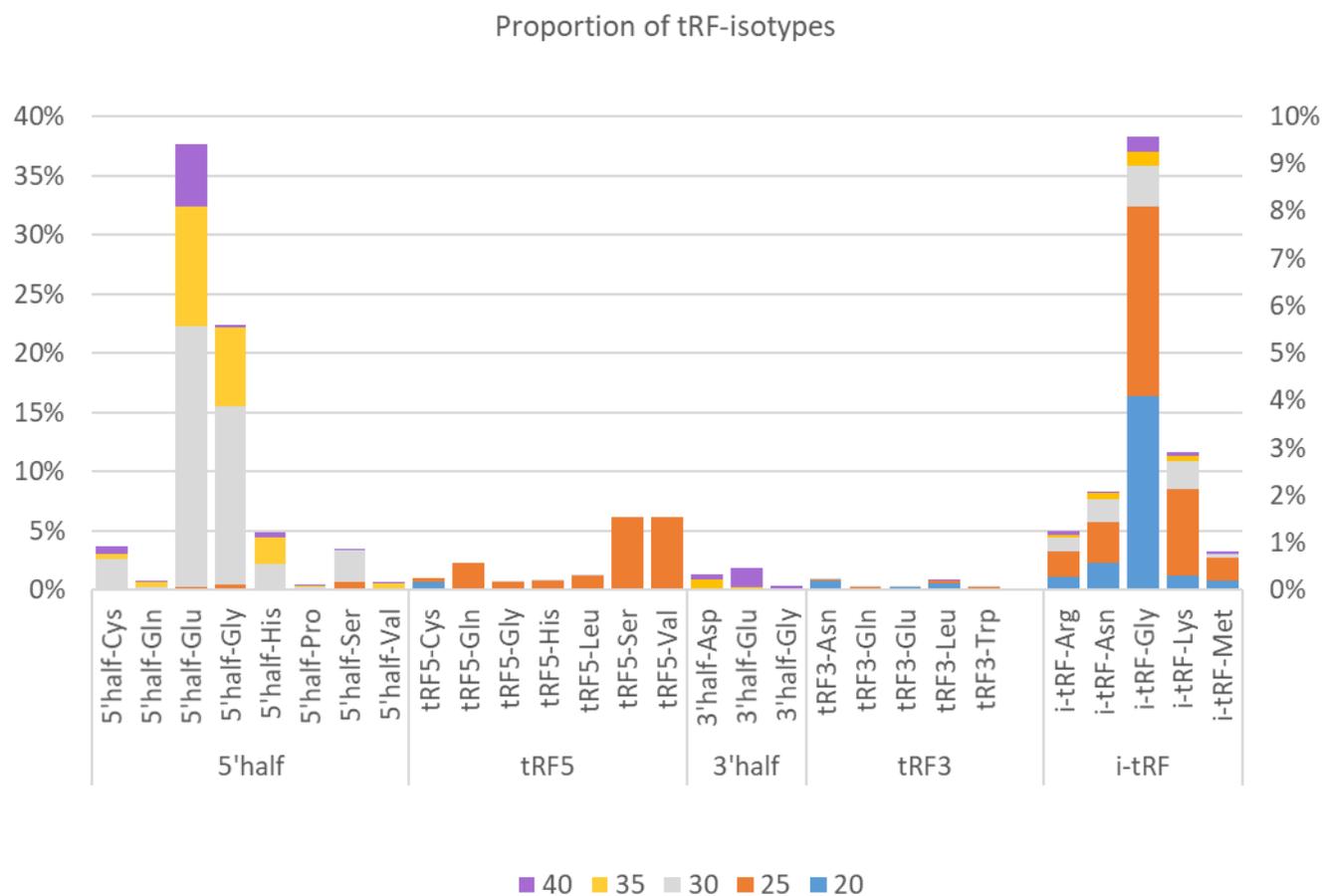


b)

