



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 16485

To cite this version :

Nevot, Alizée. *Prédiction des statuts métabolique et de santé de chevaux d'endurance via l'intégration de données « omiques » : vers l'identification de biomarqueurs*. Thèse d'exercice, Médecine vétérinaire, Ecole Nationale Vétérinaire de Toulouse - ENVT, 2016, 176 p.

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@inp-toulouse.fr.

PRÉDICTION DES STATUTS MÉTABOLIQUE ET DE SANTÉ DE CHEVAUX D'ENDURANCE VIA L'INTÉGRATION DE DONNÉES «OMIQUES» : VERS L'IDENTIFICATION DE BIOMARQUEURS

THESE
pour obtenir le grade de
DOCTEUR VÉTÉRINAIRE

DIPLOME D'ÉTAT

*présentée et soutenue publiquement
devant l'Université Paul-Sabatier de Toulouse*

par

NEVOT Alizée

Née le 16 mars 1991, à Rennes (35)

Directeur de thèse : M. Alain DUCOS

JURY

PRESIDENT :
M. Patrick CALVAS

Professeur à l'Université Paul-Sabatier de TOULOUSE

ASSEESSEURS :
M. Alain DUCOS
Mme Sophie PRADIER

Professeur à l'École Nationale Vétérinaire de TOULOUSE
Maître de Conférences à l'École Nationale Vétérinaire de TOULOUSE

MEMBRE INVITE :
Mme Núria MACH

Chargée de Recherches à l'INRA (UMR GABI, Jouy-en-Josas) Maître de stages

Répartition des Enseignants-Chercheurs par Département.

Mise à jour : 06/09/2016

DIRECTRICE : ISABELLE CHMITELIN

ELEVAGE ET PRODUITS/SANTÉ PUBLIQUE VÉTÉRINAIRE	SCIENCES BIOLOGIQUES ET FONCTIONNELLES	SCIENCES CLINIQUES DES ANIMAUX DE COMPAGNIE, DE SPORT ET DE LOISIRS
<p>Responsable : M. SANS</p> <p><u>ALIMENTATION ANIMALE :</u> M. ENJALBERT Francis, PR Mme PRIYMENKO Nathalie, MC Mme MEYNADIER Annabelle, MC</p> <p><u>EPIDEMIOLOGIE :</u> Mathilde PAUL, MC</p> <p><u>MALADIES REGLEMENTEES-ZOONOSES- MEDECINE PREVENTIVE DES CARNIVORES DOMESTIQUES-DROIT VETERINAIRE :</u> M. PICAUVET Dominique, PR</p> <p><u>PARASITOLOGIE-ZOOLOGIE :</u> M. FRANC Michel, PR M. JACQUIET Philippe, PR M. LIENARD Emmanuel, MC Mme BOUHSIRA Emilie, MC</p> <p><u>HYGIENE ET INDUSTRIE DES ALIMENTS :</u> M. BRUGERE Hubert, PR M. BAILLY Jean-Denis, PR Mme BIBBAL Delphine, MC Mme COSTES Laura, AERC Mme DAVID Laure, MCC</p> <p><u>PATHOLOGIE DE LA REPRODUCTION :</u> M. BERTHELOT Xavier, PR M. BERGONIER Dominique, MC Mme CHASTANT-MAILLARD Sylvie, PR Mme HAGEN-PICARD Nicole, PR M. NOUVEL Laurent-Xavier, MC Mme MILA Hanna, MC</p> <p><u>PATHOLOGIE DES RUMINANTS :</u> M. SCHELCHER François, PR M. FOUCRAS Gilles, PR M. CORBIERE Fabien, MC M. MAILLARD Renaud, MC M. MEYER Gilles, PR</p> <p><u>PRODUCTION ET PATHOLOGIE AVIAIRE ET PORCINE :</u> Mme WARET-SZKUTA Agnès, MC M. JOUGLAR Jean-Yves, MC M. GUERIN Jean-Luc, PR M. LE LOC'H Guillaume, MC</p> <p><u>PRODUCTIONS ANIMALES AMELIORATION GENETIQUE ECONOMIE :</u> M. DUCOS Alain, PR M. SANS Pierre, PR M. RABOISSON Didier, MC</p>	<p>Responsable : Mme GAYRARD</p> <p><u>ANATOMIE :</u> M. MOGICATO Giovanni, MC M. LIGNEREUX Yves, PR Mme DEVIERS Alexandra, MC</p> <p><u>ANATOMIE PATHOLOGIQUE - HISTOLOGIE :</u> M. DELVERDIER Maxence, PR Mme LETRON-RAYMOND Isabelle, MC Mme BOURGES-ABELLA Nathalie, PR Mme LACROUX Caroline, PR</p> <p><u>BIOLOGIE MOLECULAIRE :</u> Mme BOUCLAINVILLE-CAMUS Christelle, MC</p> <p><u>MICROBIOLOGIE – IMMUNOLOGIE - MALADIES INFECTIEUSES :</u> M. MILON Alain, PR M. BERTAGNOLI Stéphane, PR M. VOLMER Romain, MC Mme BOULLIER Séverine, MC Mme DANIELS Héléne, MC</p> <p><u>BIostatISTIQUES :</u> M. CONCORDET Didier, PR M. LYAZRHI Faouzi, MC</p> <p><u>PHARMACIE-TOXICOLOGIE :</u> M. PETIT Claude, PR Mme CLAUW Martine, PR M. GUERRE Philippe, PR M. JAEG Philippe, MC</p> <p><u>PHYSIOLOGIE –PHARMACOLOGIE THERAPEUTIQUE :</u> M. BOUSQUET-MELOU Alain, PR Mme GAYRARD-TROY Véronique, PR Mme FERRAN Aude, MC M. LEFEBVRE Hervé, PR</p> <p><u>BIOCHIMIE :</u> Mme BENNIS-BRET Lydie, MC</p> <p><u>ANGLAIS :</u> M. SEVERAC Benoît, PLPA Mme MICHAUD Françoise, PCEA</p>	<p>Responsable : Mme CADIERGUES</p> <p><u>ANESTHESIOLOGIE</u> M. VERWAERDE Patrick, MC</p> <p><u>CHIRURGIE :</u> M. AUTEFAGE André, PR M. ASIMUS Erik, MC M. MATHON Didier, MC Mme MEYNAUD-COLLARD Patricia, MC Mme PALIERNE Sophie, MC</p> <p><u>MEDECINE INTERNE :</u> Mme DIQUELOU Armelle, MC M. DOSSIN Olivier, MC Mme LAVOUE Rachel, MC Mme GAILLARD-THOMAS Elodie, MCC</p> <p><u>OPHTALMOLOGIE :</u> M. DOUET Jean-Yves, MC</p> <p><u>DERMATOLOGIE :</u> Mme CADIERGUES Marie-Christine, PR</p> <p><u>IMAGERIE MEDICALE</u> M. CONCHOU Fabrice, MC</p> <p><u>BIOLOGIE MOLECULAIRE :</u> Mme TRUMEL Catherine, PR</p> <p><u>PATHOLOGIE DES EQUIDES :</u> M. CUEVAS RAMOS Gabriel, MC Mme PRADIER Sophie, MC Mme LALLEMAND Elodie, AERC</p>

REMERCIEMENTS

A Monsieur le Professeur Patrick Calvas

Professeur de l'Université Paul Sabatier de Toulouse
Praticien Hospitalier
Génétique Médicale

Qui nous a fait l'honneur d'accepter la présidence de ce jury de thèse.
Hommages respectueux.

A Monsieur le Professeur Alain Ducos

Professeur de l'Ecole Nationale Vétérinaire de Toulouse
Productions Animales - Génétique

Qui nous a fait l'honneur de diriger cette thèse et qui nous a guidé dans son élaboration,
Pour son soutien, sa patience et sa gentillesse,
Qu'il trouve ici le témoignage de notre plus sincère reconnaissance.

A Madame le Docteur Sophie Pradier

Maître de Conférences de l'Ecole Nationale Vétérinaire de Toulouse
Médecine Interne des Equidés

Qui a aimablement accepté de faire partie de notre jury de thèse.
Sincères remerciements.

A Madame le Docteur Núria Mach

Chargée de Recherches à l'Institut National de Recherche Agronomique de Jouy-en-Josas
Génomique Animale et Biologie Intégrative – Biologie Intégrative et Génétique Equine

Qui a initié la partie expérimentale de ce travail, nous a fait l'honneur de nous associer à son étude, et nous a encadré dans l'élaboration de cette thèse.
Pour sa gentillesse, sa disponibilité, son soutien et ses précieux conseils.
Remerciements sincères et chaleureux.

A Madame le Professeur Céline Robert

Professeur et Maître de Conférences de l'Ecole Nationale Vétérinaire d'Alfort
Médecine Sportive des Equidés

Qui nous a aimablement permis de participer au projet GenEndurance en nous orientant vers cette étude.
Pour ses conseils avisés et sa disponibilité.
Sincères remerciements.

A Monsieur le Docteur Eric Barrey

Directeur de Recherche à l'Institut National de Recherche Agronomique de Jouy-en-Josas
Génomique Animale et Biologie Intégrative – Biologie Intégrative et Génétique Equine

Qui nous a aimablement intégré à l'équipe de Biologie Intégrative et Génétique Equine et qui nous a également permis de participer au projet GenEndurance.
Sincères remerciements.

TABLE DES MATIERES

INTRODUCTION	9
I. ETUDE BIBLIOGRAPHIQUE	13
1. L'intégration de différentes sources de données en vue d'une prédiction phénotypique	13
a. Les données phénotypiques	13
i. Le phénotype et ses différentes échelles	13
ii. Le phénotype et les approches phénotypiques	16
iii. Des exemples illustrant les approches de phénotypage à haut débit	17
> Le Mouse Phenome Project (M.P.P.)	17
> Le Human Phenome Project (H.P.P.)	21
iv. Les implications des campagnes de phénotypages actuelles : standardisation des mesures et ontologie des caractères	22
b. Les autres données « omiques » et leur production	24
i. Les données génomiques (génomique)	25
> La constitution d'un assemblage de référence dans les différentes espèces (séquence de référence)	26
> Le « re-séquençage » des espèces animales	28
> Le développement d'outils de génotypage à très haut débit : puces <i>SNP</i> (<i>SNP beadchips</i> , analyse de la variabilité du génome), à distinguer des puces d'expression / puces de gènes (<i>DNA microarrays</i> , analyse de l'expression du génome)	29
> Des connaissances affinées grâce aux innovations dans le domaine du séquençage	31
ii. Les données épigénomiques (l'épigénomique)	33
iii. Les données transcriptomiques et microtranscriptomiques (le transcriptome et le microtranscriptome)	35
> Le transcriptome	35
Encart : les micro ARN	40
> Les outils de la transcriptomique	41
iv. Les données protéomiques (le protéome)	48
v. Les données métabolomiques (le métabolome)	53
c. Un zoom sur les données de génomique, transcriptomique, et autres « omiques » équine, et leur application en médecine vétérinaire (génomique clinique et autres sciences omiques équine appliquées)	56
i. Les débuts de la génomique équine : séquence de référence, assemblage	57
ii. Les outils disponibles	59

iii.	Les méthodes pertinentes pour utiliser ces outils biomoléculaires dans l'étude de caractères génétiques	63
iv.	Les applications principales des outils génomiques chez le cheval	64
	> La génomique au service du dépistage d'affections génétiques	65
	o L'identification de gènes responsables d'affections monogéniques	65
	o Un manque d'informations épidémiogénétiques	68
	o L'opportunité du conseil génétique grâce au testage	69
	> La génomique au service du diagnostic clinique et de la médecine personnalisée	70
	o La prédisposition aux affections « complexes »	70
	o Des outils de diagnostic et de pronostic	71
	o Les outils thérapeutiques du futur ?	78
	> La génomique au service de la gestion du cheval athlète	79
	o La caractérisation des aptitudes	79
	o Des biomarqueurs pour le suivi de l'état physiologique et le contrôle antidopage	82
	o La nutrigenomique et l'optimisation de la nutrition	83
d.	Quelques applications (exemples) d'utilisation des données « omiques » dans d'autres espèces que l'espèce équine	84
e.	Les méthodologies statistiques utilisées pour l'intégration de différentes sources de données « omiques »	88
	Encart : La biologie intégrative	88
2.	La prédiction phénotypique appliquée aux performances sportives d'endurance	95
a.	Chez l'Homme : à la recherche de biomarqueurs de l'effort d'endurance	95
	i. Les effets d'un exercice physique intense de type « endurance »	96
	ii. Les biomarqueurs de l'effort d'endurance, ou quand les –omiques se mettent au service (de la prédiction) des performances sportives d'endurance	98
	iii. Un nouveau concept : la sportomique	102
b.	Chez l'animal athlète : exemple du cheval	103
	i. La détermination d'un phénotype sportif associé à l'endurance équine : les premiers pas du projet GenEndurance (2011-2014 ; C. Robert, E. Barrey)	105
	- Présentation des objectifs et synthèse du projet	105
	- Pertinence du cheval d'endurance comme modèle	106
	- Déroulement des 3 étapes successives et complémentaires du projet	107
	- Résultats préliminaires du projet	108
	o Première étape	109
	o Deuxième étape	113
	o Troisième étape	114
	ii. Etude sur les transcriptome (ARNm), microtranscriptome (miARN), métabolome (métabolites) et leurs liens lors d'un effort d'endurance chez le cheval	116

II. ETUDE EXPERIMENTALE	123
1. Introduction et objectifs de l'article	123
2. Article final publié dans <i>Scientific Reports</i>	124
3. Résultats, conclusion et perspectives	139
CONCLUSION	140
REFERENCES BIBLIOGRAPHIQUES	142
ANNEXES	
Glossaire	168

TABLE DES ILLUSTRATIONS

<i>Figure I : Les domaines conceptuels requis pour construire un phénotype.....</i>	<i>14</i>
<i>Figure II : L'intégration de sources de données phénotypiques murines diverses (The Mouse Phenotype Database Integration Consortium, 2007)</i>	<i>19</i>
<i>Figure III : Améliorations récentes apportées à la MPD (sur les études d'intervention et la classification des méthodologies) (Grubb & al., 2014).....</i>	<i>20</i>
<i>Figure IV : Les membres de l'IMPC (http://www.mousephenotype.org/about-imp/imp-members).....</i>	<i>21</i>
<i>Figure V : Schéma récapitulatif des grandes étapes de la construction d'ATOL (Le Bail & al., 2014).....</i>	<i>23</i>
<i>Figure VI : Représentation schématisée de l'intégration des données issues des différents concepts en « omes » et des réseaux « multi-omiques » regroupant différentes étapes de l'expression génomique, ici appliquée aux études phénotypiques à cellule unique (Hossain & al., 2015 ; Hocquette & al., 2009).....</i>	<i>24</i>
<i>Figure VII : Les objectifs de la génomique (Schibler, 2012).....</i>	<i>25</i>
<i>Figure VIII : Les puces de génotypage (Schibler, 2012; http://www.hendrix-genetics.com/fr-fr/).....</i>	<i>30</i>
<i>Figure IX : Séquenceur MiSeq, Illumina (http://www.illumina.com).....</i>	<i>31</i>
<i>Figure X : Séquenceur PacBio RS II, Pacific Biosciences (http://www.pacb.com)</i>	<i>32</i>
<i>Figure XI : Représentation quantitative des gènes d'ARN non codants parmi les gènes identifiés dans le génome humain (Lamoril & al., 2011).....</i>	<i>39</i>
<i>Figure XII : Puce Affymetrix (GeneChip®), zoom sur l'aspect du microarray (https://www.affymetrix.com)</i>	<i>41</i>
<i>Figure XIII : Principe d'analyse sur microarray (Bogard & al., 2008)</i>	<i>42</i>
<i>Figure XIV : Aspect d'un microarray.....</i>	<i>42</i>
<i>Figure XV : Caractéristiques d'une puce GeneChip®, société Affymetrix (Bogard & al., 2008).....</i>	<i>43</i>
<i>Figure XVI : Analyse sur puce Affymetrix (Bogard & al., 2008).....</i>	<i>44</i>
<i>Figure XVII : Analyse comparative de l'expression de gènes par hybridation compétitive de deux échantillons de sources différentes (Bogard & al., 2008).....</i>	<i>45</i>
<i>Figure XVIII : Analyse des images dans le cadre d'une hybridation compétitive (Bogard & al., 2008)</i>	<i>46</i>
<i>Figure XIX : Séquenceur Ion Proton™, Thermo - Life Technologies (https://www.thermofisher.com).....</i>	<i>47</i>
<i>Figure XX : Les principaux outils de la transcriptomique / génomique fonctionnelle (Schibler, 2012).....</i>	<i>48</i>
<i>Figure XXI : Les technologies protéomiques contemporaines et leur organisation (Cho, 2007).....</i>	<i>51</i>
<i>Figure XXII : Exemple de spectre métabolomique RMN d'un échantillon plasmatique (San Cristobal & al., 2014)</i>	<i>54</i>
<i>Figure XXIII : Etude métabolomique utilisant la RMN dans le cadre d'un traitement de bovins aux anabolisants (JC Portais, Cours de Métabolomique, Université Paul Sabatier (Toulouse)).....</i>	<i>56</i>
<i>Figure XXIV : Twilight, 1^{ère} jument séquencée (https://www.uky.edu/Ag/Horsemap/hgphorses.html)</i>	<i>57</i>
<i>Figure XXV : Illumina EquineSNP50 Beadchip (http://www.illumina.com)</i>	<i>59</i>
<i>Figure XXVI : Puce d'expression commercialisée par Agilent (http://www.agilent.com).....</i>	<i>61</i>
<i>Figure XXVII : Gene expression custom microarray, Agilent (http://www.agilent.com)</i>	<i>62</i>
<i>Figure XXVIII : Spectre métabolomique de l'urine équine en RMN 1H 700 MHz 1D (Escalona & al., 2015)</i>	<i>76</i>

<i>Figure XXIX : Diagramme de Venn mettant en évidence les similitudes et différences métaboliques entre compartiments biologiques (Escalona & al., 2015)</i>	<i>77</i>
<i>Figure XXX : Histogramme représentant le génotype MSTN optimal par rapport à la distance de course (Hill & al., 2010a)</i>	<i>81</i>
<i>Figure XXXI : Différence de profil d'expression des miARN dans le sang en fonction de l'exercice</i>	<i>92</i>
<i>Figure XXXII : Principe de classification graphique par une méthode utilisant les machines à vecteur de support</i>	<i>93</i>
<i>Figure XXXIII : Détermination de genres bactériens potentiellement clés pour le poids vif et la production d'sIgA chez les porcelets (Mach & al., 2015)</i>	<i>94</i>
<i>Figure XXXIV : Mensurations moyennes des chevaux arabes d'endurance (Robert & al., 2012)</i>	<i>109</i>
<i>Figure XXXV : Mise en évidence via la méthode de spectroscopie RMN de différences entre les profils métabolomiques sanguins avant l'exercice (BE) et après l'exercice (PE) (Le Moyec & al., 2014)</i>	<i>111</i>
<i>Figure XXXVI : Score plot des échantillons avant course calculé par le modèle sur des chevaux classés en fonction de leur vitesse moyenne pendant la course (Le Moyec & al., 2014)</i>	<i>112</i>
<i>Figure XXXVII : Schéma bilan explicatif d'une analyse intégrée de données -omiques dans le cadre de l'étude et de la prédiction de statuts métabolique et de santé de chevaux d'endurance</i>	<i>121</i>
<i>Tableau I : Offre industrielle pour les appareils de NGS comparée à du séquençage Sanger (Guéguen & al., 2015; http://sequencing.roche.com, http://www.illumina.com, https://www.thermofisher.com, http://www.pacb.com)</i>	<i>31</i>
<i>Tableau II : Principales affections monogéniques à mutation causale identifiée (Schibler, 2012; Bailey, 2014; site internet OMIA de l'Université de Sydney)</i>	<i>65</i>
<i>Tableau III : Tableau bilan des effectifs en fin d'année 2014 (Compte Rendu GenEndurance 2014)</i>	<i>108</i>

INTRODUCTION

L'exercice d'endurance peut être défini comme un exercice cardiovasculaire – tout comme la course, le ski de fond, le vélo, l'aérobic, la natation – réalisé pendant un laps de temps étendu (Joyner, Coyle, 2008). Les besoins physiologiques et biochimiques permettant la réalisation de ce type d'exercice induisent à la fois des réponses musculaires et systémiques. Les athlètes d'endurance exposent leur corps à des situations physiologiques extrêmes qui rompent l'homéostasie corporelle interne en bouleversant le fonctionnement normal d'organes et de systèmes. La capacité à réaliser un exercice physique sur des laps de temps prolongés ne requiert pas uniquement une machinerie métabolique et hormonale adéquate, mais aussi de l'énergie et des sources électrolytiques, ainsi qu'une réponse immunitaire contrôlée (Branth & al., 2009). Exercer une activité physique à très haut niveau pendant une longue période implique que le corps entier s'engage dans une réponse de défense via la synthèse de protéines de la phase aiguë (*acute phase proteins*, APP), la libération d'hormones, et les variations de l'équilibre métabolique et des fluides corporels. Il serait intéressant de connaître finement tous ces mécanismes.

Dans le milieu du cheval, l'endurance est une discipline sportive équestre regroupant de plus en plus d'adeptes. Ce sport équestre, qui connaît un essor dans les pays européens et arabes depuis la fin des années 1970, a pour objectif de parcourir de longues distances avec son cheval, avec ou sans limitation de vitesse. Il s'agit d'une discipline sportive comparable au marathon chez l'Homme, demandant un effort prolongé parfois extrême. En France, l'endurance équestre revêt une importance de plus en plus considérable. En effet, l'équipe de France d'endurance équestre s'illustre depuis de nombreuses années à un niveau international, avec une médaille d'argent par équipe aux Jeux Equestres Mondiaux (J.E.M.) de 2010 et aux Championnats du Monde en 2012. Par ailleurs, l'élevage français de chevaux d'endurance se situe actuellement parmi les meilleurs au monde : lors des J.E.M. de 2010, 8 des 12 premiers chevaux étaient nés en France, dont le 1^{er}, le 2^e, et le 4^e, et parmi les chevaux ayant participé aux Championnats du Monde en 2012, 45 des 147 chevaux participants (31%) étaient nés en France, dont 14 classés parmi les 20 premiers (70%).

Comme dans les autres disciplines équestres, les résultats dépendent du couple cheval / cavalier. Concernant le cheval, comme chez n'importe quel individu, son aptitude pour une discipline donnée (ou toute autre caractéristique phénotypique) résulte de la combinaison entre un patrimoine génétique particulier (son génotype), d'une part, et l'expression de ce dernier

(son phénotype) dans un contexte environnemental particulier (effets de milieu), d'autre part. Pour l'instant, l'évaluation du potentiel d'un jeune cheval repose sur le jugement de son phénotype (morphologie et allures essentiellement). Les éleveurs de chevaux d'endurance, les cavaliers d'endurance ainsi que leurs entraîneurs souhaitent disposer de chevaux présentant des caractéristiques favorables à la pratique de ce sport. La race Pur-Sang Arabe et les croisements qui en sont issus (demi-sang arabe, anglo-arabe...) semblent présenter ces caractéristiques, aussi bien en termes de morphologie (ils sont en général légers et de petite taille, et de conformation musculaire fine), que d'allures (capacité à tenir des vitesses moyennes allant jusqu'à 15 km/h sur des distances de 20 à 160 km), ou encore de métabolisme (capacités de résistance à l'effort et de récupération rapide et efficace, compatibles avec des efforts de longue durée, à la fois sur les plans cardio-vasculaire, respiratoire, musculaire, digestif...). Ces caractères sont effectivement très recherchés pour la pratique de l'endurance équestre. Enfin, leur tempérament alerte, volontaire et plein d'allant la plupart du temps, se prête bien à la pratique de cette discipline. Cependant, pour le moment, le morphotype à rechercher ou éviter chez le cheval d'endurance est encore bien mal déterminé, en raison du faible nombre d'études menées sur ce sujet, de la jeunesse de la discipline, et de l'évolution très importante de cette dernière depuis les années 1990, modifiant considérablement les efforts demandés au cheval et donc les qualités requises pour être performant.