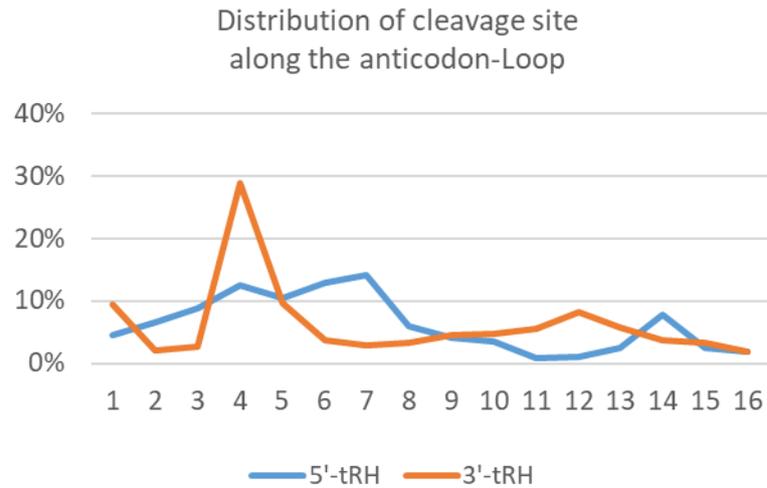


Additional Fig S5: **(a)** Biogenesis of tRFs. The cloverleaf structure of a tRNA typically contains a D-Loop, an Anticodon-Loop, a variable loop, a T-loop, and an amino acid acceptor stem. Several endonucleases can cleave tRNAs at specific sites, generating tRFs of different categories: cleavage of an RNase at the D-loop or T-loop of a tRNA can generate a 5' or 3' tRFs, respectively. Dicer (DCR) endonucleases has been reported to cleave tRNA at D-Loop, T-Loop and the amino acid acceptor stem. RNase Z cleavage has been shown to produce tRF3s in a Dicer-independent manner. Moreover, Angiogenin (ANG), a member of the RNase A superfamily, was shown to cleave the Anticodon-Loop to produce tRNA halves upon stress stimuli. ANG is also able to cleave the T-loop of tRNAs. Other unknown RNases might also participate in tRFs generation. **(b)** Multiple tRNAs alignments along the tRNA4174-LeuAAG sequence (D-Loop in red, T-Loop in yellow, Variable-Loop in green and amino acid acceptor stem in violet), illustrating the five tRFs categories: tRF5 and tRF3 (~15–32nt fragments), 5'-tRHs and 3'-tRHs (30–35nt) and i-tRFs. Reads that didn't fall into these categories were classified as other, possibly including full-length tRNAs (limited to the first 50th nucleotides due to the sequencing protocol). Post-transcriptional modifications and or genetic polymorphisms may also affect tRFs (nucleotide depicted in red). Criteria used to define the categories are depicted below the multiple alignment. Full line indicate the mandatory region that should be covered by the tRF, while dot line indicate the range that can be covered by the tRF.

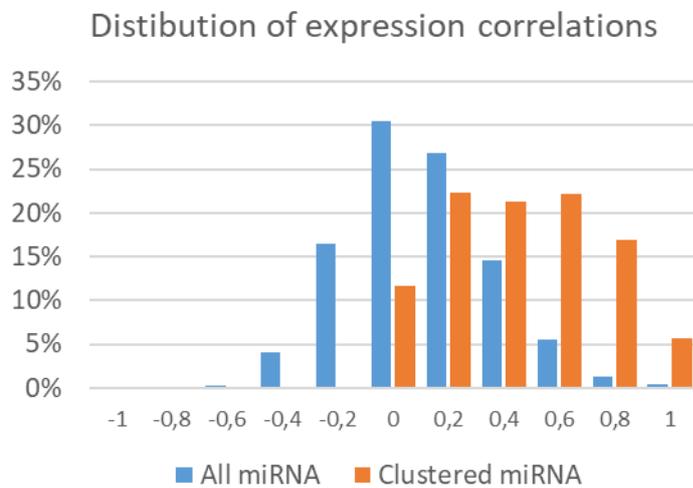
a)



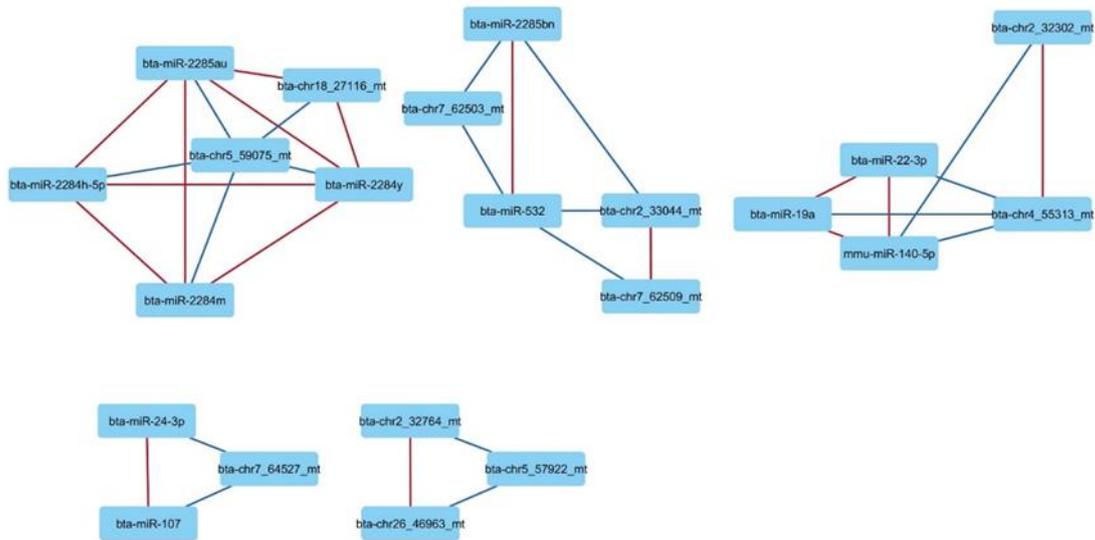
Additional Fig S6 : Frequency of tRFs isotypes. The percentage of read counts was computed for each tRF and each associated anticodon, according to the tRF length. Only the most expressed tRFs are reported in the histogram. The left axis (0-40%) refers to 5p-tRHs, which are the most expressed tsRNAs. The right axis (0-10%) refers to the other tsRNAs.



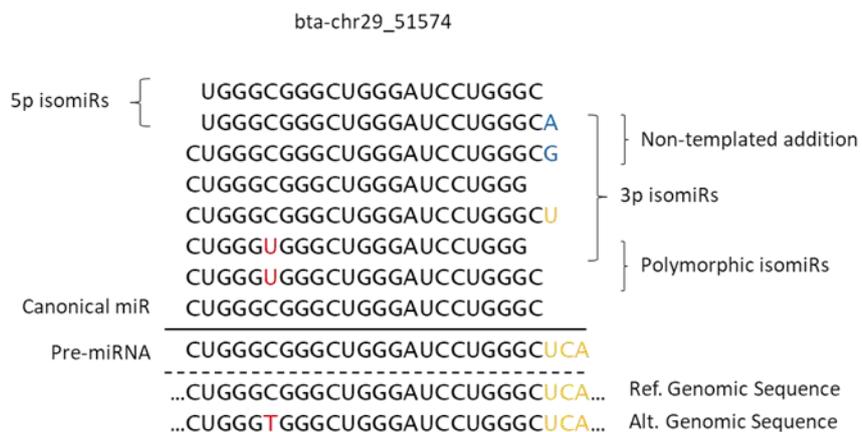
Additional Fig S7 Distribution of cleavage site along the anticodon-Loop, for both 5p-tRHs and 3p-tRHs, showing a bias towards 5p of the anticodon and a high frequency of cleavage at the 4<sup>th</sup> and 7<sup>th</sup> nucleotides.