A la différence d'autres sports équestres faisant intervenir principalement force et souplesse naturelle équine, l'endurance sollicite très fortement le métabolisme. En effet, lors d'une course d'endurance (de 160 km notamment), les chevaux sont soumis à une réponse de stress résultant principalement de l'ajustement à la déshydratation, au déséquilibre métabolique et à l'inflammation. Il est donc très probable qu'un effort physique d'endurance soit à l'origine d'effets d'adaptation mesurables sur les paramètres « omiques » (transcriptome, microtranscriptome et métabolome) de différents tissus de ces chevaux. Ce sont précisément sur ces effets et leurs relations que je vais me pencher, c'est-à-dire sur l'analyse fonctionnelle des paramètres « omiques » et leur intégration, dans le cadre du projet de recherche en génétique GenEndurance (C. Robert, E. Barrey). Ce projet a globalement pour objectifs de déterminer les combinaisons de variants génétiques, l'expression du génome et les caractéristiques phénotypiques (biochimie, métabolisme, morphologie, allures) favorables ou défavorables à la performance en course d'endurance, afin de proposer aux éleveurs, cavaliers et entraîneurs des outils pratiques et peu invasifs d'aide à une caractérisation du potentiel de leurs chevaux la plus précise et la plus précoce possible (avant 6 ans ou avant la castration). Cela permettrait de conserver la suprématie de la filière endurance équine en France, de valoriser au mieux la

production de ce type de chevaux, et d'offrir aux cavaliers français des montures de qualité, les plus adaptées à leurs demandes.

La première partie de la Thèse, bibliographique, permettra de présenter l'état actuel des connaissances sur l'intérêt et les méthodes d'intégration de différentes sources de données « omiques » afin de prédire un phénotype particulier, notamment concernant l'endurance : on présentera les diverses données « omiques » dont on dispose et la manière dont elles sont produites, puis on donnera quelques exemples d'utilisation de ces données, avant de s'intéresser aux méthodes statistiques nécessaires à leur analyse et à leur intégration. On se concentrera ensuite sur les principes d'étude et de prédiction d'un phénotype sportif associé à l'endurance, comme c'est le cas dans le projet GenEndurance. On s'intéressera pour cela aux études déjà réalisées chez l'Homme dans l'objectif d'identifier des biomarqueurs de l'effort d'endurance, puis à celles menées sur le modèle animal équin, en détaillant le projet GenEndurance et en passant en revue les études concernant les modifications du transcriptome (ensemble des ARNm transcrits à partir des gènes de l'individu), du microtranscriptome (ensemble des microARN codés à partir de l'information génétique de l'individu) et du métabolome (ensemble des métabolites présents dans un échantillon biologique) et leur lien lors d'un effort d'endurance chez les chevaux.

Dans une deuxième partie, nous présenterons les résultats expérimentaux d'un travail expérimental auquel j'ai été associée. L'article scientifique original, publié le 10 Mars 2016 dans la revue *Scientific Reports* et intitulé « *Integrated mRNA and miRNA expression profiling in blood reveals candidate biomarkers associated with endurance exercise in the horse* » est intégré dans la Thèse, accompagné d'un texte introductif précisant le contexte et les objectifs de ce travail.

Les limites actuelles de ce genre de travaux, les développements attendus, les perspectives qu'ils ouvrent, seront détaillés dans la conclusion de cette Thèse.

PREMIERE PARTIE :

Etude bibliographique

I. ETUDE BIBLIOGRAPHIQUE

1. L'intégration de différentes sources de données en vue d'une prédiction phénotypique

Le phénotype à prédire peut être considéré à différentes échelles, qui seront présentées. Les données « omiques » utilisées pour la prédiction sont par ailleurs de multiples natures (phénomiques, génomiques, épigénomiques, transcriptomiques, microtranscriptomiques, protéomiques, métabolomiques...). Nous définirons ces différentes « strates » d'information et présenterons les principales stratégies mises en œuvre jusqu'à présent pour les produire. Ces « omiques » ainsi définis, nous étudierons leurs domaines d'application ; ils peuvent effectivement être utiles dans la production de connaissances diverses et variées : comment s'en sert-on, quel est leur intérêt ? Enfin, on explicitera les méthodologies statistiques les plus couramment utilisées afin d'organiser, d'assembler, d'intégrer ces multiples sources d'information.

a. Les données phénomiques

i. Le phénotype et ses différentes échelles

Le phénotype, terme créé par Johannsen en 1909 (*Wanscher, 1975*), désigne l'ensemble des caractères observables d'un individu. Plus précisément, il s'agit de la manifestation observée de l'expression de l'information génomique propre à un individu, dans le milieu où ce dernier évolue (*Freimer, Sabatti, 2003*). Le phénotype peut donc être défini comme l'ensemble apparent des performances résultant de l'expression du génome et des effets de milieu sur différents caractères (*Faucon-Lahalle, Montagu, 2011*). Beaucoup d'efforts ont été consacrés à l'identification des bases génotypiques expliquant la variabilité phénotypique, et désormais les études se concentrent sur les points à améliorer concernant la façon d'intégrer ces efforts à l'investigation de l'influence environnementale sur les phénotypes (*Freimer, Sabatti, 2003*).

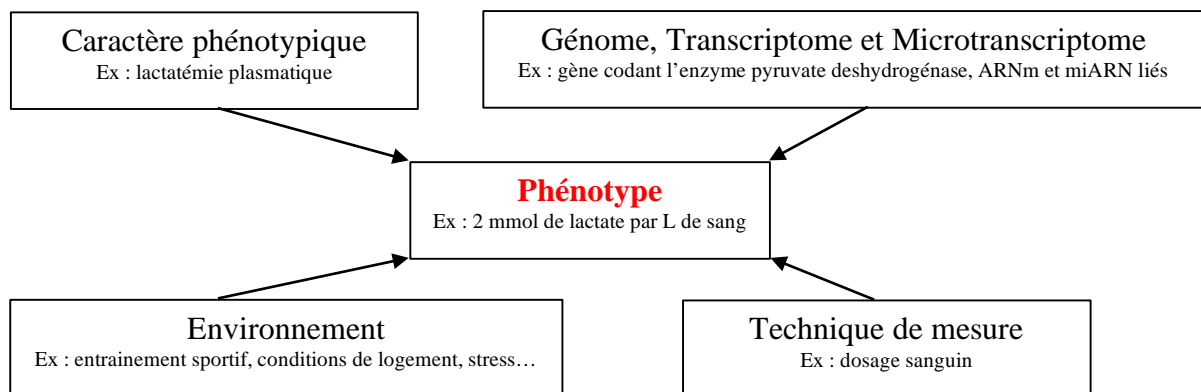


Figure 1 : Les domaines conceptuels requis pour construire un phénotype

Au-delà de cette définition du phénotype au sens large, on peut utiliser ce terme de manière plus restrictive et classer de différentes manières les phénotypes : selon l'échelle considérée (de ce qui est visible à ce qui nécessite des outils d'observation microscopique), selon le critère d'étude (caractère d'intérêt lui-même ou bien paramètre qui lui est lié), selon la nature du phénotype (en fonction du déterminisme génétique sous-jacent), selon la structuration du phénotypage (horizontal ou vertical en fonction du nombre d'individus considérés et de la finesse de leur caractérisation), selon l'intérêt du phénotype, autrement dit la finalité de son utilisation (zootechnique, médical, sportif...). Intéressons-nous donc à ces différents classements et à leurs caractéristiques.

On peut considérer le phénotype concernant un seul caractère (phénotype macroscopique), à l'échelle cellulaire (phénotype cellulaire) ou encore à l'échelle moléculaire (phénotype moléculaire). Le terme « phénotype » peut donc se rapporter à toute caractéristique morphologique, biochimique, physiologique ou comportementale d'un organisme (Freimer, Sabatti, 2003).

Un caractère phénotypique peut être le caractère d'intérêt lui-même (caractère « vitesse maximale atteinte au cours d'une course d'endurance donnée » par exemple), un bio-marqueur (ou indicateur) qui lui est corrélé (caractère « lactatémie », lié au métabolisme anaérobie des fibres musculaires), ou un effecteur qui le contrôle (expression du gène codant l'enzyme pyruvate déshydrogénase pour cet exemple).

Si l'on s'intéresse au cheval, au caractère phénotypique « robe de base » peut correspondre l'état de phénotype « bai ». Ce caractère phénotypique est discret (ou discontinu) car il présente un nombre de phénotypes limité (13 robes de base chez le cheval) : noir, bai,

isabelle, souris, alezan, gris... Dans ce cas, le phénotype est régi par un déterminisme génétique relativement simple (Rieder, 2009). D'autres caractères phénotypiques sont qualifiés de continus car ils varient de manière continue dans un intervalle borné par une valeur basse et une valeur haute : il s'agit par exemple de la taille, ou du poids d'un individu. Un même phénotype peut dépendre de l'expression de plusieurs gènes et de leurs interactions, ainsi que des marques épigénétiques et du milieu ; c'est donc la résultante de l'expression du génome (incluant génome et épigénomes) modulée ou non par le milieu environnant. C'est par exemple le cas du potentiel d'un cheval à être performant en endurance : cela dépend en effet de paramètres physiologiques intrinsèques (capacités pulmonaires, cardio-respiratoires, musculaires...) mais aussi des conditions de vie du cheval (entraînement, alimentation, logement...), de ses interactions avec les congénères et avec les humains (niveau de stress, pathologies...).

Il est également possible de structurer le phénotype de deux façons différentes mais complémentaires : horizontale et verticale. Le phénotypage horizontal est caractérisé par des mesures réalisées aisément sur un très large panel d'animaux. On peut citer, par exemple, les dispositifs officiels de contrôle de performance en ferme dans les différentes espèces de ruminants (contrôle laitier ou contrôle de croissance des jeunes), ou le programme de phénotypage « PhénoFinLait », qui consiste, entre autres, en une mesure de la teneur en protéines et matières grasses du lait des ruminants par spectrométrie moyen infrarouge (MIR) et une mise au point d'équations de prédiction, puis une mise en lien de ces phénotypes prédictifs avec le génotype des animaux (Esvan & al., 2010). Le phénotypage vertical, quant à lui, réunit de nombreuses mesures de plusieurs phénotypes de plus en plus fins concernant une (ou plusieurs) famille(s) de caractères (phénotypage très approfondi) sur peu d'individus. Le programme européen SafeBeef, dans le domaine de la qualité de la viande bovine, en est un exemple : il s'agit de mesures simultanées des caractéristiques des carcasses (conformation, engraissement), des critères de qualité sensorielle des viandes (tendreté, flaveur, jutosité), de la composition fine du muscle (type de fibre, teneur en collagène et en lipides, composition en acides gras) et de l'expression des gènes et des protéines impliquées dans le déterminisme de la qualité de la viande (Miles, Caswell, 2008; O'Connor, Wyness, 2012).

On peut aussi faire la distinction entre phénotypes d'intérêts différents : on peut par exemple étudier le phénotype d'aptitude au saut d'obstacle d'un cheval de sport, qui est un phénotype macroscopique particulier. On peut aussi mentionner la capacité du cheval à se déplacer à certaines allures particulières, comme l'amble. En effet, la plupart des chevaux ne se déplacent naturellement qu'à trois allures : pas, trot et galop. Certaines races équines présentent la capacité de se déplacer de manière naturelle à des allures supplémentaires. Ainsi,

le cheval Islandais dispose de deux allures supplémentaires : le tölt et l'amble. L'amble est une allure caractérisée par le poser alternatif de chaque bipède latéral (et non pas diagonal comme c'est le cas au trot) avec une phase de projection entre le poser de chaque bipède. Une étude récente a montré que cette allure avait un déterminisme génétique : une mutation causale dans le gène *DMRT3* (*DMRT3_Ser301STOP*) est ainsi responsable de l'expression d'un codon stop prématuré et donc d'une protéine Dmrt3 raccourcie (c'est une situation où le mécanisme de dégradation de l'ARN après un codon stop prématuré, appelé « nonsense-mediated mRNA decay » (NMD), n'est pas déclenché. (*Silva, Romão, 2009*)). Or, cette protéine joue un rôle majeur dans l'établissement des connexions (lors du développement embryonnaire) puis dans la régulation des neurones spinaux contrôlant la locomotion et la coordination du mouvement des membres chez les vertébrés. La présence de cette protéine raccourcie chez les chevaux islandais homozygotes A/A pour la mutation non-sens dans *DMRT3* entraîne à l'échelle phénotypique la capacité naturelle de se déplacer à une allure supplémentaire qu'est l'amble (*Andersson & al., 2012*). Les chevaux peuvent également apprendre à ambler, et maîtriser cette allure par le dressage : on voit donc qu'un phénotype (« capacité de se déplacer à l'amble ») dépend à la fois du génotype (homozygotie pour la mutation *DMRT3_Ser301STOP*) et des conditions du milieu (apprentissage locomoteur par le dressage).

ii. Le phénomène et les approches phénomiques

Le « phénomène » d'un individu ou organisme représente l'ensemble des « caractères phénotypiques », ou « traits phénotypiques » observables aux niveaux éthologique, morphologique, anatomique, physiologique, moléculaire, cellulaire, tissulaire, etc., qui le caractérisent. Le phénotypage d'un individu consiste à déterminer l'ensemble de ses caractères observables. Un « langage de référence » (une ontologie), sur lequel nous reviendrons par la suite, est nécessaire pour décrire ces derniers de manière précise et standardisée, afin de pouvoir ensuite comparer différents animaux et/ou résultats d'études indépendantes sur la base de ces caractères. En effet, les phénotypes, tout comme les génotypes (ensemble des combinaisons alléliques d'un individu) peuvent, par définition, différer d'un individu à un autre, et c'est sur cette variabilité que se penche la phénomique (*Freimer, Sabatti, 2003*). Les approches phénomiques reposent donc sur la collecte d'informations phénotypiques pour chaque individu à différents niveaux (moléculaire, cellulaire, tissulaire, de l'organisme entier) puis sur la détermination de la manière dont ces caractères peuvent être étudiés ensemble de manière profitable.

iii. Des exemples illustrant les approches de phénotypage à haut débit

➤ Le Mouse Phéno Project (M.P.P.)

Le modèle murin s'est avéré jusqu'à présent fort utile dans la compréhension de la biologie en général, et de la biologie humaine en particulier. En effet, la facilité d'expérimentation qu'offre cette espèce (intervalle de génération très court, prolificité élevée, coût individuel faible...), la possibilité de réalisation de nombreuses analyses phénotypiques simultanées sur un même animal, le contrôle aisé des facteurs environnementaux contribuant aux phénotypes, la possibilité d'opérer des modifications ciblées du génome dans des cellules embryonnaires souches et de générer de nouveaux phénotypes par mutagenèse, la disponibilité du génome murin désormais intégralement séquencé, ainsi que la relative similarité entre les génomes humain et murin, font de la souris un modèle expérimental exceptionnel et sont autant de paramètres favorables justifiant l'exploitation de ces modèles animaux dans le cadre de l'investigation en biologie humaine.

La contribution des données murines aux avancées de la recherche en biologie, génétique et médecine humaine a donc été majeure, et est particulièrement bien illustrée par le Mouse Phenome Project. La souris est la seule espèce animale disposant d'un programme de phénotypage à haut débit aussi abouti. Ce projet a été initié il y a une quinzaine d'années pour compléter les efforts de séquençage du génome murin en promouvant de nouvelles initiatives de phénotypage dans des conditions standardisées et en rassemblant les informations dans une base de données centralisée à accès universel, la Mouse Phenome Database (*MPD*). Le M.P.P. regroupait, dès ses débuts en 2003, environ 400 mesures concernant différents domaines phénotypiques liés à l'athérosclérose, aux troubles sanguins, à la susceptibilité aux cancers, aux troubles nerveux et comportementaux, aux défauts de fonction sensorielle, à la vulnérabilité aux maladies, à la réactivité pulmonaire, à l'hypertension, à l'ostéoporose et à l'obésité, et recensait déjà plus de 150 000 *SNP* (*Single Nucleotide Polymorphism*, mutation ponctuelle d'une base nucléotidique dans la séquence d'ADN) identifiés dans un sous-ensemble de souches murines majeures, qui contribuaient à la *MPD* (Bogue, 2003).

Les données phénotypiques de souches murines sont actuellement stockées sous des formes multiples et variées (descriptions détaillées de lignées murines, données de phénotypage sur les nouvelles mutations, données sur les traits « normaux » de lignées consanguines, etc.), à de nombreux endroits : le problème typique de ces bases de données biologiques est qu'elles contiennent des informations similaires mais non identiques, sur différents sites répartis dans le monde entier (*The Mouse Phenotype Database Integration Consortium, 2007*). Un processus a donc été initié par le *Mouse Phenotype Database Integration Consortium (MPDIC)* dans le but d'une utilisation harmonieuse et la plus efficace possible de ces groupes de données : il est caractérisé par le développement de standards de description de phénotypes (utilisant un vocabulaire de description commun, sous forme d'ontologies) et de formats de fichiers pour la description des protocoles de phénotypage et des jeux de données phénotypiques. Pour cela, le *MPDIC* a établi un site web expérimental comme site coordinateur centralisant les informations et fournissant les liens vers les sites individuels donnant accès aux données de phénotypage murines (<http://www.interphenome.org/>). Ce processus de fusion, ou du moins, d'intégration de différentes bases de données phénotypiques murines est en cours ; il réunit, entre autres, les informations fournies par :

- la *MPD*, basée au Jackson Laboratory (*Grubb & al., 2004*) (<http://phenome.jax.org/>)
- le système *MuTrack*, développé par le Laboratoire National d'Oak Ridge (*Baker & al., 2004*)
- les bases de données *EuroPhenome* (<http://www.europenome.org>) et *EMPreSS* (*European Mouse Phenotyping Resource of Standardised Screens*) (<http://empres.har.mrc.ac.uk/>) du MRC (Medical Research Council) Harwell, centre international pour la génétique murine localisé près d'Oxford (*Mallon & al., 2007*)
- PhenoSITE, développé par le centre de bioressources financé par le gouvernement japonais RIKEN BRC (BioResource Center) (<http://ja.brc.riken.jp/lab/bpmp/index.html>)
- GeneNetwork, de l'université du Tennessee (<http://www.genenetwork.org>) (*Chesler & al., 2004*)
- PHENOMIN, infrastructure française développée grâce à une collaboration entre l'Institut Clinique de la Souris (ICS, Illkirch), le centre de Transgénése et Archivage d'Animaux Modèles (TAAM : CDTA-SEAT-CIPA, Orléans, Villejuif) et le Centre d'ImmunoPHÉnomique (CIPHE, Marseille) (<http://www.phenomin.fr>).

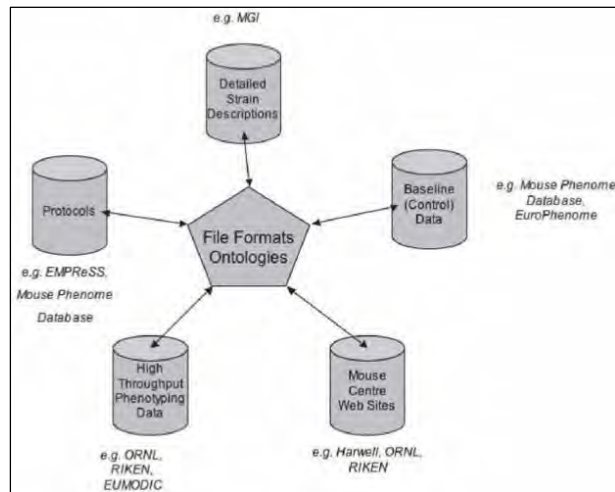


Figure II : L'intégration de sources de données phénotypiques murines diverses (The Mouse Phenotype Database Integration Consortium, 2007)

Ces données vont des descriptions détaillées fournies par le groupe Mouse Genome Informatics (MGI) aux descriptions d'études de phénotypage haut débit comme EUMODIC et descriptions synthétiques de lignées mutantes détenues par de nombreux sites, en passant par les données quantitatives sur les phénotypes « normaux » de lignées consanguines stockées dans la MPD et dans Europhenome.

Parmi ces sources d'information, nous présenterons l'exemple de la *MPD* afin d'avoir une idée plus concrète du mode de fonctionnement de ces bases de données. La *MPD* a vu le jour en 2001 au sein du Jackson Laboratory (organisation indépendante à but non lucratif) en tant que centre de coordination pour le M.P.P., et est en partie financée par les apports du NIH (National Institutes of Health). Cette base de données accessible en ligne intègre des données phénotypiques quantitatives, transcriptomiques (c'est-à-dire d'expression génétique) et génotypiques murines dans un cadre commun annoté, pour faciliter les recherches comparatives sur la santé et la pathologie humaines, en hébergeant et en mettant à disposition une grande richesse de données standardisées et soigneusement organisées caractérisant des souches murines : elle aide ainsi à s'affranchir de certains problèmes comme l'impossibilité d'expérimentation sur les humains pour des raisons pratiques ou éthiques (Maddatu & al., 2012; Grubb & al., 2014). La *MPD* contient désormais plus de 3500 mesures ou traits phénotypiques en lien avec la santé humaine, dont les processus cancéreux, le vieillissement, les troubles cardiovasculaires, l'obésité, la susceptibilité aux maladies infectieuses, les troubles sanguins, les troubles neurosensoriels, la toxicomanie... La plupart des données phénotypiques se trouve sous forme d'« enquêtes de lignées » (comparaisons de 10 à 40 souches murines couramment utilisées), et les données génotypiques sont principalement de type « génotypes *SNP* ». Les groupes de données sont fournis volontairement par des scientifiques provenant d'institutions et de cadres divers et variés, ou parfois extraits de ressources publiques par l'équipe de la *MPD*. Cette dernière entretient donc toute une palette de données de référence standardisées aidant les

chercheurs à choisir des lignées murines appropriées pour de nombreuses applications de recherche ; elle héberge des données de témoins et de traitements pour les études pharmacologiques et d'autres interventions ; elle offre une plateforme, elle-aussi standardisée, pour l'étude des relations génotype-phénotype ; et elle fournit, via une interface web, des outils d'analyse et de visualisation des données, pour les tests d'hypothèses par exemple. Les protocoles, les conditions expérimentales et l'historique environnemental de l'animal accompagnent chaque jeu de données.

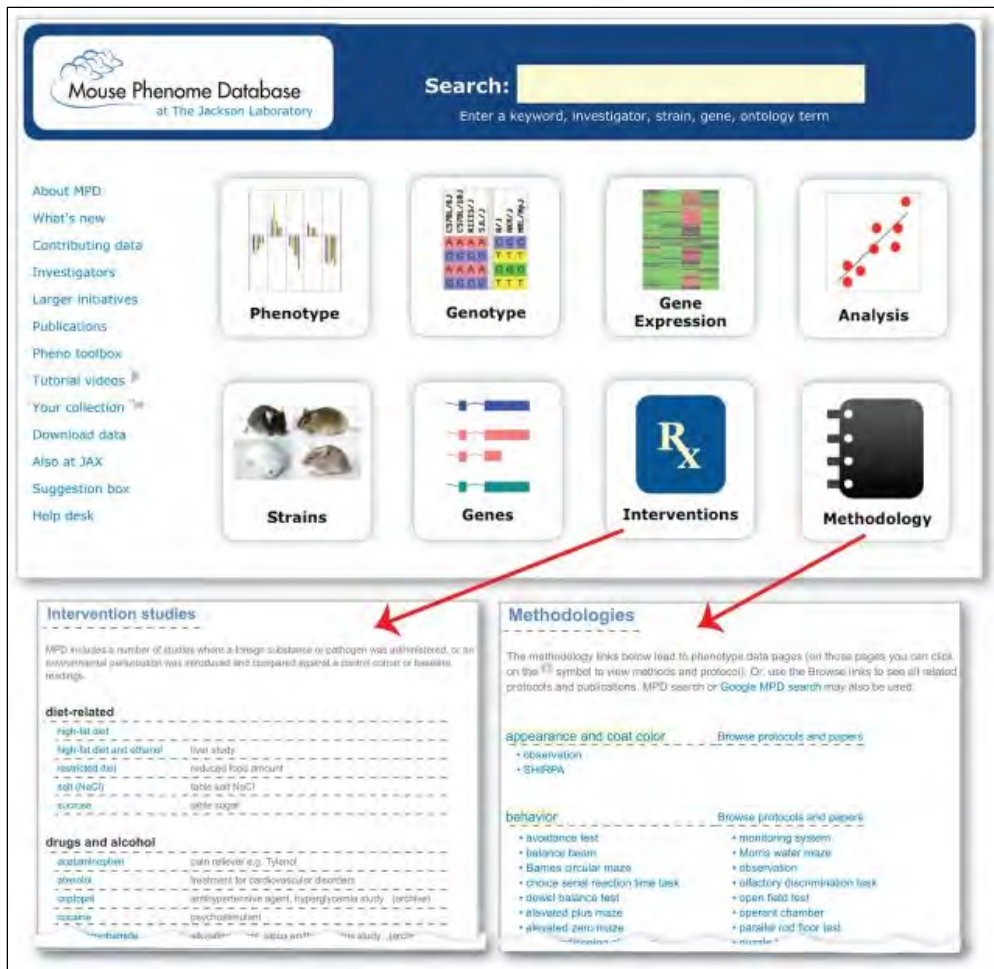


Figure III : Améliorations récentes apportées à la MPD (sur les études d'intervention et la classification des méthodologies) (Grubb & al., 2014)

Les acteurs cités précédemment sont, en majorité, regroupés au sein de l'*International Mouse Phenotyping Consortium (IMPC)* (<http://www.mousephenotype.org>). Ce réseau collaboratif en pleine expansion est actuellement composé de 18 institutions de recherche et financé par 5 organismes nationaux, et a pour but de développer un plan mobilisant les programmes et infrastructures de recherche murins internationaux majeurs, dans un effort stratégique et coordonné pour entreprendre un projet de phénotypage systématique pangénomique généralisé

de souris mutées, afin de fournir à la plus vaste communauté scientifique possible une ressource pérenne d'information génomique fonctionnelle chez les mammifères sous forme d'une base de données centralisée et à accès libre.

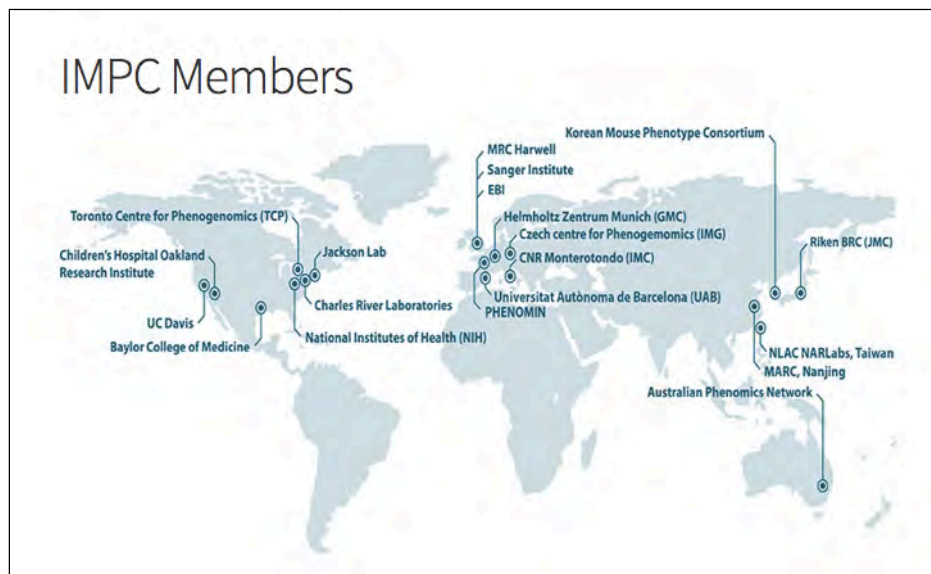


Figure IV : Les membres de l'IMPC (<http://www.mousephenotype.org/about-imp/imp-members>)

Toutes ces bases de données et les organismes à l'origine de leur développement sont en évolution continue, en parallèle aux changements affectant le paysage de la recherche.

➤ Le Human Phenome Project (H.P.P.)

Le Human Phenome Project permet, lui aussi, de comprendre le concept de « phénomique » : en effet, il s'agit d'un projet qui repose sur un effort international (collaboration entre des organismes publics et des agences privées à but non lucratif, dont les NIH des Etats-Unis, le Wellcome Trust, et les industries pharmaceutiques et de bioscience) pour la création d'une base de données phénomique (à l'instar de ce qui a été réalisé chez la souris), afin de caractériser le plus finement possible le phénoque, c'est-à-dire de disposer d'une représentation phénotypique la plus complète de l'espèce humaine (Freimer & Sabatti, 2003).

Ce projet contribue à l'objectif essentiel de la recherche en génétique humaine consistant à identifier des génotypes spécifiques (variants génétiques) associés à des phénotypes humains particuliers (ensemble des maladies humaines, courantes ou rares, ayant une base génétique). Il mobilise à cet effet de nombreuses disciplines scientifiques dont l'investigation clinique, l'épidémiologie, la génétique/génomique, la phénomique comparative

(valorisation de données produites dans des espèces modèles comme la souris, le rat ou certaines espèces de primates non-humains), la (bio)statistique, la (bio)informatique, etc.. Pour cela, il est nécessaire de disposer de bases de données phénomiques puissantes, standardisées et exhaustives.

Parmi les caractères phénotypiques composant le phénomène humain se trouvent en première ligne les **diagnostics de maladie**, qui sont basés sur les traditions et l'observation cliniques, auxquelles il est nécessaire d'ajouter des études épidémiologiques pour étudier l'héritabilité de ces caractères.

La maladie étant un phénotype « extrême », la « pointe de l'iceberg » de la variation phénotypique existant au sein d'une population pour un phénotype particulier, les **phénotypes intermédiaires ou endophénotypes** permettent d'étudier tout ce champ de variation (on peut citer par exemple la concentration sérique en immunoglobulines E, qui n'est qu'un des nombreux endophénotypes pour le diagnostic clinique de l'asthme).

Les **mesures de phénotypes quantitatifs** et les **voies biologiques** liées à tel ou tel caractère phénotypique doivent également être déterminées.

Il faut définir des méthodes de **standardisation des mesures** phénotypiques et donc des méthodes de phénotypage également.

Enfin, la **phénomique comparative** permet de compléter la définition des caractères phénotypiques : les modèles animaux apportent une aide précieuse par les mesures des paramètres vitaux (respiratoires, rénaux, cardio-vasculaires, etc.) de chaque animal sous des conditions de référence ou bien de stress, liées à des données biochimiques et histologiques ainsi qu'à des profils d'expression génétique et des données génotypiques.

iv. Les implications des campagnes de phénotypages actuelles : standardisation des mesures et ontologie des caractères

Une telle évaluation phénomique, à grande échelle et à haut débit, nécessite la mise en place de plusieurs stratégies.

Tout d'abord, l'échelle internationale implique une **standardisation des protocoles** (l'obtention d'informations phénotypiques précises, fiables, répétables et comparables entre laboratoires, pays ou entreprises) et une coopération entre chercheurs (incluant les experts qui réalisent des évaluations directes et chronophages, mais aussi les individus « lambda »

provenant de partout dans le monde, avec leurs propres évaluations). Ainsi, le programme européen Welfare Quality® définit des mesures standardisées à mettre en place dans les élevages commerciaux pour estimer douze critères associés au bien-être animal (Blokhuis & al., 2010). Un autre exemple est le programme « Meat Standards Australia » (MSA) qui a permis de développer un système de prédiction de la qualité de la viande bovine (Watson & al., 2008) en constituant et exploitant une base de données formée par les résultats de dégustation par des consommateurs de différents muscles selon des modes de cuisson variés.

La standardisation des données phénotypiques nécessite au préalable de disposer d'un **langage commun, de référence**, avec des définitions partagées et non équivoques des caractères et de leurs modes de mesure : à cet effet, le programme « Animal Trait Ontology of Livestock » (ATOL) a été développé à l'INRA en collaboration avec l'« Iowa State University », afin de définir précisément les caractères phénotypiques d'intérêt des animaux d'élevage et leur organisation hiérarchique. Ainsi, la dernière version de l'ontologie ATOL (ATOL v6.0) définit 1848 caractères phénotypiques orientés autour des 7 branches principales d'intérêt zootechnique suivantes : « Bien-être Animal et Stress », « Nutrition et Alimentation », « Croissance et Qualité de la chair », « Mamelle et Production de lait », « Reproduction et Fertilité », « Œufs » et « Foie gras » (Le Bail & al., 2014).

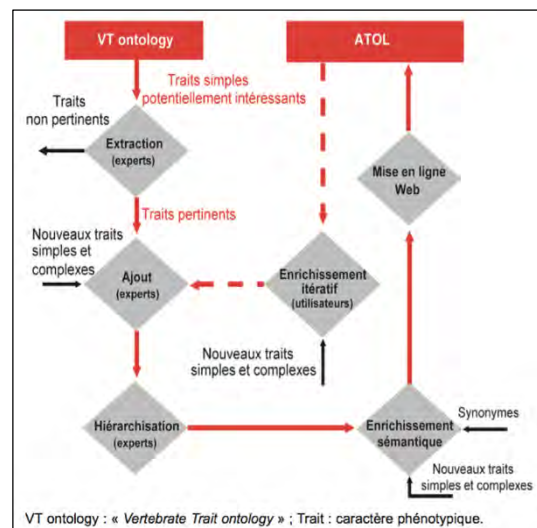


Figure V : Schéma récapitulatif des grandes étapes de la construction d'ATOL (Le Bail & al., 2014)

Quant au **haut débit**, il requiert d'augmenter le nombre de caractères phénotypiques prédicteurs de phénotypes d'intérêt (ces derniers étant parfois difficiles à appréhender par eux-mêmes) ou de phénotypes complexes (comme la robustesse ou la fertilité pour les animaux d'élevage) pouvant être mesurés par une analyse unique (donc de développer les analyses multiplexes automatisées et standardisées à partir d'échantillons biologiques comme le sérum) et d'adopter des procédures exigeant moins de travail ou une main d'œuvre moins abondante. En ce qui concerne la définition de nouveaux caractères (**précision** des relations fonctionnelles), le phénotypage à haut débit des animaux d'élevage peut être horizontal ou vertical, comme évoqué plus haut, et tout l'intérêt réside dans le croisement des données issues de ces deux approches complémentaires.

b. Les autres données « omiques » et leur production

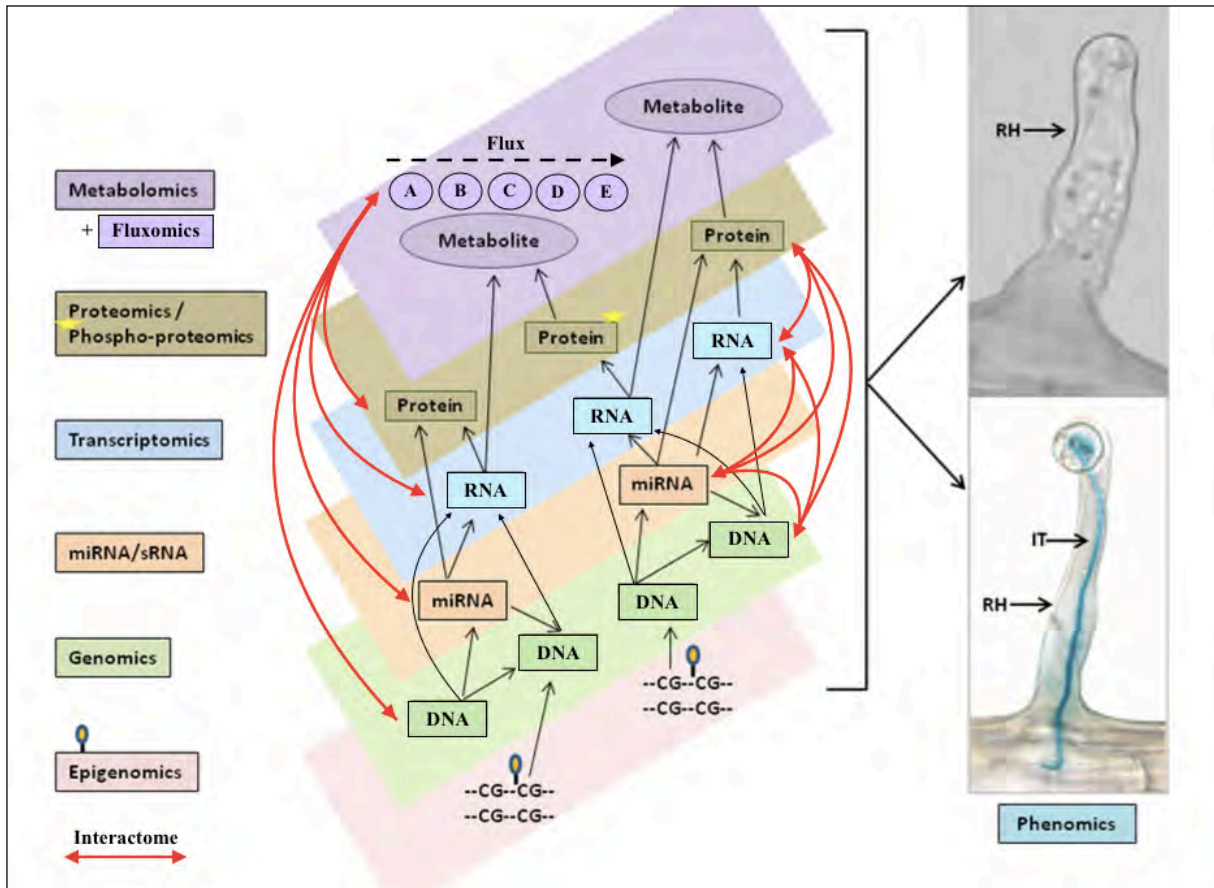


Figure VI. : Représentation schématique de l'intégration des données issues des différents concepts en « omique » et des réseaux « multi-omiques » regroupant différentes étapes de l'expression génomique, ici appliquée aux études phénomiques à cellule unique (RH = root hair cell = poil racinaire, cellule différenciée du rhizoderme (assise pilifère) d'un végétal ; IT = infection thread)

Génome = ensemble des gènes et des séquences intergéniques → **Génomique**

Epigénome = ensemble des marques épigénétiques → **Epigénomique**

Transcriptome = ensemble des transcrits (ARN) → **Transcriptomique**

Protéome = ensemble des protéines exprimées → **Protéomique**

Phospho-protéome = ensemble des protéines-kinases (protéines activatrices et activées par phosphorylation) exprimées → **Phospho-protéomique**

Métabolome = ensemble des métabolites présents dans un fluide biologique → **Métabolomique**

Fluxome = ensemble des flux métaboliques dans le contexte cellulaire → **Fluxomique**

Interactome = ensemble des interactions entre macromolécules d'une cellule (réseaux biologiques)

Phénome = ensemble des caractères phénotypiques → **Phénomique**

(Hossain & al., 2015; Hocquette & al., 2009)

Les données phénotiques étant définies et illustrées, nous allons maintenant présenter l'ensemble des autres « omiques » (données génomiques, épigénomiques, transcriptomiques, microtranscriptomiques, protéomiques et métabolomiques), en développant plus particulièrement la partie concernant les données de génomique et de transcriptomique dans le domaine équin.

i. Les données génomiques (génome)

La génomique est la science qui étudie la structure, le fonctionnement et l'évolution des génomes par des approches globales et systématiques, en bénéficiant désormais de technologies d'analyse à haut débit (Schibler, 2012). Elle se subdivise en deux disciplines complémentaires. La première est la **génomique fonctionnelle**, domaine scientifique que l'on peut situer entre la génétique et la physiologie (Hocquette & al., 2007) et qui vise à déterminer la fonction et l'expression du génome en caractérisant l'ensemble des transcrits (transcriptome), des protéines (protéome) et des métabolites (métabolome), ainsi que les éléments de la régulation de cette expression. Elle est à distinguer de la **génomique structurale**, dont l'objectif principal est de comprendre l'architecture et de l'évolution des génomes. Cette thèse s'appuie principalement sur des données issues de la génomique fonctionnelle, mais les approches de génomique fonctionnelle reposent sur des bases issues de la génomique structurale ; il est donc essentiel de les rappeler brièvement.

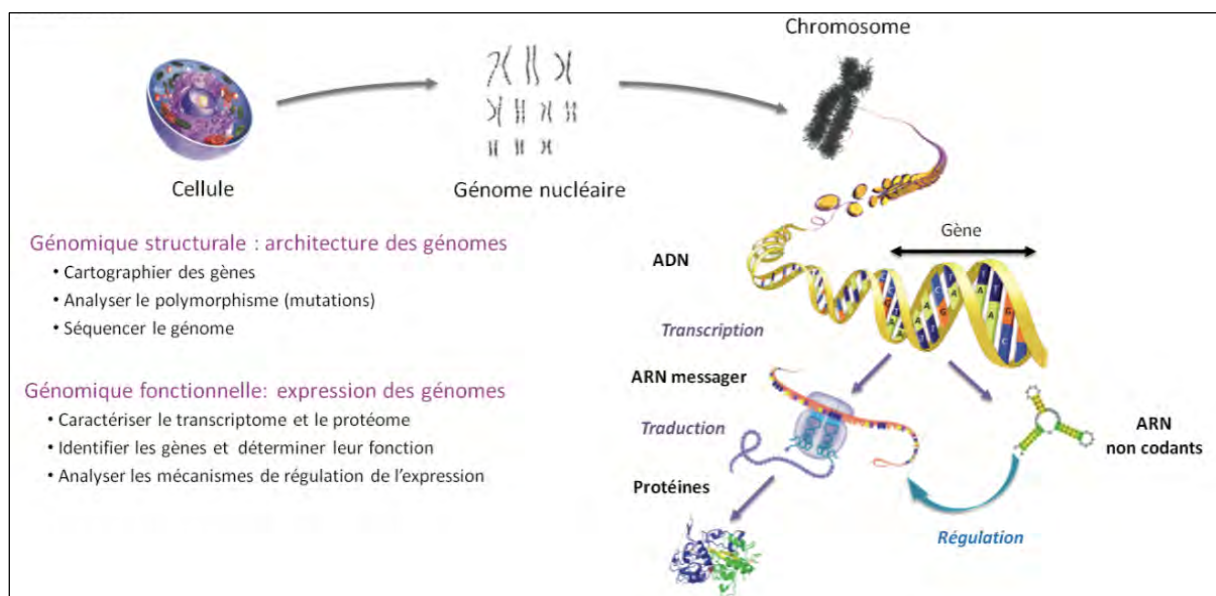


Figure VII : Les objectifs de la génomique (Schibler, 2012)

La génomique structurale s'intéresse à la localisation et à la structure des gènes et des génomes, ainsi qu'à la caractérisation de l'ensemble des variations génomiques telles que les *Single Nucleotide Polymorphisms (SNP)*, les insertions, les délétions, les insertions/délétions (« InDels »), les *Copy Number Variants (CNV)*, variations du nombre de copies de segments d'ADN pouvant ou non contenir des gènes), les réarrangements de gènes, les duplications, etc... Elle s'appuie sur différentes techniques de biologie moléculaire, dont l'utilisation croisée permet de mieux appréhender des mécanismes non encore élucidés (*Faucon-Lahalle, Montagu, 2011*). Les applications des outils et résultats de génomique structurale sont nombreuses : cartographie et identification / diagnostic de mutations responsables d'affections héréditaires (ou de prédisposition à des maladies particulières), de *QTL (Quantitative Trait Locus)*, ou locus de caractères quantitatifs, régions du génome contribuant à une partie de la variation de caractères quantitatifs), identification génétique, contrôles de filiation et médecine légale, études anthropologiques, etc...

Les données génomiques valorisées actuellement sont principalement issues de séquençage (séquençage de génomes entiers, d'exomes), et/ou de génotypage (sur puce *SNP* le plus souvent). La production de données de séquence a évolué de façon exponentielle depuis la fin des années 2000 en raison d'une rupture technologique sans précédent (développement de séquenceurs dits « de nouvelle génération », ou *NGS (Goodwin & al., 2016)*). Le développement d'outils de génotypage *SNP* à haut ou très haut débit a été permis par la constitution d'assemblages de référence pour les différentes espèces à partir du milieu des années 2000, et par le très important travail de « re-séquençage » effectué ultérieurement (projets « 1000 génomes »), permettant d'identifier de très nombreux polymorphismes génomiques (près de 30 millions de locus *SNP* ont été identifiés sur le génome bovin, par exemple (*Daetwyler & al., 2014*)). Ces outils de génotypage à (très) haute densité sont désormais utilisés de manière intensive dans les espèces d'élevage, notamment dans le cadre des programmes de sélection génomique. Ces différents points sont repris et détaillés ci-dessous.

➤ La constitution d'un assemblage de référence dans les différentes espèces (séquence de référence)

Deux niveaux de séquençage peuvent être considérés : le séquençage *de novo* (séquençage « *princeps* » d'un génome pour lequel il n'existe pas de séquence de référence), et

le « reséquençage », qui est le séquençage de tout ou partie du génome d'autres individus suivi de la comparaison du résultat obtenu avec celui de la séquence de référence connue ; les techniques utilisées sont généralement les mêmes dans les deux cas.

La stratégie utilisée pour les premiers séquençages complets *de novo* de génomes d'animaux (y compris de l'Homme) a consisté à produire un grand nombre de lectures par du séquençage aléatoire global (*Whole Genome Shotgun*, ou *WGS*), et à assembler ces lectures par des approches bioinformatiques pour former une séquence continue (contigs), puis des assemblages de contigs (supercontigs, ou *scaffolds*), tout en s'appuyant sur la cartographie génétique afin d'assigner les séquences aux chromosomes (Vignal, 2011).

Le séquençage complet du génome humain a donc été réalisé avec cette stratégie (Venter & al., 2001), et, parallèlement, à l'aide de centaines de séquenceurs classiques utilisant la méthode de synthèse enzymatique de Sanger.

Une fois établie la première séquence complète du génome humain, celles d'autres espèces de vertébrés ont rapidement été produites : les génomes de la souris et du rat, modèles majeurs pour la biologie humaine, ont ainsi été séquencés respectivement en 2002 (Chinwalla & al., 2002) et en 2004 (Gibbs & al., 2004). Cinq espèces de poissons à intérêt pour la biologie fondamentale ont également été séquencées entre 2001 et 2007. Ont suivi les génomes d'espèces à intérêt agronomique et de certaines espèces domestiques, bénéficiant de stratégies de production et d'analyse de données affinées, et de coûts de séquençage moindres, dans un laps de temps de l'ordre d'une ou deux années seulement. Ainsi, la poule (*Gallus gallus*) est le premier animal de rente à avoir été séquencé, en 2004 (*The International Chicken Genome Sequencing Consortium*, 2004), précédant les autres espèces de volaille telles que la dinde (*Meleagris gallopavo*) en 2010 (Dalloul & al., 2010), le canard (*Anas platyrhynchos*) (Kraus & al., 2011) et la caille japonaise (*Coturnix japonica*) (Kawahara-Miki & al., 2013). Le séquençage du génome du chimpanzé (*Pan troglodytes*) a été réalisé en 2005 (*The Chimpanzee Sequencing and Analysis Consortium*, 2005), avant celui de l'opossum (*Monodelphis domestica*) en 2007 (Mikkelsen & al., 2007a). Parmi nos espèces domestiques favorites, le chien (*Canis familiaris*), dont le séquençage était principalement justifié par son rôle de modèle expérimental des maladies humaines, a été séquencé en 2005 (Lindblad-Toh & al., 2005) ; le génome félin (*Felis catus*) a été séquencé à son tour en 2007 (Pontius & al., 2007) suite au *Mammalian Genome Project*, programme de séquençage partiel de 29 mammifères initié par le *Broad Institute* en 2005 (Lindblad-Toh & al., 2011). Ce projet, aussi connu sous l'appellation « *29 Mammals Project* », a permis la recherche de régions fonctionnelles grâce à la phylogénie : en alignant de multiples séquences de plusieurs espèces, des régions pour lesquelles la séquence est particulièrement conservée peuvent être

détectées (ces régions « fonctionnellement contraintes » aident à détecter des exons de gènes ou des régions non-codantes fonctionnelles). Le génome du lapin (*Oryctolagus cuniculus*) a aussi été partiellement séquencé dans le cadre de ce projet. Enfin, de plus en plus d'espèces animales ont été séquencées à leur tour : la séquence du génome bovin (*Bos taurus*) a été officiellement publiée en 2009 (*The Bovine Genome Sequencing and Analysis Consortium & al., 2009*), de même que celle du cheval (*Equus caballus*) (*Wade & al., 2009*). Le séquençage du génome des autres espèces d'élevage a rapidement suivi : la publication officielle concernant le génome porcin (*Sus scrofa*) a notamment eu lieu en 2012 (*Groenen & al., 2012*) ; un génome ovin (*Ovis aries*) virtuel initial, assemblé grâce à des séquences alignées sur les génomes humain, bovin et canin, a été publié en 2007 (*Dalrymple & al., 2007*), puis complété en 2010 grâce aux innovations dans le domaine du séquençage (*The International Sheep Genomics Consortium & al., 2010*).

➤ Le « re-séquençage » des espèces animales

La séquence de référence des différentes espèces ainsi déterminée, des variations interindividuelles de séquence (polymorphismes) peuvent être détectées en alignant des séquences d'individus ou de mélanges d'individus de populations différentes sur cette séquence de base. Les différences détectées dans ce contexte sont majoritairement de type *SNP*, mais peuvent aussi être des *InDels* (courtes insertions ou délétions nucléotidiques), des *CNV*, etc.. Bien que la plupart des polymorphismes n'aient pas ou peu de conséquences phénotypiques, leur étude fine par le re-séquençage permet d'une part de détecter puis de préciser les régions du génome impliquées dans l'expression de caractères phénotypiques, et d'autre part de comprendre l'histoire des populations naturelles ou sélectionnées. On peut ainsi identifier des régions du génome ayant subi des pressions de sélection importantes, au point d'entraîner une diminution locale de la variabilité (*Vignal, 2011*).