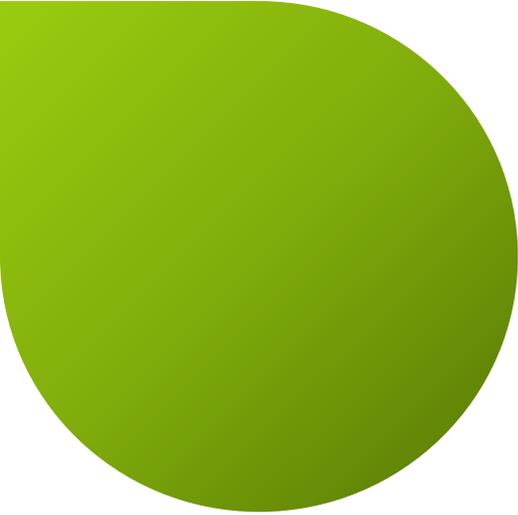
Additional Fig S8 : Distribution of Pearson correlation coefficients computed for the top 1580 miRNAs having a mean expression level above 10 in at least one breed (all miRNA) and clustered miRNAs. Correlations of a miRNA with itself were omitted. Stronger correlations were observed between genomically clustered miRNA compared to non-clustered miRNAs;



Additional Fig S9 : Five putative regulation networks, identified based on correlations above |0.7|. A search was performed to identify specific patterns of correlation (above |0.7|) indicative of putative functional regulation (miR-x → miR-y → miR-z implies negative correlation between miR-x and miR-y as well as miR-y and miR-z, while a positive correlation is expected between miR-x and miR-z).



Additional Fig S10 : Classification and examples of isomiRs for bta-chr29\_51574



# Communications

## 1) Publications hors « article(s) de thèse »

Perrier, J. P., Sellem, E., Prézélin, A., Gasselin, M., Jouneau, L., Piumi, F., ... & Le Danvic, C. (2018). A multi-scale analysis of bull sperm methylome revealed both species peculiarities and conserved tissue-specific features. *BMC genomics*, *19*(1), 1-18.

Gobé, C., Elzaïat, M., Meunier, N., André, M., Sellem, E., Congar, P., ... & Pailhoux, E. (2019). Dual role of DMXL2 in olfactory information transmission and the first wave of spermatogenesis. *PLoS genetics*, *15*(2), e1007909.

Rossitto, M., Marchive, C., Pruvost, A., Sellem, E., Ghetts, A., Badiou, S., ... & Boizet-Bonhoure, B. (2019). Intergenerational effects on mouse sperm quality after in utero exposure to acetaminophen and ibuprofen. *The FASEB Journal*, *33*(1), 339-357.

Perrier, J. P., Kenny, D. A., Chaulot-Talmon, A., Byrne, C. J., Sellem, E., Jouneau, L., ... & Fair, S. (2020). Accelerating onset of puberty through modification of early life nutrition induces modest but persistent changes in bull sperm DNA methylation profiles post-puberty. *Frontiers in Genetics*, *11*, 945.

Beaujean, N., Boutinaud, M., Devinoy, È., Jammes, H., Le Guillou, S., Le Provost, F., ... & Kiefer, H. (2020). L'épigénétique et la construction du phénotype chez le bovin. *Production animales*.

C.M. O'Meara, E. Henrotte, K. Kupisiewicz, C. Latour, M.L.W J Broekhuijse, A. Camus, L. Gavin-Plagne, E. Sellem. The effect of adjusting settings within a Computer Assisted Sperm Analysis (CASA) system on bovine sperm motility and morphology results. *En cours d'écriture*

## 2) Présentations orales lors de congrès internationaux

Congres international de la Société d'Andrologie de la Langue Française – 21 janvier 2021 / **Invited speaker**. Introduction de la session « petits ARN non codants ».