Le projet 1000 génomes illustre parfaitement ce concept : initié en 2008 et finalisé en 2015, ce programme de caractérisation du large spectre géographique et fonctionnel de la variabilité génomique humaine fournit une base solide d'aide à la compréhension de la contribution génétique aux maladies chez l'Homme. Cette description exhaustive des variations du génome humain a pu être réalisée grâce au séquençage « tout-génome » à basse couverture, au séquençage d'exomes et au génotypage dense par biopuces (ces termes seront précisés par la suite) de divers groupes d'individus issus de multiples populations. Deux articles récapitulant les différentes étapes et résultats du projet ont été publiés à son stade final, analysant les données issues du génotypage de 2 504 individus de 26 populations différentes (provenant de 5 régions

continentales distinctes), et fournissant ainsi la vue aujourd'hui la plus aboutie de la variabilité structurale génomique humaine mondiale sous forme de carte intégrée. Le premier (*The 1000 Genomes Project Consortium, 2012*) se concentre sur les variations courtes et relativement simples affectant jusqu'à 500 bases, mais expose aussi des petites mais complexes modifications (autres que des *SNP* et des *indels*) non étudiées auparavant. Le deuxième (*Sudmant & al., 2015*) explore quant à lui les modifications plus complexes affectant des portions chromosomiques plus larges, pouvant atteindre une longueur de 500 kb ; ces variants structuraux sont analysés bien plus en détail que ce qui était possible auparavant (grâce aux avancées technologiques considérables réalisées au cours de la dernière décennie), et leurs effets sont comparés à ceux des variants caractérisés par la substitution d'un nucléotide par un autre (*SNP*). L'article met également en lumière la localisation de variants structuraux au sein de régions génomiques associées à des caractères complexes ou à des maladies : ces nouvelles données fournissent donc un point de départ pour des études mécanistiques ultérieures, facilitant ainsi l'exploration de rares variants structuraux qui n'étaient pas accessibles auparavant. De plus, les individus étant issus de cinq régions continentales différentes, les données issues de leur génotypage sont à l'origine d'une vision beaucoup plus riche de la variabilité humaine physiologique que celle issue des jeux de données « euro-centrées » précédents (*The 1000 Genomes Project Consortium, 2012*). La création en 2015 de l'*International Genome Sample Resource* a permis la mise au point d'un centre de coordination de ces jeux de données à accès public, mais cet accès est fortement susceptible de devenir contrôlé à l'avenir (*Birney, Soranzo, 2015*).

En ce qui concerne les animaux, un projet 1000 génomes bovins a été calqué sur le modèle humain et 234 taureaux ont ainsi été séquencés, fournissant une base pour la compréhension de l'architecture génétique des caractères monogéniques et complexes dans cette espèce (*Daetwyler & al., 2014*). Ce programme a été construit dans l'optique d'accélérer le progrès génétique chez les animaux de rente tout en tenant compte de la santé et du bien-être animal par l'annotation des variants de séquence et le génotypage de taureaux ancestraux clés.

- Le développement d'outils de génotypage à très haut débit : puces *SNP* (*SNP beadchips*, analyse de la variabilité du génome), à distinguer des puces d'expression / puces de gènes (*DNA microarrays*, analyse de l'expression du génome)

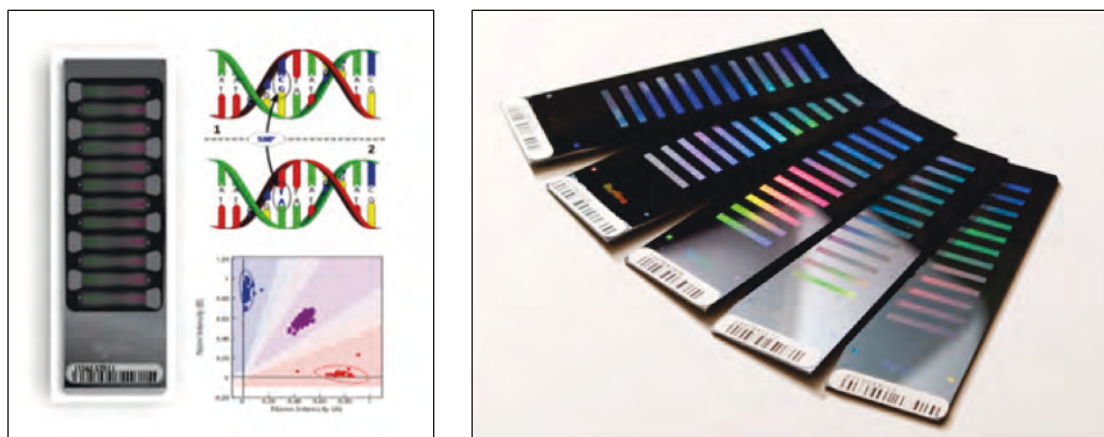


Figure VIII : Les puces de génotypage (Schibler, 2012) (<http://www.hendrix-genetics.com/fr-fr/>)

La détection de plusieurs millions de *SNP* puis leur intégration dans des outils de génotypage à grande échelle et à coût raisonnable comme les biopuces (*SNP beadchips*) permettent d'une part l'étude des polymorphismes (variations interindividuelles de séquence) répartis sur l'ensemble du génome (*genome wide SNP array*) dans les études d'association pangénomiques (*Genome Wide Association Studies, GWAS*, qui comparent de larges cohortes d'individus pour identifier des variants associés à une maladie ou à d'autres caractères), et d'autre part de mettre en œuvre des méthodes de sélection prenant en compte l'information moléculaire, telles que la sélection génomique ou *GWMAS (Genome Wide Marker Assisted Selection)* (Hayes, Goddard, 2010). Cette application a pu être mise en place grâce à la grande densité en marqueurs *SNP* existant dans les génomes et aux nouvelles technologies de génotypage permettant l'étude simultanée de plusieurs centaines de milliers de ces marqueurs (Vignal, 2011). Les puces *SNP* permettent donc désormais de génotyper précocement (dès leur naissance) la plupart des individus avec une précision remarquable, de manière relativement rapide et à moindre coût. Par exemple, le génotypage de bovins femelles sur puce 10K (c'est à dire comportant 10 000 *SNP*) dans le cadre de la sélection génomique peut varier de 45 à 60€ selon le nombre d'analyses demandées. Le génotypage de chevaux sur puce 54K (*Illumina EquineSNP50 Beadchip*) dans le cadre de projets de recherche est de l'ordre de 150€; son coût avoisine les 200€ si l'on utilise la puce 74K (*Illumina EquineSNP70 Beadchip*). Concernant les espèces ovine, caprine, porcine, et la poule, le coût du génotypage sur puce de 50 à 60K varie autour d'une centaine d'euros. Bien évidemment, le prix des puces augmente avec leur résolution (autrement dit, le nombre de marqueurs (*SNP*) qu'elles comportent).

Les puces d'expression (*microarray* d'ADN), permettant d'analyser les profils d'expression du génome, sont à distinguer des puces de génotypage (puces *SNP*) et seront présentées par la suite, avec les autres outils d'analyse transcriptomique.

➤ Des connaissances affinées grâce aux innovations dans le domaine du séquençage

Depuis 2007, on constate un essor important du séquençage massif parallèle (*NGS : Next Generation Sequencing*, séquençage de nouvelle génération) par des séquenceurs de haut débit, très utilisé actuellement (*Lamoril & al., 2008*). Ces séquenceurs de nouvelle génération permettent la lecture parallèle d'un très grand nombre de fragments d'ADN (ou d'ARN, dans le cas de l'*ARN-seq*, sur lequel nous reviendrons par la suite) immobilisés, par lecture des bases au fur et à mesure de leur incorporation, et ainsi de s'affranchir de certains biais de la méthode de Sanger (*Pinaudeau, 2015*).

Pour le séquençage de nouvelle génération, plusieurs technologies existent, correspondant à différents fournisseurs sur le marché : le pyroséquençage (utilisé par *Roche*), le séquençage par terminateur réversible (utilisé par *Illumina*), le séquençage par mesure ionique (utilisé par *Thermo Fisher Scientific*), le séquençage de molécule unique en temps réel



(technologie *Single Molecule Real Time (SMRT) ®*, utilisée par *Pacific Biosciences*)... Ci-dessous sont récapitulés les différents instruments disponibles et leurs caractéristiques d'acquisition de données, en comparaison avec un séquençage « classique » de Sanger (première ligne du tableau).

Figure IX : Séquenceur MiSeq, Illumina (<http://www.illumina.com>)

Tableau I : Offre industrielle pour les appareils de NGS comparée à du séquençage Sanger

(Guéguen & al., 2015; <http://sequencing.roche.com>, <http://www.illumina.com>, <https://www.thermofisher.com>, <http://www.pacb.com>)

Fournisseur (Technique)	Instrument / Modèle	Durée (h)	Nombre max de séquences (millions)	Longueur moy (max) des séquences (pb)	Taille cumulée max (max output) (Gb)	
Séquençage Sanger (Capillaire)	Applied Biosystems 3730	2	0,000096	650	0,00006	
Roche 454 (Pyroséquençage)	GS Junior System	10	0,100	400 (600)	0,035	
	GS Junior+ System	18	0,100	700 (800)	0,070	
	GS FLX+ System	23	1	700 (1000)	0,700	
Illumina (Séquençage par terminateur réversible)	MiniSeq System	4 – 24	25	2 x 150	7,5	
	MiSeq Series (v3)	4 – 55	25	2 x 300	15	
	NextSeq Series	12 – 30	400	2 x 150	120	
	HiSeq Series	*	5000	2 x 150	1500	
	HiSeq X Series	< 3 j	6000	2 x 150	1800	
Thermo Fisher Scientific :	Ion S5 System	Ion 520 Chip	2,5 – 4	(3 –) 5	200	(0,6 –) 1
					400	(1,2 –) 2

Ion Torrent (Séquençage par mesure ionique)		Ion 530 Chip	2,5 – 4	(15 –) 20	200	(3 –) 4
					400	(6 –) 8
		Ion 540 Chip	2,5 – 4	(60 –) 80	200	(10 –) 15
	Ion PGM™	Ion 314™	2,3	(0,400 –) 0,550	200	(0,03 –) 0,05
		Chip v2	3,7		400	(0,06 –) 0,1
	(Personal Genome Machine)	Ion 316™	3,0	(2 –) 3	200	(0,3 –) 0,6
		Chip v2	4,9		400	(0,6 –) 1
	System	Ion 318™	4,4	(4 –) 5,5	200	(0,6 –) 1
		Chip v2	7,3		400	(1,2 –) 2
		Ion Proton™ System	2 – 4	(60 –) 80	200	10
	Chip PI					
Pacific Biosciences (Séquençage de molécule unique en temps réel SMRT)	RS II System	0,5 – 4	0,060 par cellule	3000 (20000)	(0,750 –) 1,25 par cellule, max 16 cellules	
	Sequel System	0,5 – 6	0,370 par cellule	> 20000 (60000)		

* : < 1 – 3,5 j (HiSeq 3000/4000), 7 h – 6 j (HiSeq 2500)

Les industries *PacBio* (*Pacific Biosciences*) ont même développé des séquenceurs permettant de lire des séquences de 3 000 (*RS II : Read-Sequencer II*) à 20 000 pb, voire jusqu'à 60 000 pb (technologie *SMRT*) (*Guéguen & al., 2015*). Les outils proposés par *PacBio* sont idéaux pour la création de nouveaux génomes d'assemblage, à savoir les génomes d'espèces pour lesquelles il n'y a pas de référence (par exemple, les génomes bactériens). Ils sont également idéaux pour l'amélioration de génomes de référence existants chez les mammifères, en particulier parce qu'ils augmentent la précision dans les régions riches en G-C et dans les régions avec de nombreuses répétitions. De toute évidence, l'un des inconvénients reste leur coût plus élevé.



Figure X : Séquenceur PacBio RS II, Pacific Biosciences (<http://www.pacb.com>)

L'une des caractéristiques les plus importantes de la *NGS* est la « couverture », définie comme nombre de fois où chaque base est présente dans les fragments obtenus par séquençage du génome à une position donnée ; la valeur de couverture est l'un des facteurs déterminants pour l'évaluation de la fiabilité avec laquelle des nucléotides sont affectés à une position dans le génome, ainsi que la valeur de qualité de séquençage. Cela peut permettre de choisir le type de séquenceur *NGS* que l'on souhaite utiliser. Dans des expériences pharmaco-oncologiques et de médecine clinique détectant des mutations associées à des maladies ou quantifiant le nombre de copies d'une variante de structure (par exemple de duplication), cette valeur de couverture est encore plus importante.

Les données de séquences des génomes complets sont stockées dans les bases de données internationales DDBJ (Japon) / EMBL (Etats-Unis) / GenBank (Europe) et les principaux browsers de génome utilisés pour la visualisation des données de séquence et d'annotation sont *Ensembl*, *UCSC (University of California Santa Cruz)* et *NCBI (National Center for Biotechnology Information)*. (Vignal, 2011).

ii. Les données épigénomiques (l'épigénome)

On abordera de manière succincte les données épigénomiques, puisqu'elles n'interviennent pas directement dans cette thèse ; leur présentation est tout de même justifiée en ceci qu'elles sont également intégrées dans les études et analyses, au même titre que les autres données « omiques ».

L'épigénomique est une science « omique » récente et en plein essor : il s'agit de l'étude des modifications réversibles et héritables de l'ADN qui régulent l'expression génomique, telles que la méthylation des cytosines, la modification post-traductionnelle (méthylation, acétylation, phosphorylation notamment) des extrémités amino-terminales des histones, protéines intimement liées à l'ADN, le remodelage des nucléosomes (complexe nucléoprotéique formé d'un ensemble de protéines histones structurant l'ADN en un premier niveau de compaction ; il représente l'unité de base de la chromatine), les variations d'architecture tridimensionnelle de la chromatine et toute autre modification ayant pour conséquence un changement d'accessibilité de l'ADN. Les modifications des histones sont une des « voies » épigénétiques les plus étudiées (Kouzarides, 2007) ; toutes ces modifications biochimiques, quelle que soit leur nature, influent sur la transcription de l'ADN, certaines affectent la réparation de l'ADN, ou sa réplication ; la condensation de l'ADN est, quant à elle, impactée uniquement par l'acétylation et la phosphorylation des histones (Kouzarides, 2007). L'épigénomique regroupe donc un ensemble de mécanismes aboutissant à des modifications phénotypiques sans modification de la séquence d'ADN mais avec des changements de structure du génome. Des mécanismes de nature épigénétique contribuent à la régulation de différents phénomènes biologiques, comme par exemple l'« empreinte génomique » (aussi appelé « empreinte parentale », phénomène épigénétique par lequel certains gènes peuvent être exprimés de manière « parent-spécifique »), l'inactivation du chromosome X, le développement de l'embryon et la reprogrammation cellulaire, etc. ; de nombreuses modifications épigénétiques induites par l'environnement ont également été décrites, certaines étant

transmises d'une génération à l'autre sans nécessité de maintien du stimulus d'origine (Kouzarides, 2007). Certaines marques épigénétiques pourraient constituer une nouvelle classe de biomarqueurs à valeur diagnostique, pronostique ou thérapeutique (Guéguen & al., 2015).

Les méthodes d'étude des modifications des histones peuvent faire intervenir des anticorps spécifiques et/ou la spectrométrie de masse, mais aussi, désormais, le séquençage.

La méthode d'évaluation de l'état chromatinien fait ainsi souvent intervenir une phase d'immunoprécipitation de chromatine (*Chromatin ImmunoPrecipitation, ChIP*) : des anticorps ciblant l'ADN de régions génomiques portant un épitope (déterminant antigénique) spécifique sont utilisés. On obtient ainsi une collection de fragments d'ADN de courte taille (dont on sait qu'ils interagissent avec la protéine sélectionnée par l'anticorps). Cette phase immunologique peut être suivie d'une *PCR* classique (***ChIP puis PCR***), mais une approche plus récente la couple à l'utilisation d'une biopuce (*microchip*), accélérant l'identification des séquences (car plusieurs milliers de séquences sont testées simultanément) : c'est la méthode ***ChIP on chip*** (ou *ChIP-chip*) ; l'ADN enrichi est hybridé à une puce à ADN. Une alternative à l'utilisation de puces est le séquençage des fragments issus de l'immunoprécipitation : il s'agit du ***ChIP-seq***, méthode couplant une phase immunologique à un séquençage, utilisée dans de nombreuses études, dont celle de Mikkelsen & al. (Mikkelsen & al., 2007b).

Une autre étude place le séquençage de fractions définies d'ADN génomique ayant été traité au bisulfite de sodium (***Reduced Representation Bisulphite Sequencing, RRBS***) comme technologie la plus efficace et puissante pour le profilage épigénétique (évaluation de la méthylation de l'ADN) de populations cellulaires intervenant dans la biologie du développement, l'oncologie et la médecine régénérative (dont les cellules embryonnaires souches, les cellules dérivées de ces dernières, et les cellules nerveuses primaires) (Meissner & al., 2008). Elle génère, en utilisant cette méthode, des cartes de méthylation d'ADN couvrant la plupart des îlots *CpG* (séquences à concentration élevée en dinucléotides de bases nucléiques cytosine et guanine, dont la méthylation constitue autant de marques épigénétiques dynamiques) ainsi qu'un échantillon représentatif d'éléments non codants conservés, de transposons et d'autres éléments génomiques particuliers dans ces cellules. Cette étude montre également que les motifs de méthylation de l'ADN sont mieux corrélés aux motifs de méthylation des histones qu'à la séquence génomique sous-jacente.

Une hypothèse majeure révolutionnant le domaine de l'épigénétique est que les modifications des protéines histones seraient les exécutants de phénomènes épigénétiques

plutôt que les véritables porteurs de mémoire épigénétique : la transmission des processus de modifications chromatiniques est en effet réalisée via les petits ARN, véritables « déterminants » de transmission (Kouzarides, 2007). De multiples études appuient cette hypothèse, notamment en montrant que certains petits ARN présents dans le sperme murin peuvent être transmis à la descendance, et que la présence de ces ARN particuliers est associée à une conversion épigénétique stable et récurrente, phénomène appelé paramutation (Rassoulzadegan & al., 2006). Ce transfert héritable d'informations épigénétiques entre deux copies (allèles) d'un même gène, découvert initialement chez les végétaux, n'a été étudié et compris que bien plus tard chez les animaux ; c'est chez la drosophile que l'on a objectivé son lien avec l'émergence d'un locus producteur de piARN (*PIWI-interacting RNA*, petits ARN particuliers que l'on présentera par la suite) (de Vanssay & al., 2012). Epigénomique et transcriptomique apparaissent ainsi très étroitement intriquées.

iii. Les données transcriptomiques et microtranscriptomiques (le transcriptome et le microtranscriptome)

La transcriptomique est une approche permettant de quantifier et décrire le transcriptome, c'est-à-dire les différents transcrits d'une cellule (ARN codés par le génome), reflet direct du niveau d'expression des gènes. Elle se mêle étroitement à la génomique fonctionnelle, puisque cette dernière s'intéresse en partie à la régulation de l'expression génomique, et que certains transcrits interviennent dans ce fin contrôle.

Notons que le transcriptome peut aussi être considéré comme un phénotype moléculaire ; ici, nous le traitons comme « omique » à part entière.

➤ Le transcriptome

Les ARN, chefs d'orchestres majeurs dans les réseaux métaboliques et en biologie, sont très divers, aussi bien en termes de structure que de fonction (Lamoril & al., 2010a). Ce sont des molécules intermédiaires (adaptateurs) se fixant sur une séquence nucléique (molécule cible) et provoquant ainsi l'activité catalytique d'une ou plusieurs molécules partenaires (Lamoril & al., 2010a). Parmi les ARN, on trouve notamment :

- Les **ARN codants**, dont la traduction aboutit à la synthèse d'une protéine : il s'agit des **ARN messagers** (ARNm ou mRNA).

- Les **ARN non codants** : ils guident en général des complexes protéiques vers des séquences d'acides nucléiques ; de nombreuses protéines partenaires sont impliquées et constituent une unité fonctionnelle : le complexe des ribonucléoprotéines.
 - Les **ARN ribosomiaux** (ARNr, ou rRNA) : ces composants constitutifs des ribosomes (organelles cytoplasmiques permettant la traduction des ARNm en protéines) participent à la reconnaissance des ARN de transfert (présentés par la suite) au site A du ribosome.
 - Les **ARN de transfert** (ARNt ou tRNA) : ils participent à la synthèse protéique en jouant le rôle de transporteur d'acides aminés, permettant le transfert de l'acide aminé spécifique qui leur est attaché au peptide en cours de synthèse au sein du complexe ARNm / ribosome / ribonucléoprotéines / ARNr, lors d'une réaction complexe, la « traduction ». Il est important de souligner que les erreurs de traduction participent à la diversité protéique, donc phénotypique.
 - Les **ribozymes** : ce sont des ARN à activité catalytique (ARN-enzymes), comme par exemple la RNase P (ribozyme clivant spécifiquement les précurseurs d'ARNt (pré-ARNt)), les introns auto-épissants, et les petits ARN catalytiques.
 - Les **petits ARN non codants** (de 20 à 200 nucléotides) :
 - Les **petits ARN nucléaires** (snARN ou snRNA, *small nuclear RNA*) : ils guident la reconnaissance d'une séquence nucléique cible (par exemple lors de l'épissage alternatif des pré-ARNm).
 - Les **petits ARN nucléolaires** (snoARN ou snoRNA, *small nucleolar RNA*) : ils contribuent aux modifications de séquence des ARN ribosomiaux en assurant le lien entre un ou des composants catalytiques (les ribonucléoprotéines, RNP) et leur cible. En dégradant ou en réprimant leur cible par fixation à une extrémité 3'UTR (Kutter, Svoboda, 2008), ils interviennent ainsi dans la traduction en protéines, ils contribuent à la stabilité du génome, et certains joueraient un rôle dans la régulation de l'épissage alternatif des ARNm.
 - Les **petits ARN régulateurs/répressifs** : ces petits ARN de 20 à 30 pb présentent la particularité d'exercer un contrôle négatif sur

d'autres ARN en s'associant à des complexes protéiques comprenant des protéines de la famille Argonaute (*RNA silencing*).

- ✓ Les **microARN** (miARN ou miRNA) : ce sont de courts ARN de 20 à 24 pb, dont le rôle de régulateur de l'expression génétique physiologique et au cours du développement est fondamental. Associés aux protéines AGO, ils provoquent la dégradation de l'ARNm cible ou inhibent la traduction. Le mode de fonctionnement précis et les particularités de ces ARN seront détaillés plus tard.
- ✓ Les **petits ARN interférents** (siARN ou siRNA, *small interfering RNA*) : ARN de 21-22 pb exogènes ou endogènes (endo-siARN exprimés dans les oocytes et les cellules des lignées germinale et somatique), leur rôle est similaire à celui des miARN mais ils dégradent systématiquement l'ARN cible, et leur association aux protéines AGO guide l'interférence soit sur l'ARN, soit sur l'ADN. Ils régulent ainsi l'expression génomique, notamment par répression des transposons ; ils interviennent également dans le développement de la lignée germinale.
- ✓ Les **ARN interagissant avec les protéines PIWI** (piARN ou piRNA, *PIWI-interacting RNA*) : ces ARN de 24 à 31 pb s'associent aux protéines de la famille PIWI (*p-element induced wimpy testis*), sous classe de la famille des protéines Argonautes, et régulent ainsi l'expression génomique, avec pour cible principale les transposons. Ils diffèrent des miARN et siARN par une biogénèse indépendante de l'enzyme DICER entre autres. Chez les Mammifères, les piARN ne sont présents que chez les mâles, dans les cellules germinales ou somatiques en contact avec des cellules germinales : ils ne sont pas essentiels dans la lignée germinale femelle mais sont nécessaires pour une spermatogénèse correcte (ce n'est pas toujours le cas, la situation diffère par exemple chez le poisson-zèbre (*Kutter, Svoboda, 2008*)). Les fonctions des piARN sont diverses : protection de la lignée

germinale et somatique, maintien de l'intégrité du génome, régulation de l'expression de l'ARNm, contrôle de l'expression génomique dans la lignée germinale, rôle dans l'empreinte génomique...

- On trouve d'autres petits ARN non codants tels que les « riboswitches » par exemple, intervenant dans le contrôle de la traduction (*Mattick, Makunin, 2006*).
- Les **longs ARN non codants** (longs ARNnc ou lncRNA, *long non-coding RNA*) : de taille supérieure à 200 nucléotides, allant jusqu'à 100 kb, leur origine et leurs modes d'actions sont multiples, ce qui rend leur identification et leur étude complexes (*Mercer & al., 2009; Rinn, Chang, 2012; Geisler, Collier, 2013*). Parmi eux, on trouve les ARN ribosomiaux 18S et 28S (ARNr), les longs (ou larges) ARN non codants intergéniques (lincARN), les régions transcrites ultraconservées (T-UCR), les pseudogènes (séquence génomique de structure proche d'un gène connu mais non fonctionnelle), les ARN à séquence répétée GAA (GRC-ARN), les longs ARN introniques non codants, les ARN antisens (aARN, encore appelés transcrits antisens, dont on commence tout juste à entrevoir les fonctions biologiques et leurs implications sur la régulation, l'architecture, et l'évolution génomiques (*Pelechano, Steinmetz, 2013*)), les longs ARN associés aux promoteurs (PALR), les transcrits en amont du promoteur (PROMPT), les ARN stables issus d'introns, les longs transcrits non codants induits par le stress (LSINCT)... Ce sont notamment des régulateurs épigénétiques, en tant qu'intermédiaires favorisant l'association de complexes protéiques agissant sur le remodelage chromatinien donc sur l'expression génomique ; un paradigme émergent serait de les considérer comme chefs d'orchestre des actions coordonnées dans l'espace et dans le temps des complexes agissant sur le remodelage de la chromatine (*Geisler, Collier, 2013; Mercer, Mattick, 2013; Lee, 2012*). On peut par exemple citer leur implication dans l'empreinte parentale (seul un des chromosomes est exprimé par inactivation de l'autre par un long ARNnc) et l'inactivation du chromosome X chez les Mammifères via *XIST* (*X-inactive specific transcript*) (*Chow, Heard, 2009*). Leurs nombreuses fonctions incluent la régulation de l'épissage alternatif, la répression de la transcription via les ARNnc Alu (*Brosnan, Voinnet, 2009*), la régulation du cycle cellulaire et de

l'apoptose, et le contrôle du développement via les loci HOX (*homeobox*) (Brosnan, Voinnet, 2009; Sauvageau & al., 2013). Au-delà de leurs nombreux rôles sur la transcription (Ponting & al., 2009; Wilusz & al., 2009; Nagano, Fraser, 2011), les longs ARNnc interviennent dans la régulation post-transcriptionnelle et post-traductionnelle de l'activité protéique (par leur capacité d'inhibition d'activités enzymatiques diverses et multiples), l'organisation des complexes multiprotéiques (par leur potentiel à servir de base à la formation (autrement dit, de point de nucléation) de complexes protéiques d'ordre supérieur), le signalement intercellulaire (par leur présence dans les exosomes en tant qu'*exosomal shuttle RNAs*, comme certains miARN et ARMm), la recombinaison de l'ADN (notamment la recombinaison V(D)J, réaction de recombinaison site-spécifique à l'origine de la grande diversité de récepteurs des cellules immunitaires T et d'immunoglobulines, essentielle à la reconnaissance d'anticorps divers) (Geisler, Collier, 2013)... De nombreuses études restent à effectuer pour appréhender la complexité de ces longs ARNnc (Mattick, 2009).

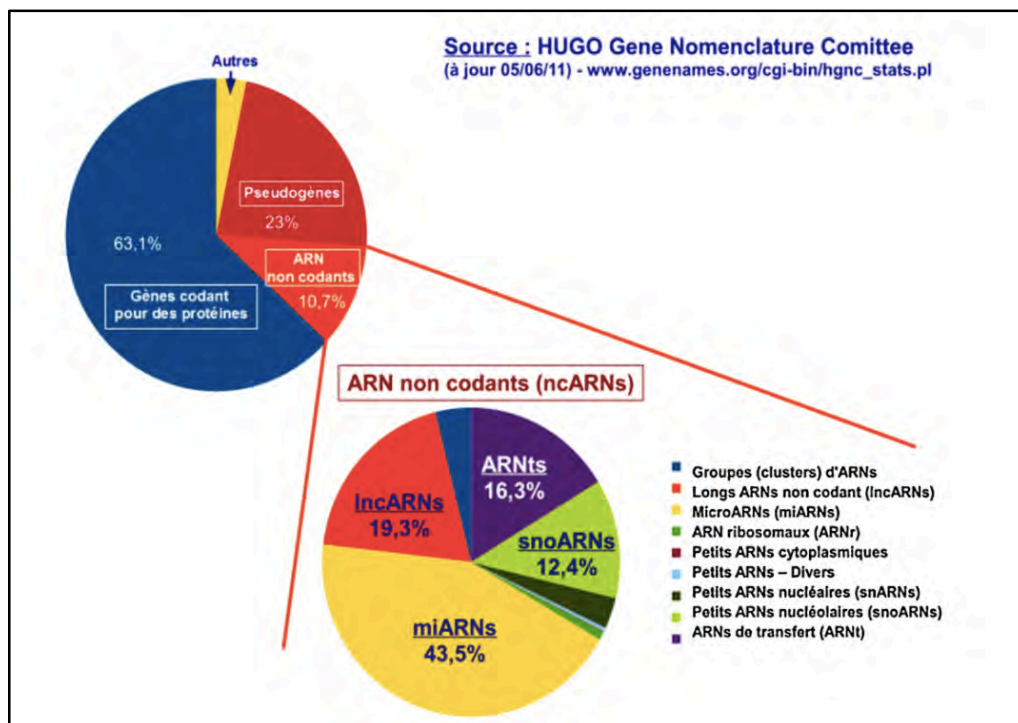


Figure XI : Représentation quantitative des gènes d'ARN non codants parmi les gènes identifiés dans le génome humain (Lamoril & al., 2011)

Encart : les micro ARN

L'importance toute particulière des miARN dans le cadre de la partie expérimentale de cette thèse justifie leur présentation de manière détaillée dans cet encart.

Ce sont une des trois classes de petits ARN simple brin non codants, constitués de 20 à 24 (ou 21 à 23, selon les sources) nucléotides, assurant la régulation post-transcriptionnelle de l'expression génétique en se liant aux protéines de la famille argonaute. L'étude de ces petits ARN particuliers est à l'origine d'une science à part entière : la microtranscriptomique ou *microRNomics*, sous-discipline de la génomique fonctionnelle et de la transcriptomique, se concentrant sur l'identification des miARN et l'analyse de l'expression, la biogenèse, la structure, les cibles, ainsi que les fonctions biologiques des miARN à une échelle génomique, ainsi que leur rôle dans la régulation de l'expression génomique. L'expression des miARN est tissu-spécifique. Cette spécificité associée à la possibilité de les mesurer dans le sang circulant, ainsi que leur stabilité (liée à l'association à des complexes particuliers ou protéiques assurant leur protection contre les RNases) en font de bons biomarqueurs (la notion de biomarqueur sera développée par la suite). Depuis Juin 2014, 28 645 miARN sont répertoriés dans la base de données miRBase (version miRBase 21) (<http://www.mirbase.org/>). Les miARN sont impliqués dans de nombreux mécanismes essentiels à la vie cellulaire tels que le métabolisme et le cycle cellulaire, l'apoptose, le développement et la différenciation cellulaire, la réponse au stress et la signalisation hormonale. Toute variation qualitative ou quantitative de leur expression peut être associée au développement de nombreuses affections, ce qui explique en partie leur utilisation en tant que biomarqueurs (de cancers, de prédisposition à la résistance aux thérapeutiques, etc.). Les miARN sont des répresseurs post-transcriptionnels, c'est-à-dire qu'ils guident la dégradation ou la répression de la traduction en protéine des ARN messagers cibles en s'y appariant de manière spécifique par quelques bases : en s'associant à des protéines de la famille Argonaute (AGO), ils guident des complexes appelés *RNA-induced silencing complexes* (RISC) aux ARNm auxquels ils s'apparient, avec une spécificité absolue ou relative. Ils peuvent donc réguler les fonctions cellulaires mais aussi contrôler la structure de la chromatine, la méthylation des acides nucléiques et l'expression de gènes codants, notamment celle de gènes codant des facteurs de transcription. La régulation biologique par les miARN prend donc différents degrés : il s'agit soit d'une inactivation totale de la cible (par répression totale de l'expression de l'ARNm), soit d'une modulation d'expression (l'ARNm reste fonctionnel mais moins qu'en l'absence du/des miARN le ciblant), soit d'un effet neutre (pas d'effet) ; cela se traduit en des modulations phénotypiques, ou bien en un changement

phénotypique total, ou encore par un effet quantitatif sur le phénomène (pour les miARN ayant une action sur plusieurs cibles). Les miARN sont responsables d'un « tuning » : ils peuvent moduler l'activité de leur cible, et ainsi la concentration en miARN influe sur l'expression de l'ARNm cible et donc sur la quantité de protéines, ce qui modifie les effets biologiques et est à l'origine de la modulation phénotypique. Les miARN diffèrent des facteurs de transcription principalement par leur vitesse d'action (plus grande), la réversibilité de la régulation qu'ils opèrent, ainsi que leur compartimentation cellulaire (ils affectent donc l'expression locale d'un gène, par exemple dans les synapses) ; ils sont ainsi à l'origine d'une réponse ajustée, rapide et spécifique, permettant le maintien de l'homéostasie cellulaire et l'obtention d'une réponse rapide à un stimulus environnemental. (Savagner & al., 2015; Lamoril & al., 2010b). L'efficacité de ciblage et donc de fonctionnement des miARN (et siARN) est principalement liée à l'accessibilité du miARN à son/ses site(s) cible(s) (donc à la possibilité d'interaction ARN/cible) (Kutter, Svoboda, 2008).

➤ Les outils de la transcriptomique

La quantification des transcrits utilise différents outils et techniques, dont les plus importants sont :

- **La RT-PCR** (reverse transcriptase-polymerase chain reaction) ou PCR quantitative en temps réel, qui permet de dénombrer les transcrits (ARN) issus d'un gène donné. Elle est surtout utilisée pour valider les différentiels d'expression observés entre deux conditions données avec les puces à ADN.
- **Les puces à acides nucléiques (ADN, ARN, ADNc, ARNc)** (*microarray*, biopuces, puces d'expression) : lames (de verre, dans la majorité des cas) sur lesquelles sont déposées de manière ordonnée des fragments d'ADN ou ARN appelés « sondes ». En raison de son utilisation dans la partie expérimentale de la thèse, cet outil issu des innovations technologiques dans le domaine de la biologie moléculaire se doit d'être présenté en détail.

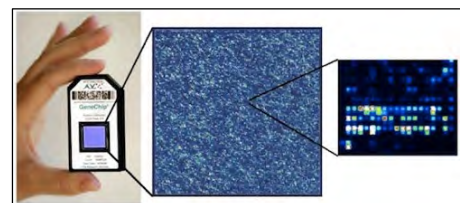


Figure XII : Puce Affymetrix (GeneChip®), zoom sur l'aspect du microarray (<https://www.affymetrix.com>)

Principe d'analyse sur *microarray* :

Les *microarrays* sont une application du principe d'hybridation inverse (représenté sur la *Figure XIII*) : les sondes sont utilisées pour repérer de manière spécifique des séquences d'intérêt dans un milieu complexe contenant des milliers de séquences nucléotidiques différentes (Bogard & al., 2008). Chaque fragment d'acide nucléique peut représenter un gène, un ARN non codant (tels que les miARN ou les ARN longs non codants), ou une entité encore inconnue. Les hybridations sont hautement spécifiques dans les conditions expérimentales utilisées. Les puces permettent ainsi de détecter des acides nucléiques extraits d'un échantillon biologique, et de les quantifier dans certaines conditions. Le marquage de ces extraits est généralement effectué à l'aide de fluorophores. Grâce aux *microarrays*, on peut donc comparer l'expression de milliers de transcrits entre plusieurs conditions en une seule manipulation opératoire, ce qui en font un outil d'une grande puissance. Par exemple, la puce *GeneChip® human genome U133 Plus 2.0 array*, commercialisée par la société Affymetrix, permet l'analyse simultanée de plus de 47 000 transcrits. Les puces d'expression permettent de découvrir des réseaux de gènes exprimés sous certaines conditions et donc de détecter des *QTL* d'expression (*eQTL* : régions du génome polymorphes, dont le polymorphisme explique une part de la variation du niveau d'expression de gènes de la même région – cis *eQTL* – ou d'autres régions – trans *eQTLs*).

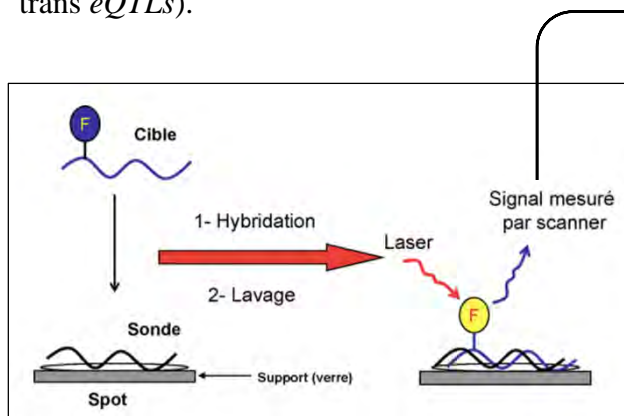


Figure XIII : Principe d'analyse sur microarray

Technique d'hybridation inverse : sonde fixée sur le support, cible capturée marquée par un fluorophore dont l'excitation par un laser entraîne l'émission d'un signal quantifié, si la cible est présente dans le milieu analysé.
(Bogard & al., 2008)

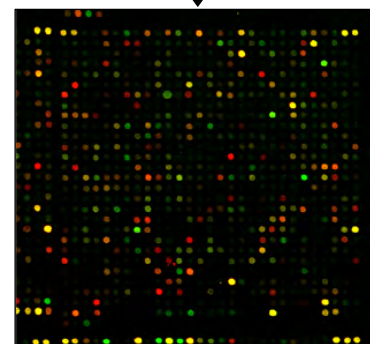


Figure XIV : Aspect d'un microarray

Sondes déposées de manière ordonnée (chaque spot correspond à une sonde) ; analyse simultanée de centaines voire de milliers de cibles différentes.

En général, les études portent sur des acides nucléiques obtenus à partir d'ARNm ; l'array mesure alors l'expression des gènes dont sont issus les messagers (le signal mesuré sera proportionnel au nombre de messagers transcrits).

Fabrication d'un *microarray* :

Il existe trois technologies principales de création de ce type d'outil transcriptomique : le dépôt de sondes sous forme de microgouttes, la synthèse de sondes in situ (systèmes de synthèse d'oligonucléotides de type « jet d'encre », ou de techniques photolithographiques avec « masque de protection » des acides nucléiques sondes (utilisée par Affymetrix pour sa puce *GeneChip*®, cf *Figure XV* ci-dessous) ou sans masque de protection (système *Maskless Array Synthesizer* développé par Nimblegen™ Systems)), et l'adressage électronique de sondes sur des microélectrodes de platine (développé par la société Nanogen™ avec son système *NanoChip*® *Electronic Microarray*).

L'article de M. Bogard et ses collègues (*Bogard & al., 2008*) présente le détail de fabrication de ces différentes puces ; on peut également accéder à de nombreuses informations via les sites internet des différentes sociétés commercialisant ce type d'outils (Affymetrix et sa puce *GeneChip*®

http://www.affymetrix.com/estore/browse/level_one_category_template_one.jsp?category=35796&categoryIdClicked=35796&parent=35796,

Agilent et sa puce *SurePrint G3*®

<http://www.genomics.agilent.com/campaign.jsp?id=5000008&CID=G012085...>).

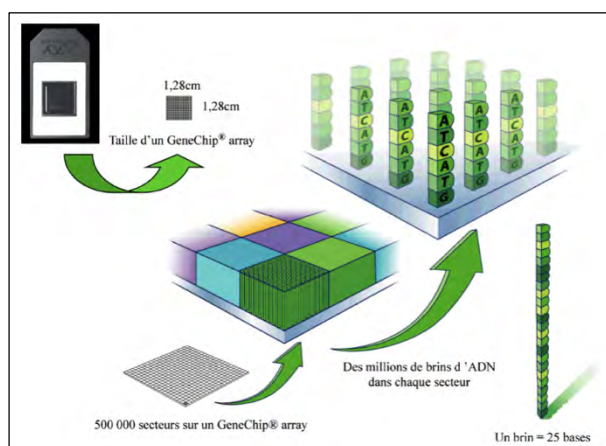


Figure XV : Caractéristiques d'une puce GeneChip®, société Affymetrix

Puce de 1,28 cm x 1,28 cm, à 500 000 secteurs (dans chacun : des milliers de sondes identiques de 25 bases de long). (Bogard & al., 2008)

Concernant les puces d'expression, comme celles utilisées pour notre étude expérimentale (*Agilent custom equine miRNA microarray 8x60K* et *4x44K*), les sondes que l'on utilise sont préparées à partir d'ADNc (ADN complémentaire des ARNm) obtenus par transcription inverse.

Utilisation des *microarrays* d'expression :

Cette application particulière des *arrays* présente un caractère innovant par la réalisation d'un « instantané » de ce qui se passe à l'échelle cellulaire : on observe dans sa globalité l'expression sous forme d'ARNm ou d'ADNc d'un ensemble de gènes, voire de tous les gènes d'une culture cellulaire ou d'un tissu, et on compare cette expression dans différentes circonstances. Ainsi, produire un profil d'expression de gènes revient à étudier un ensemble de gènes dans deux conditions différentes : par exemple, on analyse l'expression d'un tissu après une phase d'exercice musculaire de l'individu (conditions de stress cellulaire et tissulaire) par rapport au même tissu avant l'effort, au repos (condition qui sert de référence).

Les trois étapes principales d'une analyse sur microarray sont représentées sur la *Figure XVI* : (a) Préparation de l'échantillon avec vérification de sa qualité par électrophorèse, mesure d'absorbance ou étude de l'ARN total sur *biochip* (Agilent™) combinant électrophorèse et mesure quantitative de fluorescence ; (b) Hybridation de l'extrait sur la puce ; (c) Phase de lecture, c'est-à-dire excitation au laser et enregistrement des signaux correspondant aux molécules hybridées.

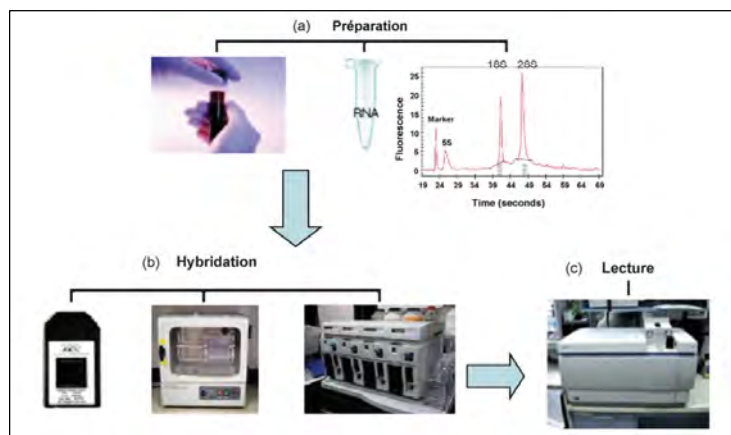
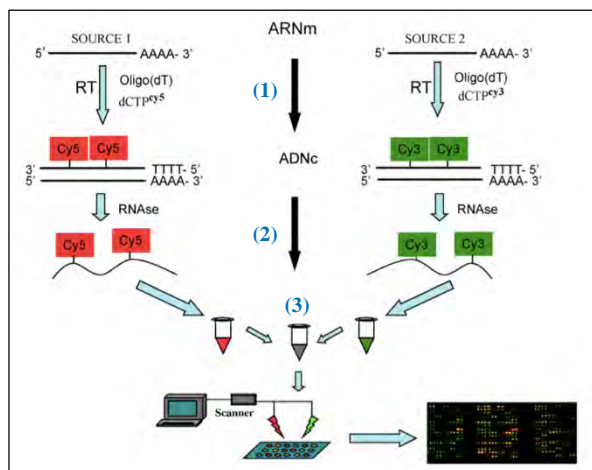


Figure XVI : Analyse sur puce Affymetrix (Bogard & al., 2008)

La méthode de marquage est fonction de la technologie de l'*array* utilisé. Soit on réalise deux *arrays* indépendants, chacun par hybridation des transcrits dans une des deux conditions : il s'agit alors d'un marquage unique (méthode utilisée par

Affymetrix par exemple) ; soit on hybride sur le même *array* les deux extraits (cibles, transcrits à tester) marqués de manière différente dans chacune des deux conditions : c'est une hybridation compétitive. Dans ce dernier cas, présenté sur la *Figure XVII*, les marqueurs fluorescents Cy3 et Cy5 sont classiquement utilisés. L'approche par hybridation compétitive facilite les analyses comparatives et la méthode limite la variabilité entre les expérimentations, car l'utilisation d'une seule puce au lieu de deux limite tous les biais potentiels issus de la comparaison de deux *arrays*.



- (1) Formation des ADNc en présence d'une transcriptase inverse et de précurseurs (dNTP) dont l'un (le dCTP) est marqué avec une molécule fluorescente. L'ADNc de la source 1 est marqué au Cy5, celui de la source 2 au Cy3.
- (2) Elimination du brin d'ARN, ADNc retrouvés sous forme simple brin.
- (3) Mélange des deux sources, placées à hybrider sur la puce d'expression.

Figure XVII : Analyse comparative de l'expression de gènes par hybridation compétitive de deux échantillons de sources différentes (Bogard & al., 2008)

En ce qui concerne l'exploitation des images, ces dernières sont traitées par un logiciel affectant, à partir de deux images monochromes, des couleurs aux signaux mesurés : vert pour Cy3 et rouge pour Cy5. La fluorescence est d'abord mesurée aux deux longueurs d'onde (mesure des signaux) puis le rapport d'intensité de fluorescence (Cy5/Cy3) des deux échantillons hybridés de façon compétitive sur le *microarray* est calculé (analyse des signaux). Ainsi, si le spot est vert, l'expression du gène est plus forte dans la source marquée par Cy3. S'il est rouge, l'expression du gène est plus forte dans la source marquée par Cy5. S'il est jaune ou gris, les deux sources ont le même niveau d'expression.

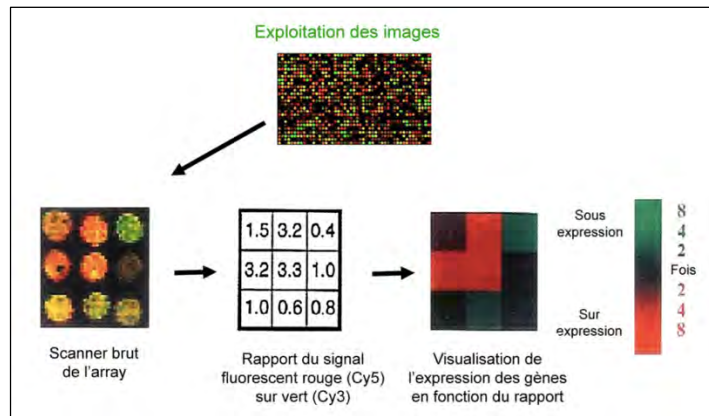


Figure XVIII : Analyse des images dans le cadre d'une hybridation compétitive (Bogard & al., 2008)

L'analyse des données fournies par un *microarray* d'expression nécessite une normalisation préalable de l'intensité des signaux obtenus, l'extraction d'ARN et le marquage fluorescent étant effectués de manière indépendante. On élimine les mauvaises images (décalées, noires, de mauvaise qualité), on soustrait le bruit de fond (pour faire ressortir les différences entre les deux échantillons comparés ; cela peut être réalisé de plusieurs manières différentes) et enfin on procède à la normalisation proprement dite (traitement du signal) : elle a pour objectif de corriger les biais systématiques des données (liés à la préparation des sondes, l'impression des arrays / alignement des sondes, la quantité et la qualité de l'ARN initial, les problèmes de lecture du scanner, le logiciel d'analyse, l'expérimentateur, etc.) et de supprimer dans la mesure du possible toute influence non biologique sur ces données, afin de tirer de vraies informations biologiques des données et de pouvoir comparer ces dernières d'un *array* à l'autre, voire d'une plateforme d'*array* à une autre. La normalisation passe par une transformation du signal : l'intensité de chaque spot (généralisé par laser) est en général exprimée en échelle logarithmique (\log_2 le plus souvent), ce qui permet de traiter de manière « symétrique » les gènes surexprimés ou sous-exprimés, de rendre les variations d'intensité plus indépendantes de leur valeur absolue, et de faciliter l'emploi d'outils statistiques qui pourront alors recentrer la distribution. On utilise alors un MA plot, qui exprime graphiquement la fonction $y = f(x)$ où x est l'intensité moyenne de fluorescence de chaque spot et y le log du rapport de fluorescence Cy5/Cy3. Afin de rectifier l'image en forme de « banane » reflet des courbures dues à une non-linéarité des intensités de fluorescence, on utilise fréquemment une correction des courbures appelée transformation de Loess ou Lowess (*locally weighted scatter plot smoothing*) qui

permet la production d'un nuage de points centrés sur une droite. Ainsi normalisées, les données peuvent être exploitées par analyse statistique. (Bogard & al., 2008)

- **Le séquençage d'ARN** : comme dans le champ de la génomique structurale, le séquençage de masse (NGS) est entré fortement dans le domaine de la transcriptomique. Si les puces sont actuellement encore largement utilisées dans de nombreux contextes, l'ARN-seq (ou RNA-seq, séquençage de l'ARN, analyse transcriptomique par NGS) a commencé à remplacer cette technologie.