QualiVets, octobre 2020 – Animation de conférences en tant que **Président**. Standardisation Européenne de la mesure de la qualité de la semence.

Animal Sperm Research Center (ASRC) – conference. Jerusalem University. April 2020 / **Invited speaker**. A comprehensive overview of bull sperm-borne small non-coding RNAs and their diversity across breeds.

AI-Vet 2019, 4-5 octobere 2019, Bilund Denmark / **Invited speaker**. Standardization of semen quality assessment.

Alberta Epigenetic Network, 8 March 2019, Calgary, Canada / **Invited speaker**. Bull sperm sncRNAs: A new source for potential fertility biomarkers?

International Bull Fertility Conference, 27-30 May 2018, Westport, Ireland / **Speaker**. Bull sperm sncRNAs: A new source for potential fertility biomarkers?



**Titre :** Les petits ARN non codants du spermatozoïde bovin : de potentiels biomarqueurs de la fertilité mâle ?

**Mots clés :** sncRNA, spermatozoïde, bovin, fertilité

**Résumé :** L'utilisation de taureaux subfertiles entraîne des pertes de production, une augmentation du taux de réforme des vaches pour infertilité et un gaspillage des ressources. C'est pourquoi le contrôle qualité de la semence utilisée en Insémination Animale est une préoccupation majeure des entreprises de sélection. Divers protocoles de contrôle qualité ont été développés au cours des dernières décennies, exploitant des biomarqueurs et des paramètres fonctionnels clés mesurés par cytométrie en flux. Toutefois, la capacité prédictive de ces tests reste limitée et ne permet pas d'identifier les taureaux subfertiles. C'est pourquoi des travaux ont été entrepris dans le cadre du LabCom SeQuaMol, afin d'explorer l'intérêt de l'épigénome spermatique comme biomarqueurs de fertilité chez les taureaux.

Ainsi, les travaux menés au cours de cette thèse ont porté sur les petits ARNs non codants (sncRNA), dont le rôle au cours de la spermatogenèse et de la fécondation a été suggéré / démontré. L'amélioration des connaissances sur le contenu en sncRNA du spermatozoïde bovin éjaculé avait comme objectif pratique de développer un outil permettant de fiabiliser la prédiction en routine de la fertilité des taureaux. Les travaux ont été organisés selon trois axes : l'établissement du contenu exhaustif en sncRNA du sperme bovin éjaculé, l'étude de la dynamique d'expression de ces sncRNAs au cours de la maturation épидидymaire et l'identification *in fine* de biomarqueurs associés à la fertilité des taureaux.

**Title:** The bovine sperm small non-coding RNAs: Potential biomarkers of male fertility?

**Keywords:** sncRNA, sperm, bovine, fertility

**Abstract:** Most dairy cows are inseminated and subfertile bulls can have an overall dramatic negative impact on the sustainability of the dairy sector. The bull fertility prediction, through efficient quality control (QC) procedures is thus crucial for the cattle breeding sector. Over the last decades, several biomarkers and quality control procedures and have been proposed to guarantee semen fertility, including assessment of key functional parameters using flow cytometry. However, their relevance for routine QC has yet to be ascertained and their predictive value is too limited to confidently discard subfertile bulls. Thus, studies have been conducted by LabCom SeQuaMol to explore to what extent the bull sperm epigenome may be relevant to improve fertility prediction.

In this respect, this PhD thesis was focused on small noncoding RNAs (sncRNAs), which have been shown to play a role in spermatogenesis and fertilization. We aimed at improving the scientific knowledge on bull sperm sncRNA content in order to develop a routine tool to predict bull fertility. The work has been organized according to three main themes: providing a comprehensive overview of cattle sperm sncRNA, studying their expression dynamics during sperm maturation along epididymis and identifying biomarkers associated with bull fertility.