Figure XIX : Séquençeur Ion Proton™, Thermo - Life Technologies (<https://www.thermofisher.com>)

Les avantages que présente l'ARN-seq sont qu'il permet : i) de détecter de nouveaux gènes et isoformes ; ii) de détecter l'expression spécifique d'allèles différents ; iii) de détecter des sites d'épissage (« splice junctions ») ; iv) de détecter des longs ARN non codants qui peuvent avoir un rôle important dans la régulation épigénétique (tels que les lincARN, T-UCR, etc.), comme exposé plus haut (Ozsolak, Milos, 2011). En outre, l'ARN-seq ne nécessite pas la sélection d'une sonde (séquence nucléotidique) à hybrider sur une puce ; elle prend en compte la totalité de séquence du transcriptome d'une cellule ou d'un tissu. Cependant, l'étude d'une séquence d'ARN implique le développement de nouveaux algorithmes pour l'analyse et la normalisation des données ainsi que la quantification des transcrits de gènes, et d'autres algorithmes également afin d'ajuster la distribution de l'expression ou de corriger des erreurs de détection de gènes différentiellement exprimés. Contrairement aux ARN codants, les petits ARN non codants sont rarement détectés avec le protocole classique d'ARN-seq. Bien que l'on dispose des biopuces pour étudier l'expression de ces petits ARN dans les différents tissus, l'utilisation des technologies de séquençage massif par de petits fragments d'ARN offre la possibilité de découvrir de nouveaux petits ARN qui ne sont pas encore décrits sur miRBase (<http://www.mirbase.org>). Les miARN sont désormais des outils puissants de diagnostic en tant que biomarqueurs parce que : i) ils sont conservés entre espèces, ce qui suggère l'importance de leur rôle biologique ; ii) le nombre relatif de miARN connus est encore faible, bien que leur gamme d'expression soit très variable ; iii) ils sont sécrétés dans la circulation, ce qui permet de les étudier dans des fluides

biologiques tels que le sang et donc de s'affranchir d'une biopsie ou d'autres méthodes invasives ; iv) ils ne se dégradent pas aisément, et cette grande stabilité est une valeur importante dans l'analyse d'échantillons cliniques ou d'échantillons d'intérêt ancien et historique ; v) ils sont impliqués dans la régulation d'environ un tiers du transcriptome ; vi) ils présentent la particularité d'avoir une expression tissu-dépendante ; vii) on dispose maintenant de nombreuses données bibliographiques sur leur rôle dans la physiopathologie des maladies.

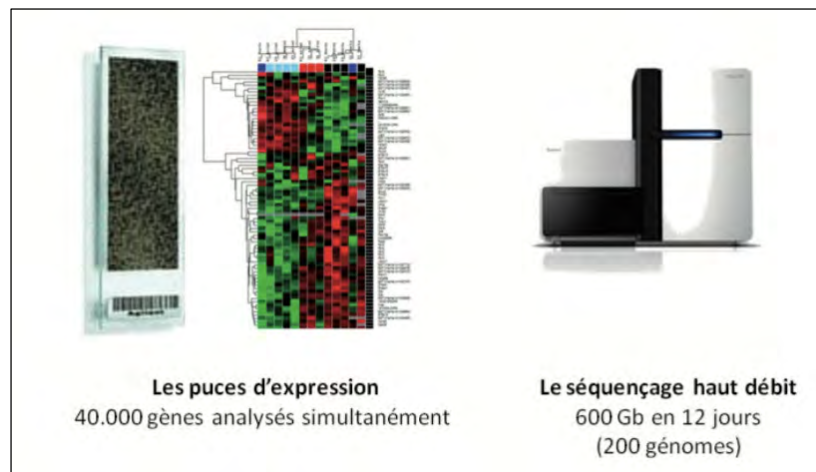


Figure XX: Les principaux outils de la transcriptomique / génomique fonctionnelle (Schibler, 2012)

- Et enfin des **techniques génomiques pour l'étude de l'expression antisens** : cette dernière nécessite des approches « brin spécifique » pour mesurer ces transcrits particuliers et minimiser les artefacts environnementaux. L'immunoprécipitation de la chromatine étant une méthode non « brin-spécifique », on utilise désormais pour quantifier et séquencer les transcrits antisens la **GRO-seq** (*Global Run-on sequencing*), qui mesure la présence et l'orientation de polymérase active réalisant l'élongation d'un brin, ainsi que la **NET-seq** (*Native Elongating Transcript sequencing*), qui mesure les transcrits en formation (transcrits « naissants ») à partir d'ARN polymérase engagées dans la transcription ; ces deux techniques sont complétées d'un séquençage des ARN produits. (Pelechano, Steinmetz, 2013)

iv. Les données protéomiques (le protéome)

C'est le généticien australien Marc R. Wilkins qui donne naissance en 1994 à la notion de protéome (Wilkins, 2009). Suite à l'achèvement du *Human Genome Project*, l'intérêt s'est

reporté sur l'aspect protéique de l'organisme humain ; ainsi s'est développée la protéomique. Ce domaine scientifique est défini comme l'étude à grande échelle des protéines, en particulier de leurs expression, structures, fonctions et interactions, dans un système biologique donné (Zhang & al., 2010) : cette approche, d'émergence récente, permet d'identifier, de décrire, de caractériser et de quantifier les protéines présentes (exprimées), dans un compartiment cellulaire, une cellule, un tissu, ou même un organisme entier. On pourrait dire que la protéine est à la protéomique ce que le gène est à la génomique. Alors que la séquence de référence équine était encore incomplète, une carte de référence en deux dimensions des protéines sériques équines avait déjà été réalisée (Miller & al., 2004). Avec la finalisation du séquençage des génomes entiers de nombreux procaryotes et eucaryotes, la protéomique a connu un regain d'intérêt ; le protéome reflète plus précisément l'état dynamique d'un système biologique que le génome, ce qui fait de la protéomique un domaine complémentaire à la génomique fonctionnelle (Cho, 2007). L'importance du protéome est incontestable : ce sont les protéines cellulaires qui sont responsables de la structure de la cellule et de la production d'énergie, ce sont elles qui assurent la communication, le mouvement et la reproduction cellulaires ; en d'autres termes, les protéines fournissent le cadre structural et fonctionnel de la vie cellulaire (Cho, 2007).

A l'instar des technologies utilisées en génomique, celles utilisées en protéomique ont connu de nombreuses avancées depuis leur première utilisation, mais les méthodes initiales restent des techniques de choix dans certains cas.

La séparation des protéines se fait soit par électrophorèse bidimensionnelle (la plus souvent utilisée), soit par chromatographie en phase liquide, puis leur identification est réalisée par spectrométrie de masse (digestion des protéines et identification des peptides obtenus, selon leur masse moléculaire, par comparaison avec des bases de données).

L'électrophorèse bidimensionnelle sur gel (*2-Dimensional (Polyacrylamide Gel) Electrophoresis, 2-D(PAGE)*) est la méthode traditionnelle, puissante, de séparation de protéines sur la base de leur point isoélectrique et de leur poids moléculaire (Klose, Kobalz, 1995) ; elle a connu des améliorations au fil des années depuis sa première utilisation, lui conférant désormais une haute résolution (Zabel, Klose, 2009).

L'approche par chromatographie en phase liquide (*liquid chromatography, LC*) lui est parfois préférée dans les études protéomiques à grande échelle (*shotgun proteomics*) en raison de sa sensibilité et de son rendement supérieurs ; l'objectif étant de fractionner les mélanges protéiques afin de maximiser les capacités des spectromètres de masse à identifier et quantifier

les peptides composant les protéines (Zhang & al., 2010). La *LC-MS (Liquid Chromatography coupled with Mass Spectrometry)* est ainsi considérée par de nombreux laboratoires comme la méthode de choix en protéomique, pour sa rapidité et sa précision (Xu & al., 2005). Le développement récent de la chromatographie multidimensionnelle en phase liquide (*Multidimensional Liquid Chromatography, MDLC*) permet désormais une séparation encore plus puissante des peptides (Zhang & al., 2010).

Pour exploiter les profils protéiques exprimés sous différentes conditions physiologiques et pathologiques, la spectrométrie de masse est classiquement utilisée comme plateforme puissante en protéomique (Cho, 2007). La phase d'ionisation des protéines peut être réalisée de différentes manières, les plus fréquemment utilisées étant l'ionisation par électroébuliseur (*Electrospray Ionisation, ESI*), l'ionisation par désorption au laser d'une matrice (*Matrix-Assisted Laser Desorption/Ionization, MALDI*), et l'ionisation par désorption au laser activée par une surface (*Surface-Enhanced Laser Desorption/Ionization, SELDI*), car ce sont celles qui ne causent que peu ou pas de fragmentation des molécules durant le processus d'ionisation/désorption. Le spectromètre de masse utilisé comportant souvent un détecteur triant les ions selon leur « temps de vol » (*Time-Of-Flight, TOF*), on parle de spectrométrie de masse *ESI-TOF*, *MALDI-TOF* ou *SELDI-TOF*. En effet, lorsque les ions sont accélérés dans un champ électrique et/ou magnétique, la vitesse avec laquelle ils atteignent le détecteur est conditionnée par leur masse, on peut ainsi séparer les peptides (ensemble de ces ions) selon leur vitesse, qui est fonction du rapport m/z (masse/charge ou nombre de protons). Alternative aux spectromètres à détecteurs de type *TOF*, la technologie *LTQ-Orbitrap* de spectroscopie de masse (*Linear ion Trap Quadrupole – Orbitrap Mass Spectroscopy, LTQ-Orbitrap MS*) peut être utilisée, prenant en compte non pas le temps de vol des ions, mais leur courbure de trajectoire (le point d'impact des ions dépend lui aussi du rapport m/z , ce qui permet leur tri).

Les innovations dans les biotechnologies ont permis la mise au point d'outils à présent disponibles pour les analyses protéomiques, pouvant compléter l'utilisation des spectromètres de masse ou bien être utilisés seuls. On trouve notamment des puces à protéines, robustes et stables, permettant de détecter les interactions moléculaires avec d'autres protéines, peptides, acides nucléiques, sucres, etc. (Figeys, Pinto, 2001), mais aussi des outils de protéomique quantitative faisant intervenir des isotopes stables, permettant de surmonter le problème de la variabilité d'ionisation de différents peptides (Cho, 2007). Parmi ces outils, la technique de marqueurs d'affinité contenant un isotope d'identification (*Isotope-Coded Affinity Tag labeling, ICAT*) est la plus développée ; elle permet de comparer l'abondance relative des

protéines entre deux échantillons (par exemple entre une condition normale et une condition pathologique).

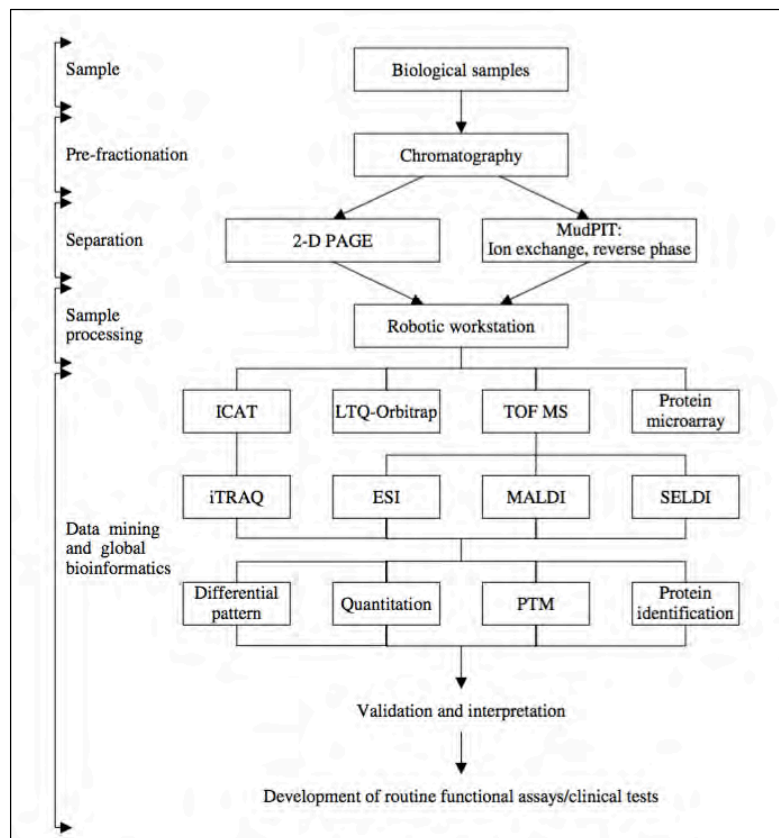


Figure XXI : Les technologies protéomiques contemporaines et leur organisation (Cho, 2007)

2-D PAGE : 2-Dimensional Polyacrylamide Gel Electrophoresis ; ESI : Electrospray Ionization ; ICAT : Isotope-Coded Affinity Tag ; iTRAQ : isobaric Tags for Relative and Absolute Quantification ; LTQ : Linear ion Trap Quadrupole ; MALDI : Matrix-Assisted Laser Desorption/Ionization ; MudPIT : Multidimensional Protein Identification Technology, PTM : Post-Transcriptional Modification ; SELDI : Surface-Enhanced Laser Desorption/Ionization ; TOF MS : Time-Of-Flight Mass Spectrometry.

Comme pour les autres données « omiques », les outils bioinformatiques sont essentiels pour convertir des données protéomiques brutes en connaissances et donc applications utiles. Nous n’aborderons pas ici en détails les applications de la protéomique, puisqu’elles seront évoquées dans le « zoom » sur les « omiques » équin, mais il est important de souligner que la protéomique est à l’origine de nouveaux horizons dans de nombreux domaines de recherche biomédicale et de médecine.

La protéomique clinique peut être définie comme sous-discipline de la protéomique appliquée au domaine de la médecine, et s’avère prometteuse dans l’identification de nouvelles cibles thérapeutiques et de marqueurs protéiques d’affections utiles au diagnostic *in vitro* (In

Vitro Diagnostics, IVD, tests cliniques applicables à des échantillons tels que les tissus et fluides corporels : sérum, plasma, urine...) (Vitzthum, 2005).

En effet, les technologies protéomiques à haut débit combinées aux approches bioinformatiques sont utilisées de manière considérable dans l'identification de signatures moléculaires d'états physiologiques ou pathologiques basée sur les voies du métabolisme protéique et les cascades de signalisation protéique. Citons le domaine de l'oncologie, où les études protéomiques visent à l'identification de biomarqueurs dans l'optique d'une détection précoce d'évolution cancéreuse (Conrad & al., 2008). Combinée à l'utilisation de la transcriptomique (par des analyses sur *microarrays* par exemple), la protéomique aide à gagner en précision et peut être à l'origine de l'identification de biomarqueurs dans des contextes où l'on n'en avait pas déjà repéré en utilisant seule une des deux sciences « omiques » (Devaux & al., 2010). Au delà de son utilisation dans le diagnostic au sens strict, la protéomique permet aussi la conception de nouveaux traitements pharmaceutiques sur la base de l'identification de cibles thérapeutiques protéiques. On comprend aussi son intervention dans la médecine personnalisée, puisqu'elle pourrait permettre l'élaboration d'un traitement optimal de chaque sous-population de patients si ces derniers sont discriminés sur la base de leur machinerie cellulaire (Gulmann & al., 2006) et de leur expression protéique. Elle offre la possibilité de suivre l'évolution des affections et d'émettre un pronostic en suivant un ou des marqueurs protéiques suite à un traitement ou dans le cadre d'un non-traitement. Enfin, la précocité de détection des biomarqueurs protéiques justifie le fait que l'on considère la protéomique comme moyen de prévention contre certaines affections. Prévention, diagnostic, pronostic, traitement et suivi sont donc des champs couverts par ce domaine scientifique en plein essor.

Il faut préciser que la protéomique « différentielle » est utilisée dans le cadre de la protéomique clinique : elle détecte les protéines associées à une maladie grâce à des niveaux d'expression diminués ou augmentés entre un état témoin et un état pathologique ; des corrélations sont alors établies entre l'ensemble des protéines produites par un système biologique et l'initiation ou la progression d'une affection (Cho, 2007).

Les possibilités d'intervention de la protéomique en pharmaco-toxicologie paraissent évidentes : en effet, de nombreuses drogues et/ou substances dopantes étant de nature protéique, des analyses protéomiques semblent pertinentes afin de les détecter dans un organisme. Elle trouve à ce titre des applications à la fois dans les domaines médical, sportif et même juridique.

Enfin, les données protéomiques viennent enrichir les connaissances issues des données génomiques : par exemple, des séquences génomiques équine ont été validées et annotées par modélisation fonctionnelle du protéome d'un fluide issu de lavage broncho-alvéolaire équin

(Bright & al., 2011), et on a pu montrer des différences substantielles entre le protéome sérique de chevaux et d'ânes, liées à des modifications post-traductionnelles spécifiques d'espèce, par une étude utilisant des puces de protéines (ProteinChips®) (Henze & al., 2011).

v. Les données métabolomiques (le métabolome)

Le terme « métabolome » a été proposé dans une étude anglo-saxonne en 1998 (Oliver & al., 1998), par analogie avec les termes « génome », « transcriptome » et « protéome ». La métabolomique est un domaine scientifique étudiant l'ensemble des métabolites (molécules de faible poids moléculaire synthétisées par un organisme : sucres, acides aminés, acides gras, etc.) présents et quantifiables dans un échantillon biologique (cellules, tissus, plasma, urine). Elle repose actuellement sur l'utilisation de deux grandes techniques que sont la spectrométrie de masse et la résonance magnétique nucléaire (RMN) et permet de visualiser l'ensemble de l'activité d'un système biologique (compartiment cellulaire, cellule, tissu, organe...) (Faucon-Lahalle, Montagu, 2011).

Une définition plus étroite de la métabolomique est la caractérisation des phénotypes métaboliques sous des conditions parfaitement définies (stade évolutif de l'organisme (âge pour un mammifère), modifications génétiques et/ou conditions environnementales) et du lien entre ces phénotypes et les génotypes correspondants. Cette définition a quelque peu évolué puisque les communautés scientifiques ont eu tendance à associer la notion de métabolomique à l'identification et à la quantification de groupes de métabolites (biomarqueurs) présents dans un système biologique, sans relation avec la notion de génotypage (Combourieu, 2007). Toujours est-il que le profilage métabolique permet de prendre en compte des informations d'origine à la fois intrinsèque (génomique et expression protéique) et extrinsèque (régime alimentaire, microbiote intestinal...), fournissant une vue du système dans sa globalité (Escalona & al., 2015).

Que cela soit pour l'étude ciblée de voies métaboliques, l'analyse des flux métaboliques associés ou l'analyse globale (statistique) du métabolome, les outils privilégiés restent la spectrométrie de masse et la RMN, parfois utilisés de manière complémentaire afin d'obtenir une couverture maximale du métabolome (Combourieu, 2007; Escalona & al., 2015). La spectrométrie de masse ayant déjà été présentée comme outil protéomique, il convient de décrire brièvement la RMN. Cette méthode est utilisée depuis les années 1980 dans l'étude du métabolisme énergétique et permet d'explorer un répertoire de quelques centaines de

métabolites de petits poids moléculaires à un coût relativement abordable, et donc de comparer le métabolome de plusieurs centaines d'individus ou tout simplement de très nombreux échantillons différents (San Cristobal & al., 2014). Le suivi de molécules endogènes par RMN ^1H , ^{31}P ou ^{13}C constitue d'ailleurs un des seuls moyens d'accéder *in situ* ou *in vivo* aux voies et aux flux métaboliques (Combourieu, 2007). La RMN s'applique surtout à des biofluides (urine, plasma sanguin...) et fournit en sortie des spectres tels que celui présenté ci-après (Figure XXII), ce qui donne une image qualitative et quantitative de ces fluides biologiques sans aucun *a priori*. Un spectre peut être considéré comme la signature métabolique d'un système biologique observé à un temps donné, dans un état physiologique ou pathologique donné. En spectroscopie RMN ^1H à une dimension (1D), les métabolites sont identifiés par la position des pics, due à leurs protons non-échangeables. La position (déplacement chimique par rapport à la fréquence de résonance d'une substance de référence) et le motif (structure en multiplet) d'un pic donné reflètent tous deux l'environnement chimique du ou des protons constituant le métabolite. Ces informations, associées à de vastes bases de données, sont utilisées pour l'identification des métabolites. L'intensité d'un pic est proportionnelle à la concentration de métabolite dans l'échantillon, mais un pic peut correspondre à un ou plusieurs métabolites, et un métabolite précis peut être représenté par différents pics sur le spectre, ce qui complexifie l'analyse du spectre. Pour pallier le manque de résolution lors d'analyse de mélanges complexes par RMN du proton sans accroître la valeur du champ magnétique (onéreux), il est possible de coupler une détection RMN avec une séparation chromatographique (LC, présentée auparavant).

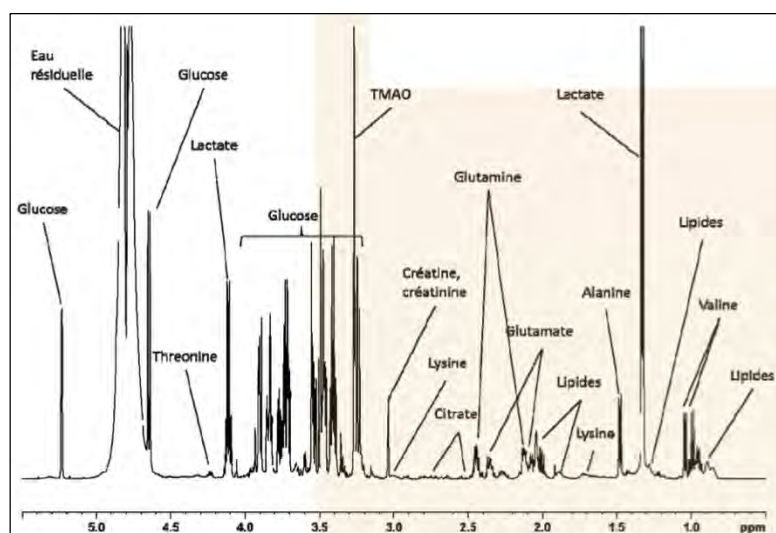


Figure XXII : Exemple de spectre métabolomique RMN d'un échantillon plasmatique (San Cristobal & al., 2014)

Le déplacement chimique est représenté en abscisse.

Chaque valeur du spectre en ordonnée est susceptible de varier d'un échantillon à un autre.

La spectroscopie RMN ^1H présente l'avantage, par rapport à la spectrométrie de masse, d'être une méthode de forte reproductibilité, et dont la sensibilité ne dépend pas des propriétés physiques et chimiques des molécules (*Pan, Raftery, 2007*). En revanche, elle nécessite une concentration en métabolites plus élevée afin que ces derniers soient détectés.

Tout comme les autres sciences « omiques », la métabolomique trouve des applications et ouvre des horizons potentiels dans des domaines multiples et variés : recherche, médecine, pharmacologie, toxicologie, nutrition, environnement, sport, zootechnie, domaine législatif / judiciaire... Dans tous les cas, son utilisation pour comprendre les mécanismes biologiques sous-jacents d'un phénotype particulier (maladie, conformation, performance sportive...) ou pour prédire ce dernier, consiste en une recherche de biomarqueurs (ce terme sera défini de façon précise par la suite) pronostiques, précoces ou descriptifs ; l'utilisation de la métabolomique plutôt que d'autres sciences -omiques en tant que prédicteur de phénotypes est particulièrement justifiée en partie en raison de la proximité forte entre métabolome et phénomène dans la cascade biologique (génomome et ADN, transcriptome et ARN, protéome et protéines, métabolome et métabolites, phénomène et phénotypes). Ainsi, certaines études sont menées afin de quantifier le pouvoir prédictif du métabolome sur des phénotypes d'intérêt : par exemple, M. San Cristobal et ses collègues se sont récemment concentrés sur la quantification du pouvoir prédictif du métabolome sanguin sur les phénotypes de production particuliers que sont la composition corporelle porcine (taux de muscles des pièces ou poids de certaines pièces de la carcasse comme le jambon) ainsi que le rendement de carcasse et la consommation moyenne journalière de ces animaux, et ont montré qu'un spectre métabolomique peut être envisagé comme source de biomarqueurs intéressants dans certaines conditions (*San Cristobal & al., 2014*). Le principe d'une étude métabolomique est schématisé sur la figure suivante, ici dans le cadre de bovins traités par l'administration d'anabolisants.

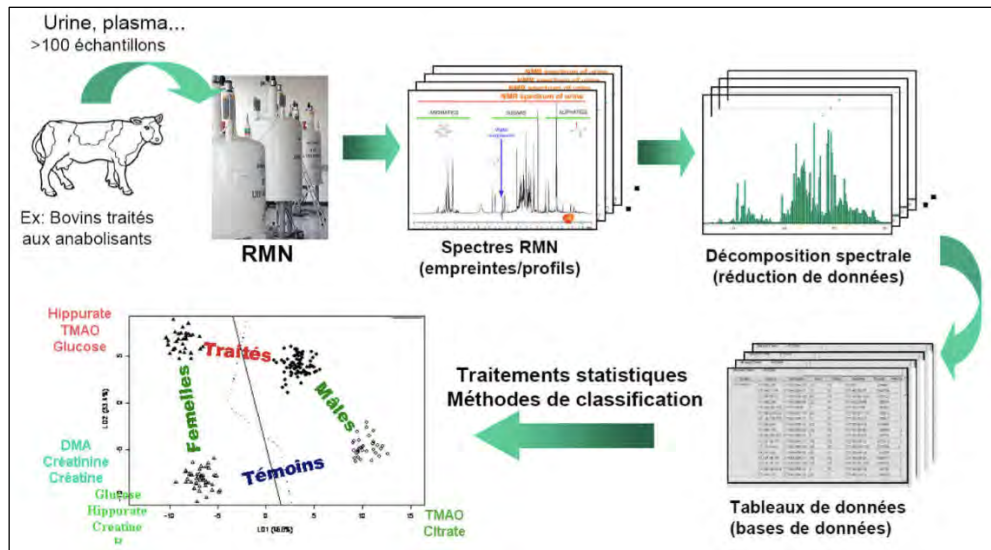


Figure XXIII : Etude métabolomique utilisant la RMN dans le cadre d'un traitement de bovins aux anabolisants (JC Portais, Cours de Métabolomique, Université Paul Sabatier (Toulouse))

(Figure issue d'un diaporama en ligne : http://genoweb1.irisa.fr/OGP/ftp/Gen2Bio2012/Atelier27_Giraudeau_Gen2Bio2012.pdf)

La métabolomique se révèle tout particulièrement efficace dans la compréhension des interactions métaboliques complexes entre un hôte mammifère et sa flore commensale intestinale, ce qui est intéressant pour la recherche équine au vu du large impact du microbiote intestinal sur la biodisponibilité d'aliments, de médicaments et de l'énergie. Comme les autres -omiques, elle montre un potentiel de plus en plus important dans les paramètres cliniques à la fois comme outil de détection/dépistage et comme moyen de compréhension des mécanismes de voies métaboliques associées à une maladie (Escalona & al., 2015).

Puisque nous reviendrons sur certaines applications de la métabolomique en lien avec le sport, et que ce domaine scientifique sera abordé dans la suite de la thèse, nous n'entrerons pas plus dans les détails ici.

- c. Un zoom sur les données de génomique, transcriptomique, et autres « omiques » équines, et leur application en médecine vétérinaire (génomique clinique et autres sciences omiques équines appliquées)

Les connaissances dans le domaine de la génomique et transcriptomique et des autres sciences « omiques » équines présentent quelques particularités qu'il est intéressant de souligner. Afin d'avoir un aperçu de la situation chez le cheval, il semble important de rappeler

les principales étapes clés dans les investigations menées depuis quelques décennies, de présenter les outils de biologie moléculaire spécifiquement utilisés pour les études chez cet animal, et de s'attarder sur quelques applications usuelles de ces outils à des fins cliniques.

i. Les débuts de la génomique équine : séquence de référence, assemblage (Finno, Bannasch, 2014; Brosnahan & al., 2010)

L'aventure « génomique » commence chez le cheval il y a une vingtaine d'années (Octobre 1995) avec la mise au point des **premières cartes génétiques équines** par le *First International Equine Gene Mapping Workshop* (Premier Atelier International de Cartographie Génétique Equine), au sein du *Horse Genome Project (H.G.P.)* regroupant généticiens et cliniciens vétérinaires de 22 laboratoires répartis dans 12 pays. Les cartes de première génération du génome équin, qui recensaient 733 marqueurs polymorphes distribués uniformément, ont rapidement été suivies par celles de deuxième génération, qui regroupaient 4104 marqueurs. Ces cartes ont fourni les outils initiaux nécessaires à l'assemblage de la séquence génomique entière de notre animal préféré.

La première **séquence génomique** du cheval domestique est publiée en **Novembre 2009** suite au travail collaboratif de la communauté mondiale de chercheurs équins (Wade & al., 2009). Elle est construite à partir de l'ADN d'une jument pur-sang nommée Twilight, présentant un niveau élevé d'homozygotie de son Complexe Majeur d'Histocompatibilité (C.M.H., région de haute diversité génétique impliquée dans le système immunitaire, en général difficile à assembler) comparé aux autres chevaux



Figure XXIV : Twilight, 1^{ère} jument séquencée
(<https://www.uky.edu/Ag/Horsemap/hgphorses.html>)

candidats. La méthode utilisée est le WGS (présentée précédemment). Cette version préliminaire (« draft sequence ») de haute qualité a été enrichie par l'ajout de séquences supplémentaires (extrémités de *BAC*, *bacterial artificial chromosome*) à partir d'un hongre pur-sang du nom de Bravo, apparenté à Twilight. L'assemblage résultant, **EqCab2.0**, présente une couverture de 6,8X du génome (autrement dit, chaque portion d'ADN est séquencée 6,8 fois en moyenne) ce qui est certes faible mais parmi les meilleures valeurs de couverture que l'on pouvait obtenir il y a quelques années (plus importante que celle des autres génomes de mammifères hormis l'Homme et la souris). Cette base de données a été rendue publique et accessible à la communauté de la recherche équine. La séquence génomique équine a été estimée à cette occasion à une longueur de l'ordre de **2,7 Gb**, avec **32 paires de chromosomes**

(31 paires d'autosomes et les chromosomes sexuels, soit une formule génétique de $2n = 64$ pour une cellule diploïde de cheval domestique), **20 322 gènes** ont été annotés, et un taux de polymorphisme global de 1/1000 pb a été déterminé.

La base de données de marqueurs génétiques a été étoffée par l'apport de séquences génomiques partielles de sept chevaux supplémentaires, chacun de race différente (Akhal-Teke, Andalou, Arabe, Islandais, Pur-Sang, Quarter Horse, Trotteur Américain) (Wade & al., 2009). Une carte *SNP* de plus d'un million de marqueurs a été générée à partir des 400 000 *SNP* supplémentaires découverts suite à une centaine de milliers de lectures *WGS* des séquences partielles de ces chevaux, ajoutés aux 700 000 *SNP* découverts dans le génome de Twilight. En complément aux marqueurs microsatellites, cette carte est alors disponible comme outil génomique pour l'étude de caractères et d'affections génétiques.

En 2011, le génome entier d'une jument Quarter Horse Américain a été séquencé en utilisant le séquençage massif parallèle « paired-end » (séquençage de deux extrémités de fragments courts, de taille inférieure à 1 kb) ; la couverture du génome de cette jument était de 24,7X (Doan & al., 2012). 19,1 Mb de séquence génomique équine inédite ont été générés à l'occasion de ce séquençage de novo. Ce dernier permettait en outre de passer d'un répertoire de 1 163 580 *SNP* équins, sans polymorphismes d'insertion/délétion ni *CNV* annotés, à un total de 3,1 millions de *SNP*, avec 193 000 insertions/délétions et 282 *CNV* détectés et consécutivement répertoriés. Des analyses de voies biologiques ont mis en évidence chez le Quarter Horse des *SNP* à l'origine d'une utilisation de voies biologiques impliquées dans la proprioception, les processus cellulaires et la transduction de signaux. Etant donné que la jument n'avait pas été sélectionnée pour le séquençage sur des critères d'homozygotie et qu'elle était d'une race différente de celle ayant fourni la séquence de référence, son génome a fourni une ressource précieuse pour les études de variabilité génétique.

En plus du séquençage de races contemporaines, la séquence génomique de chevaux préhistoriques a été étudiée récemment, révélant que la lignée *Equus* a donné naissance à tous les chevaux modernes, zèbres et ânes, et que cette lignée est née il y a 404,5 millions d'années (Orlando & al., 2013). Le cheval de przewalski possède 66 chromosomes (donc deux de plus que les chevaux domestiques modernes) et son ADN mitochondrial forme un groupe génétique distinct ; il proviendrait d'un groupe génétique régional différent de celui dont sont issus les chevaux domestiques modernes (Jansen & al., 2002). Le séquençage supplémentaire par *NGS* de races de chevaux domestiques et d'un przewalski n'a pas montré de preuve de brassage génétique récent entre ces deux types de chevaux, ce qui conforte l'hypothèse selon laquelle les chevaux de przewalski représenteraient les derniers survivants de la population équine sauvage

et l'étude ayant révélé cela montre que le cheval de przewalski a divergé de la lignée du cheval moderne il y a environ 50 000 ans (Orlando & al., 2013) et non pas 160 000 ans comme le suggérait une étude antérieure (Goto & al., 2011).

Les efforts sont actuellement concentrés sur l'amélioration de la séquence équine de référence via la création de **EquCab3.0** en enrichissant la séquence génomique de Twilight.

ii. Les outils disponibles (Finno, Bannasch, 2014)

Le séquençage du génome équin et les avancées consécutives dans le domaine de la génomique équine ont mené à des progrès conséquents dans le développement d'outils visant à cartographier des caractères et des maladies et à évaluer l'expression génétique.

Ainsi, les puces *SNP* équines permettent l'étude de caractères génétiques « simples » et « complexes », tandis que les *microarrays* d'ADN se révèlent essentiels dans l'étude de l'expression génétique de divers tissus et sous certaines conditions pathologiques. La *NGS* étant devenue plus abordable depuis peu, le séquençage d'ADN *tout-génome* et l'*ARN-seq* sont des méthodes désormais couramment appliquées à la recherche génomique équine, domaine qui continue à s'étendre rapidement au fur et à mesure de la diminution des coûts de génotypage et de séquençage.



Figure XXV :
Illumina
EquineSNP50
Beadchip
(<http://www.illumina.com>)

Deux puces *SNP* équines sont actuellement disponibles. La puce de première génération, **EquineSNP50 Beadchip**, mise au point en 2008 à San Diego en Californie et commercialisée par *Illumina*, rassemble 54 602 marqueurs *SNP* (puce *SNP* 50K) et couvre le génome équin entier à l'exception du chromosome Y, avec une distance de 43,1 kb entre *SNP* sur les 31 autosomes. Parmi ces 54 602 *SNP*, 53 524 révèlent un polymorphisme, (c'est-à-dire qu'on trouve au moins un individu hétérozygote en ce locus au sein du groupe échantillon).

Dans l'article présentant la séquence de Twilight, des estimations de puissance reposant sur la longueur du déséquilibre de liaison (*linkage disequilibrium (LD)*, niveau d'association entre marqueurs) chez le cheval, le nombre de combinaisons alléliques au sein d'un haplotype (séquences adjacentes d'ADN sur un chromosome) et le taux de recombinaison, ont suggéré que plus de 100 000 *SNP* seraient nécessaires pour cartographier des caractères au sein de chaque race et entre les différentes races (Wade & al., 2009). La puce *SNP* de première génération a été validée sur un panel d'échantillons représentatifs de 14 races de chevaux domestiques modernes et 18 espèces apparentées (McCue & al., 2012). En raison du grand

déséquilibre de liaison dans des races telles que le Quarter Horse et le cheval mongol, il a été recommandé d'augmenter le nombre de marqueurs afin d'obtenir une cartographie fine dans les races préhistoriques et celles caractérisées par une taille efficace de population (N_e , aussi appelée population efficace ou effectif efficace de population) importante (McCue & al., 2012). La puce *SNP* 50K de première génération ne possède donc que la moitié de la densité de marqueurs estimée nécessaire pour mener à bien des études d'association sur des races avec un LD moyen ou élevé. Elle a cependant été utilisée pour étudier la structure de la population chez 744 individus de 33 races équines. La variabilité relevée parmi les races a été utilisée afin d'identifier des gènes et des variants génétiques ciblés par la sélection génétique (c'est-à-dire des « signatures » de sélection). Cette étude a permis d'identifier : i) des variants du gène de la myostatine (*MSTN*) sur le chromosome 18 (ECA18) du cheval Paint américain (*American Paint Horse*) et du Quarter Horse américain, associés de manière significative à des proportions modifiées d'un type de fibres musculaires impliquées dans les capacités de sprint (fibres de type 2B augmenté, type 1 diminué), ii) des variants (mutation du gène *DMRT3* sur le chromosome ECA23, présentée plus haut dans cette thèse) dans les races qui présentent des allures supplémentaires, et iii) des polymorphismes dans des régions génomiques participant à la détermination de la taille (sur le chromosome ECA11) chez des chevaux de trait et des chevaux miniatures (Petersen & al., 2013). En Janvier 2011, la puce *EquineSNP50* a été remplacée par une puce *SNP* de seconde génération, la *EquineSNP70 Beadchip*, présentant 74 500 marqueurs, avec une moyenne de 1,5 *SNP* toutes les 50 kb. Cette plateforme réunit les 53 524 marqueurs polymorphes initiaux de la puce *SNP* 50K et des *SNP* supplémentaires choisis pour combler les lacunes et améliorer la couverture sur l'ensemble du génome. Ces derniers ont été fournis à partir du séquençage des chevaux des 7 races « de découverte » (ayant contribué à étoffer la base de données de marqueurs génétiques, présentés dans le paragraphe précédent : Akhal-Teke, Andalou, etc.), de celui de Twilight, et de données issues de *RNA-seq* (sur lesquelles nous reviendrons plus tard) (Coleman & al., 2010). Par rapport à la précédente, la puce *SNP* 70K contient des *SNP* augmentant la couverture du C.M.H. sur le chromosome 20, ainsi que des *SNP* supplémentaires sur le chromosome X et 2 *SNP* sur le chromosome Y. Des études d'association utilisant les puces à *SNP* équines ont permis d'identifier une région chromosomique portant un gène candidat important dans le syndrome du poulain lavande (*Lavender Foal Syndrome, LFS*) et un séquençage consécutif a mis en évidence la mutation génétique responsable de cette affection (Brooks & al., 2010). En plus de son apport pour l'étude du *LFS*, la puce *SNP* équine 50K a été utilisée pour identifier des mutations génétiques impliquées dans le syndrome de l'immunodéficience létale du poulain (Fox-Clipsham & al., 2011)

et la mutation DMRT3 (présentée plus haut) permettant la possession d'allures supplémentaires (Andersson & al., 2012). D'autres études d'association utilisant les puces *SNP* équines ont permis de détecter : i) des *QTL* impliqués dans l'ostéochondrite disséquante chez les pur-sangs (Corbin & al., 2012) à explorer de manière plus poussée, ii) des loci associés à un risque de neuropathie laryngée récurrente (affection du nerf laryngé récurrent, causant un symptôme respiratoire appelé « cornage » chez le cheval) (Dupuis & al., 2011), iii) des loci et *QTL* associés à la taille corporelle (Makvandi-Nejad & al., 2012; Tetens & al., 2013; Metzger & al., 2013; Frischknecht & al., 2015), et iv) des régions candidates impliquées dans le tympanisme des poches gutturales (Metzger & al., 2012), l'uvéite équine (Kulbrock & al., 2013), et l'hypersensibilité aux piqures d'insectes (responsable de la derm(at)ite estivale récidivante équine, DERE) (Schurink & al., 2012). La densité de marqueurs de 100 000 *SNP* n'ayant toujours pas été atteinte, une puce *SNP* de troisième génération est en cours de développement, avec un objectif de 700 000 marqueurs ; sa disponibilité était prévue pour 2014 mais elle n'est pas encore commercialisée actuellement.

L'étude de l'expression génomique tissu-spécifique chez le cheval sous des conditions particulières et en considérant certains processus pathologiques est un domaine de recherche en plein essor. Les premiers outils développés pour étudier cette expression, par l'évaluation du microtranscriptome, incluaient le séquençage d'*EST* (Nagaraj & al., 2006), l'analyse en série de l'expression des gènes (*serial analysis of gene expression, SAGE*) (Velculescu & al., 1995), ainsi que les puces d'expression (*microarrays*) (Schenk & al., 1995). Jusqu'à une période récente, les puces d'expression étaient utilisées comme méthode expérimentale primaire dans l'analyse de l'expression génétique chez le cheval au niveau transcriptomique. Pour rappel, les technologies par *microarray* font intervenir une extraction d'ARN (cible, *target*) et son hybridation consécutive sur la puce d'expression à des séquences d'ADN connues et spécifiques (sondes, *probes*). Les motifs d'hybridation sont alors comparés pour permettre l'identification d'ARNm surreprésentés ou sous-représentés dans deux échantillons cibles ou plus (Murphy, 2002). Initialement, les puces humaine et murine étaient utilisées pour établir les profils d'expression génétique équins. Suite au séquençage du génome équin, quelques groupes ont initié des études afin d'améliorer les puces équines, en utilisant les modèles prédictifs d'*Ensembl* et du *NCBI*. Les *microarrays* équins ont été utilisés pour évaluer l'expression génétique dans le cadre de fourbures (Noschka & al., 2009) et de cicatrisation du cartilage articulaire (Mienaltowski & al., 2009). Une étude récente utilisant la technologie des

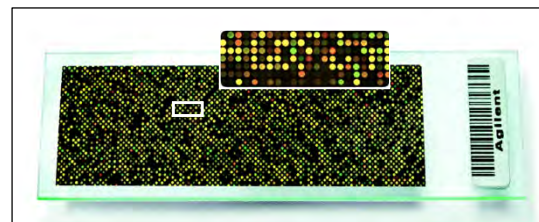


Figure XXVI: Puce d'expression commercialisée par Agilent (<http://www.agilent.com>)

microarrays sur des tissus placentaires équins a identifié une surexpression de l'ARNm codant la cytokine interleukine (IL)-22 dans la ceinture chorionique : c'est la première fois que l'IL-22 était détectée dans d'autres cellules que celles du système immunitaire (Brosnahan & al., 2012). Comme requis pour toute étude d'expression utilisant les puces à ADN, ces résultats ont été confirmés à l'aide de la *RT-PCR* quantitative. Agilent™ fournit actuellement une puce d'expression équine comportant 43 803 sondes, qui peut être personnalisée de manière à répondre à divers besoins spécifiques de recherche : *Agilent eArray custom microarray* (<http://www.genomics.agilent.com/en/Custom-Design-Tools/eArray/?cid=AG-PT-122&tabId=AG-PR-1047&pageId=1456&crumbAction=push&navAction=pop>).

Depuis peu, l'*ARN-seq* est utilisée pour affiner les modèles structuraux géniques et évaluer les motifs d'expression géniques (Coleman & al., 2010a, 2010b). Cette méthode génère des données à la fois quantitatives et qualitatives, tout en donnant un aperçu sur les transcrits alternatifs d'un même gène. Elle est actuellement appliquée dans l'étude de nombreuses affections équines. En outre, elle a été utilisée pour étudier le rôle de l'empreinte génomique chez le cheval (Wang & al., 2012; Wang & al., 2013). La *NGS* est aussi utilisée au niveau de l'ADN (séquençage massif parallèle d'ADN) chez le cheval ; ces technologies plus efficaces permettent d'obtenir des séquences représentatives du génome entier d'un cheval à un coût

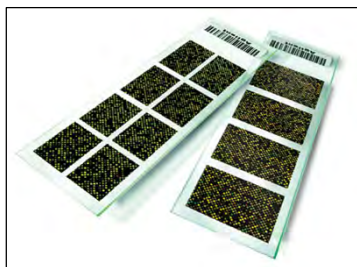


Figure XXVII: Gene expression custom microarray, Agilent (<http://www.agilent.com>)

abordable et d'offrir un aperçu sans précédent de tous les variants (*SNP*, insertions, délétions, réarrangements) du génome. De nombreuses études qui avaient initialement employé des techniques d'association pangénomiques (présentées ci-dessus) recourent désormais à la *NGS* afin d'étudier de manière plus poussée une ou des régions d'association. Grâce au fait que la *NGS* soit de moins en moins onéreuse, des séquences de génome équin entier sont générées dans des laboratoires répartis dans

différentes régions du monde. L'analyse et le stockage de ces grandes quantités de données de séquence peut devenir problématique. De nombreux programmes logiciels sont actuellement disponibles afin de réaliser des contrôles de qualité, l'alignement sur la séquence de référence, l'assemblage de séquences non alignées, et la détection de variants dans les données issues de séquençage massif parallèle (Koboldt & al., 2013). Concernant la *RNA-seq*, des algorithmes variés permettent de quantifier et comparer l'expression génétique entre différentes conditions (Seyednasrollah & al., 2015). Des connaissances poussées en bioinformatique sont désormais essentielles afin d'analyser les séquences obtenues par *NGS*.

iii. Les méthodes pertinentes pour utiliser ces outils biomoléculaires dans l'étude de caractères génétiques (Finno, Bannasch, 2014)

Avant de décrire les principales découvertes liées à l'application de ces outils novateurs dans l'espèce équine, il est important de brièvement rappeler les stratégies d'analyse appropriées pour l'étude de caractères génétiques.

Au vu de la disponibilité croissante et de la baisse des coûts des outils génomiques conçus pour les chevaux, de larges opportunités se présentent pour les mettre au service de l'étude d'un grand nombre d'affections équines. Si l'on souhaite étudier les maladies génétiques équines, il faut choisir judicieusement la population échantillon ainsi que des outils adéquats.

Pour étudier les facteurs de risque de développer une affection, il est essentiel de réaliser un phénotypage fin. Le phénotype étudié devrait être choisi *a priori* et son intensité devrait être gradée, comme cela devient utile pour la sélection de candidats au séquençage. Lorsque cela est possible, une population témoin devrait être choisie afin de maximiser les chances que les chevaux témoins ne manifestent jamais le phénotype étudié, en considérant l'âge moyen d'apparition de la maladie, les facteurs de risque environnementaux et le degré de parenté entre les populations témoin et malade.

La taille de l'échantillon disponible influe sur l'utilisation des outils génomiques. Des articles comme celui de Finno et Bannasch (Finno, Bannasch, 2014) rapportent qu'« avec des populations de taille moyenne présentant des phénotypes bien perceptibles et un mode de transmission autosomique récessif du caractère phénotypique [c'est-à-dire que le gène impliqué est porté par un autosome, et que la présence de deux allèles identiques est indispensable pour que le caractère phénotypique soit exprimé par l'individu], une association entre le phénotype et un locus chromosomique (ou des loci) peut être identifiée à l'aide des techniques disponibles actuellement via une étude d'association tout-génome utilisant les puces SNP équines sur le marché. Cependant, de faibles tailles d'échantillon ou bien un phénotype influencé par de multiples facteurs de risque, par l'imprécision du diagnostic, ou par le faible déséquilibre de liaison dans certaines races (telles que le Quarter Horse), peuvent aboutir à un échec de détection d'associations significatives en raison de la faible puissance statistique. ». Mais comment définir concrètement une « population de taille moyenne » ? Effectivement, un échantillon de population de taille importante ne permet pas de s'affranchir de bons estimateurs phénotypiques ; autrement dit, non seulement l'échantillon doit être assez grand, mais le caractère phénotypique étudié doit être clairement et précisément défini et sa variabilité doit être assez grande au sein de la population afin de pouvoir l'associer à un

déterminisme génétique. Les facteurs qui influent sur le caractère doivent être limités, ou à défaut, leurs effets doivent pouvoir être quantifiés (et corrigés). Dans le cas d'une maladie, le diagnostic doit être établi sur la base de signes cliniques non ambivalents, et/ou de critères analytiques (biochimiques, métaboliques, etc.) révélateurs, et/ou d'examen complémentaires (radiographiques, échographiques, etc.) non équivoques. Si ces conditions sont réunies, un échantillon de 500 chevaux devrait être suffisant pour de telles études. Cela dit, n'oublions pas qu'une des difficultés avec laquelle les chercheurs doivent composer dans le domaine équin est la réticence de nombreux propriétaires à consentir à la réalisation de prélèvements ou d'interventions sur leurs chevaux lorsque ces derniers sont en bonne santé : ce problème limite bien souvent les possibilités de disposer d'un échantillon de grande taille dans le cadre d'études de recherche (en plus qu'il soit véritablement dommage de ne pas disposer des valeurs de référence de son propre cheval afin d'avoir une base de comparaison si celui-là connaît une baisse de forme ultérieurement, dans le cas où il s'agit d'un simple prélèvement sanguin). Une solution peut être de mentionner l'intérêt et les applications concrètes que l'étude de recherche pourrait entraîner : par exemple, pour impliquer les propriétaires dans une étude sur le microbiote intestinal des chevaux, bien insister sur le fait qu'une connaissance plus fine de ce dernier permettrait d'évaluer le risque d'un cheval de développer un jour des affections digestives, et que des avancées sur cette question pourraient permettre de prévenir l'apparition de ce genre de maladies, ou du moins d'être conscient du danger pour être prêt à réagir en conséquence.

Le séquençage massivement parallèle d'ADN et l'*ARN-seq* peuvent être utilisés simultanément pour explorer l'expression génomique et la corréler à des variants de séquence génomique. Lors de la conception d'une expérience par *ARN-seq*, il est nécessaire de prendre en compte les aspects fondamentaux d'une méthodologie expérimentale solide : la réplication (nombre de réplicats biologiques/techniques), la randomisation et la constitution de blocs. Il est également impératif que l'ARN extrait pour tous les cas provienne d'un site standardisé du tissu d'intérêt. L'utilisation simultanée d'*ARN-seq* et de séquençage ADN tout-génome sur des échantillons appariés augmente la puissance de détection de variants pertinents d'un point de vue biologique avec des échantillons de taille plus restreinte, et devrait être envisagée pour l'étude de caractères génétiques dits « complexes ».

iv. Les applications principales des outils génomiques chez le cheval

Les applications de ces outils biomoléculaires en génomique clinique chez le cheval sont nombreuses. Les principales restent actuellement le dépistage d'affections génétiques, le diagnostic génétique associé à la médecine personnalisée, ainsi que la gestion du cheval athlète.

- La génomique au service du dépistage d'affections génétiques (Brosnahan & al., 2010; Schibler, 2012; Finno, Bannasch, 2014; Bailey, 2014)
 - L'identification de gènes responsables d'affections monogéniques (Schibler, 2012)

Historiquement, la première application des travaux de génomique a été l'identification de mutations responsables d'affections monogéniques (affections liées à la mutation d'un seul gène). A l'heure actuelle, on recense pour l'espèce équine 49 caractères et affections monogéniques, parmi lesquels 37 ont d'ores et déjà une mutation causale identifiée, notamment les mutations touchant les loci liés aux différentes robes de chevaux. Une liste régulièrement actualisée et disponible au grand public les répertorie, tout en fournissant des informations sur les publications relatives à chaque caractère ou affection monogénique : elle se trouve sur la base de données *OMIA (Online Mendelian Inheritance for Animals)* de l'Université de Sydney (<http://omia.angis.org.au/home/>). Le Conseil et Observatoire Suisse de la Filière du Cheval (*COFICHEV*) présente également des informations à ce sujet, accessibles sur internet (<http://www.cofichev.ch/fr/Connaissances/Genetique-genomique/Maladies-hereditaires.html>). A quelques exceptions près, ces mutations ont été identifiées via une approche de gènes candidats, reposant sur la connaissance de syndromes héréditaires similaires chez l'Homme et la souris (Brosnahan & al., 2010). Les mutations responsables des 12 autres affections / caractères monogéniques restent donc à déterminer ou à publier ; plusieurs programmes sont en cours à cet effet, à l'échelle internationale (listés sur le site web français de l'I.N.R.A. de Jouy-en-Josas <http://dga.jouy.inra.fr/horse.genomics/hgpprojects.html> ou sur le site web américain de l'University of Kentucky <https://www.uky.edu/Ag/Horsemap/hgpprojects.html>). Le tableau ci-dessous présente plus d'une vingtaine d'affections monogéniques pour lesquelles la mutation causale a été identifiée.

Tableau II : Principales affections monogéniques à mutation causale identifiée

(Schibler, 2012; Bailey, 2014; site internet *OMIA* de l'Université de Sydney)

Mode de transmission : D = dominant, R = récessif, C = codominant, Y-linked = lié au chromosome Y

Maladie génétique (transmission)	Description	Gène (chromo)	Sources
HYPP (D)	Hyperkalemic Periodic Paralysis Periodic Paralysis II in Equus caballus (Hyperkaliémie périodique paralysante)	SCN4A (ECA11)	<i>(Rudolph & al., 1992a ; Rudolph & al., 1992b)</i>
OT-DSD (Y-linked)	Ovotesticular Disorder of Sexual Development Sex Reversal Syndrome, 64, XY, SRY-negative XY female (Réversion du sexe, 64, XY, SRY négatif)	SRY (ECA Y)	<i>(Pailhoux & al., 1995 ; Abe & al., 1999 ; Raudsepp & al., 2010 ; Pujar, Meyers-Wallen, 2012)</i>
SCID (R)	Severe Combined Immunodeficiency Disease (Syndrome d'immunodéficience sévère combinée)	DNA-PKcs (ECA9)	<i>(Bailey & al., 1997 ; Shin & al., 1997a, 1997b ; Bernoco, Bailey, 1998)</i>
OLWFS (R)	Overo Lethal White Foal Syndrome Ileocolonic aganglionosis, Megacolon... (Syndrome létal du poulain blanc - robe pie Overo)	EDNRB (ECA17)	<i>(Metallinos & al., 1998 ; Santschi & al., 1998 ; Yang & al., 1998 ; Bellone, 2010)</i>
JEB1, H-JEB (R)	Junctional Epidermolysis Bullosa 1 Herlitz Junctional Epidermolysis Bullosa (H-JEB) (Epidermolyse bulleuse jonctionnelle (Herlitz))	LAMC2 (ECA5)	<i>(Spirito, & al., 2002 ; Baird & al., 2003 ; Milenkovic & al., 2003 ; Graves & al., 2009)</i>
JEB2, EI (R)	Junctional Epidermolysis Bullosa 2 Epitheliogenesis imperfecta neonatorum (Epidermolyse bulleuse jonctionnelle 2)	LAMA3 (ECA8)	<i>(Lieto, Cothran, 2003 ; Graves & al., 2009)</i>
GBED (R)	Glycogen Branching Enzyme Deficiency Glycogen Storage Disease VI (GSD IV), Amylopectinosis (Déficiente en enzyme branchante du glycogène)	GBE1 (ECA26)	<i>(Valberg & al., 2001 ; Ward & al., 2004)</i>
MH (D)	(Equine) Malignant Hyperthermia (EMH) (Hyperthermie maligne)	RYR1 (ECA10)	<i>(Aleman & al., 2004 ; Aleman & al., 2009 ; Nieto, Aleman, 2009 ; McCue & al., 2009)</i>
GT (R)	Glanzmann's Thrombasthenia Thrombasthenia 1 (Thrombasthénie de Glanzmann)	ITGA2B (ECA11)	<i>(Christopherson & al., 2006 ; Christopherson & al., 2007 ; Sanz & al., 2011)</i>
HERDA (R)	Hereditary Equine Regional Dermal Asthenia Ehlers-Danlos Syndrome, Hyperelastosis Cutis (Asthénie dermique héréditaire localisée équine, Hyperélastose cutanée)	PPIB (ECA1)	<i>(Tryon & al., 2007)</i>
PSSM (D)	(Equine) Polysaccharide Storage Myopathy, Type 1 (Myopathie à stockage de polysaccharides, type 1)	GYS1 (ECA10)	<i>(McCue & al., 2008a, 2008b ; McCue & al., 2009 ; Herszberg & al., 2009 ; McCoy & al., 2014)</i>
Gray horse melanoma (D)	Gray horse melanoma Progressive greying with age, Melanome, Melanomatosis (Prédisposition au mélanome, mélanose du cheval gris)	STX17 (ECA25)	<i>(Rosengren Pielberg & al., 2008 ; Bellone, 2010)</i>
MCOA (C)	Multiple Congenital Ocular Anomalies - Coat color Silver (Anomalies congénitales oculaires multiples - Robe Silver)	PMEL17 (ECA6)	<i>(Andersson & al., 2008 ; Bellone, 2010 ; Andersson & al., 2011 ; Andersson & al., 2013)</i>
CSNB (R)	Congenital Stationary Night Blindness - Leopard	TRPM1 (ECA1)	<i>(Bellone & al., 2010 ; Bellone, 2010 ; Sandmeyer & al., 2012 ;</i>

	(Leopard complex / print / spotting, Appaloosa Coat Pattern...) (Cécité nocturne / héméralopie congénitale non évolutive - robe tachetée)		<i>Bellone & al., 2013)</i>
LFS (R)	Lavender Foal Syndrome - Coat Color Dilution Lethal (CCDL) (Syndrome du poulain lavande)	MYO5A (ECA1)	<i>(Brooks & al., 2010 ; Bellone, 2010 ; Gabreski & al., 2012)</i>
CA (R)	Cerebellar (Cortical) Abiotrophy Neonatal Cerebellar Cortical Degeneration (NCCD) (Abiotrophie / atrophie (corticale) cérébelleuse)	TOE1 ou MUTYH (ECA2)	<i>(Brault & al., 2011a, 2011b ; Cavalleri & al., 2013)</i>
FIS (R)	Foal Immunodeficiency Syndrome (in the Fell and Dales Pony) Fell Pony Foal Syndrome (FPFS) (Syndrome d'immunodéficience du poulain / poney Fell)	SLC5A3 (ECA26)	<i>(Fox-Clipsham & al., 2011 ; Bailey, 2011)</i>
MY (R)	Congenital Myotonia Equine Myotonic Dystrophy, Myotonia congenita (Myotonie congénitale)	CLCN1 (ECA4)	<i>(Wijnberg & al., 2012)</i>
AIS (R)	Androgen Insensitivity Syndrome - 64, XY, SRY-positive Testicular feminization (XTfmY) (Syndrome d'insensibilité aux androgènes - 64, XY, SRY positif)	AR (ECAX)	<i>(Révay & al., 2012 ; Pujar, Meyers-Wallen, 2012)</i>
IP (D)	(Equine) Incontinentia Pigmenti - Brindle Hyperpigmentation (Incontinentia Pigmenti, Robe bringée)	IKBKG (ECAX)	<i>(Towers & al., 2013)</i>
CLD (D)	Chondrodysplasia-Like Dwarfism ACAN Dwarfism (Nanisme)	ACAN (ECA1)	<i>(Eberth, 2013)</i>
CLF (R)	Congenial Liver Fibrosis Caroli Liver Fibrosis (Fibrose hépatique congénitale, Syndrome de Caroli)	PKHD1 (ECA20)	<i>(Drögemüller & al., 2014)</i>
WFFS (R)	Warmblood Fragile Foal Syndrome (Type 1) Ehlers-Danlos-Like Syndromes (EDLS) (Syndrome du poulain de sang fragile)	PLOD1 (ECA2)	<i>(Winand, 2012 ; Monthoux & al., 2015)</i>
HWSD (R)	Hoof Wall Separation Disease Hoof Wall Separation Syndrome (HWSS) in Connemara Ponies (Maladie / syndrome de séparation de la paroi du sabot)	SERPINB11 (ECA8)	<i>(Finno & al., 2015)</i>
Hydrocephalus (R)	Hydrocephalus (Hydrocéphalie)	B3GALNT2 (ECA1)	<i>(Ducro & al., 2015)</i>

Le clonage positionnel des affections à déterminisme dominant (la présence d'une seule copie mutée suffit à entraîner le développement de la maladie) nécessite des dispositifs de quelques centaines d'animaux répartis en quelques grandes familles. L'hyperkaliémie périodique paralysante (HYPP) des chevaux Quarter Horse pure race ou croisés, American Paint Horse et Appaloosa, est une illustration de ce type d'affection sélectionnée involontairement parce qu'associée à une hypertrophie musculaire recherchée par les éleveurs. Les chevaux malades sont porteurs d'une mutation ponctuelle faux-sens (substitution de cytosine par la guanine) dans le gène *SCN4A*, entraînant le remplacement d'un acide aminé par

un autre (phénylalanine remplacée par la leucine) dans la structure protéique que code ce gène : la sous-unité α du canal sodique voltage-dépendant du myocyte squelettique, une glycoprotéine essentielle à l'activation électrique des cellules musculaires (*Rudolph & al., 1992*). Chez les chevaux atteints, la protéine ainsi altérée présente un défaut de fermeture (entrée accrue d'ions sodium), ce qui se traduit à terme par une diminution du potentiel membranaire de repos, devenant très proche du potentiel d'action, donc à un état de dépolarisation prolongée des myocytes. Ils présentent, dans certaines conditions, des phases de tremblements et fasciculations musculaires (épisode de myotonie quand le niveau de dépolarisation est tel que l'excitabilité des fibres musculaires est encore possible) pouvant aller jusqu'à la paralysie, ou une faiblesse musculaire et hypotonie généralisées (asthénie liée à l'inexcitabilité membranaire résultant de la dépolarisation accrue), cette deuxième phase faisant souvent suite à la première et caractérisée par un décubitus involontaire chez un animal encore conscient. Cette situation où une affection monogénique est accidentellement sélectionnée car associée à un phénotype d'intérêt pour les éleveurs n'est pas uniquement retrouvée dans l'espèce équine : en effet, si l'on prend l'exemple des bovins, la mutation du gène de la myostatine est connue pour le phénotype « culard » auquel elle est liée, caractérisé par un fort développement musculaire. En sélectionnant l'allèle muté de ce gène, on a involontairement sélectionné les difficultés au vêlage (déséquilibre pondéral et inadéquation morphologique foeto-maternels à la mise-bas), les troubles de la reproduction (comportement sexuel des jeunes affecté, diminution de la fertilité et de la production de lait chez les femelles culardes, diminution de la viabilité des veaux culards...) et les problèmes de susceptibilité au stress des animaux homozygotes pour cette mutation (*Allais, 2011*).

Les affections récessives sont plus insidieuses, puisque les animaux porteurs d'un seul allèle défavorable sont sains et seuls leurs descendants homozygotes pour l'allèle défavorable peuvent déclarer la maladie. L'identification des bases moléculaires de telles maladies s'avère particulièrement efficace et requiert l'analyse d'une dizaine d'individus atteints (*Schibler, 2012*). Une délétion dans le gène codant la myosine Va (*MYO5A*) a ainsi été identifiée en analysant six poulains de race arabe atteints du syndrome du poulain lavande (*LFS*) et 30 chevaux sains présentant un lien de parenté fort avec ces poulains, à l'aide de la puce *SNP 50K* (*Brooks & al., 2010*).

- Un manque d'informations épidémiogénétiques (*Schibler, 2012*)

Peu d'anomalies génétiques sont décrites chez le cheval en comparaison des autres espèces (49 anomalies mendéliennes équines recensées par *OMIA*, contre 216 chez les bovins et 277 chez le chien actuellement). Cela peut s'expliquer par la sélection sur les performances sportives dans l'espèce équine, privilégiant les individus en excellente santé et de conformation irréprochable. Cependant, de nombreuses affections résultent de mutations *de novo* (se produisant lors de la formation des gamètes) et devraient donc être répertoriées chez le cheval (*Schibler, 2012*). Leur absence dans la base de données *OMIA* suggère ainsi plutôt un déficit de signalement et de recueil des informations, à l'origine d'une méconnaissance du déterminisme génétique de nombreuses maladies. Ce déficit expliquerait le faible nombre de gènes identifiés chez le cheval comme responsables d'affections ou de caractères monogéniques, alors que 3212 gènes de ce type sont déjà recensés chez l'Homme sur *OMIM* (*Online Mendelian Inheritance in Man* (<http://www.omim.org/>)).

- L'opportunité du conseil génétique grâce au testage (*Schibler, 2012*)

La génomique fournit des outils permettant de faire face à l'émergence de certaines affections héréditaires en évitant des « situations génétiques à risque », et de générer des connaissances utiles pour prévenir la manifestation clinique de ces affections.

Des tests génétiques commerciaux sont ainsi réalisés à partir d'un simple prélèvement sanguin, afin de détecter de manière précoce les animaux porteurs. Dans le cas d'affections récessives, le dépistage permet de mettre en œuvre un conseil génétique comme alternative à l'éradication : en évitant les accouplements à risque, on réduit l'incidence de la maladie, sans pour autant perdre les qualités génétiques des animaux porteurs.

La connaissance précoce d'une prédisposition génétique à une maladie peut également permettre de prévenir son développement clinique, par l'adoption d'une conduite d'élevage ou d'une gestion des animaux adéquate (régime alimentaire, type d'exercice...). On peut par exemple réduire les crises d'HYPP en réduisant la teneur en potassium de la ration, en privilégiant un entraînement modéré régulier et en évitant d'exposer les chevaux au froid.

Dans son article, Laurent Schibler met donc en lumière le rôle des praticiens vétérinaires équins dans la recherche et le conseil génétique, à la fois en tant qu'acteurs épidémiologiques, prescripteurs de tests génétiques, et conseillers pour une gestion pertinente de chaque cheval en fonction de ses données génomiques.

➤ La génomique au service du diagnostic clinique et de la médecine personnalisée (Schibler, 2012)

○ La prédisposition aux affections « complexes » (Schibler, 2012)

De nombreuses affections présentent une composante génétique, mais sont dites complexes car déterminées par de nombreux facteurs génétiques et environnementaux, qui, considérés isolément, ont des effets modestes sur le risque d'apparition des symptômes. La génomique et la transcriptomique peuvent permettre d'identifier des biomarqueurs liés à la prédisposition à une maladie : marqueurs génétiques, protéines circulantes, niveaux d'expression de gènes... Ces données ouvrent la voie au concept novateur d'une « génomique personnalisée », c'est-à-dire à une prédiction individuelle du risque.

Quelques études des déterminants génétiques d'affections équine « complexes » sont en cours au niveau mondial ; parmi ces affections, de nombreuses sont bien connues dans le milieu équin et équestre. La *RAO* (*Recurrent Airway Obstruction*), première maladie respiratoire chronique chez le cheval, aussi appelée emphysème ou pousse, en fait partie ; cette maladie inflammatoire chronique des voies respiratoires profondes, d'origine allergique, est caractérisée par de la toux, une intolérance à l'effort et de la détresse respiratoire au repos. On pourra aussi citer le *SME* (*Syndrome Métabolique Equin*) ou *EMS* (*Equine Metabolic Syndrome*), trouble endocrinien multifacettes se manifestant par une obésité et une résistance à l'insuline, prédisposant à la fourbure. Le tympanisme des poches gutturales des poulains de race Arabe et issus de cette race est une dysfonction du clapet cartilagineux fermant l'orifice pharyngien du tube auditif qui n'est également pas étrangère aux éleveurs de ce type de chevaux. L'uvéite récidivante équine (*URE*), maladie oculaire chronique et cause la plus commune de cécité chez le cheval, serait, elle aussi, sous un certain déterminisme génétique ; l'uvéite isolée étant un vice rédhibitoire, il s'agit d'une maladie que le vétérinaire équin doit impérativement connaître et savoir reconnaître, notamment dans un contexte de visite d'achat. Les propriétaires et cavaliers connaissent aussi en général le syndrome de dermatite estivale récidivante équine (*DERE*) déjà évoqué dans cette thèse, affection dermatologique d'origine allergique à base héréditaire. L'immunité vis-à-vis de certains agents infectieux présente une composante génétique encore imparfaitement élucidée, qu'il s'agisse d'agents viraux (*Herpesvirus équin de type 1 (EHV1)*, ou *Papillomavirus* responsable de sarcoïdes par exemple), bactériens (*Rhodococcus equi* par exemple, causant des infections auxquelles les poulains sont très sensibles), ou parasitaires. Enfin, on mentionnera les affections locomotrices

présentant un déterminisme génétique « complexe » fréquentes comme : i) les lésions du tendon fléchisseur superficiel du doigt et les lésions du ligament suspenseur du boulet chez le cheval de course pur-sang anglais, ii) le syndrome podo-trochléaire ou maladie naviculaire (lié à des lésions dégénératives de l'os sésamoïdien distal, aussi appelé os naviculaire, souvent associées à des lésions des formations anatomiques voisines comme le tendon fléchisseur profond du doigt, les ligaments sésamoïdiens, les synoviales adjacentes...), une des causes les plus répandues de boiterie chronique antérieure chez le cheval, iii) l'ostéochondrose dans la race Trotteur Français, objet du programme de recherche *GenEquin* mené conjointement par l'Ecole Nationale Vétérinaire d'Alfort (ENVA), le Centre d'Imagerie et de Recherche sur les Affections Locomotrices Equines (CIRALE), l'Institut National de la Recherche Agronomique (INRA) et l'Institut Français du Cheval et de l'Equitation (IFCE) et financé par l'Agence Nationale de la Recherche, l'IFCE et le Fonds Eperon ; ou plus rares comme la lordose précoce des chevaux de selle américains (American Saddlebreds), malformation congénitale des vertèbres vraisemblablement sous contrôle génétique.

Les génomiciens ont encore de gros progrès à réaliser pour améliorer la qualité et la précision des prédictions, mais le concept répond à une attente de la filière équine dans la prévention des affections. Le développement de ces approches ne peut cependant s'envisager qu'à la faveur d'études transversales intégrant sciences en « omique », zootechnie et recherche clinique, car la prédiction individuelle n'a d'intérêt que si elle est couplée à des préconisations permettant de prévenir l'apparition de la maladie. Là encore, on aperçoit le rôle du vétérinaire clinicien et praticien comme vétérinaire conseil : c'est en effet le mieux placé pour proposer des ajustements pertinents de la conduite des élevages ou d'élaborer des stratégies de traitement novatrices au vu de l'information génomique.

- Des outils de diagnostic et de pronostic (Schibler, 2012)

Les innovations dans les sciences « omiques », notamment les progrès importants en génomique fonctionnelle, l'apport de la NGS, et l'essor de la métabolomique, ont engendré un regain d'intérêt pour la découverte et l'utilisation de biomarqueurs génomiques chez l'Homme et les animaux. Définis comme « indicateurs de processus biologiques normaux, de processus pathogéniques ou de réponses pharmacologiques à une intervention thérapeutique, mesurés objectivement, et censés remplacer un critère d'évaluation clinique (prédire l'aspect bénéfique ou nocif/préjudiciable) sur la base de preuves épidémiologiques, thérapeutiques, physiopathologiques ou d'autres preuves scientifiques » (*Biomarkers Definitions Working Group*,

2001), ils possèdent une grande diversité de fonctions et leurs applications dans le domaine médical sont multiples. Ils apportent, entre autres, une facilité d'accès à de nouveaux dosages, et sont à l'origine d'améliorations conséquentes dans la détection précoce de maladies, le suivi des traitements, le pronostic ou les études de toxicité (Beger, 2007; Hong & al., 2010; Ganesalingam, Bowser, 2010; Rosenson, 2010; Hong, Slikker, 2015; Zhang & al., 2015). Ils sont à la base de la « théranostique » (« *theranostics* »), discipline née de l'union entre la thérapie et le diagnostic (Jain, 2005; Netto, Epstein, 2010), et s'inscrivent dans une démarche de stratification des patients et de médecine personnalisée, permettant « d'ajuster les traitements médicaux à différentes sous-populations de patients » partageant des caractéristiques communes de « sensibilité vis-à-vis de maladies particulières ou de réponse à un traitement spécifique » (définition proposée en 2008 par le *President's Council of Advisors on Science and Technology* dans un rapport intitulé *Priorities for Personalized Medicine*, accessible en ligne via le lien suivant : https://www.whitehouse.gov/files/documents/ostp/PCAST/pcast_report_v2.pdf). Ces approches sont particulièrement développées dans le domaine de l'oncologie (cancers de la prostate, du sein, du poumon, du colon et du rectum), dans le but de classer et stratifier les types de tumeurs via des profils d'expression, pour mieux prédire leur évolution, ajuster les traitements en définissant la stratégie thérapeutique la plus pertinente, et prévenir les risques de rechute. Depuis quelques années, l'intérêt se porte particulièrement sur les miARN car le sérum et le plasma contiennent de grandes quantités de miARN dérivés de différents tissus ou organes (Chen & al., 2008). Ces miARN constituent ainsi d'excellents biomarqueurs non invasifs, en améliorant à la fois précocité du diagnostic et précision du pronostic chez l'Homme, non seulement dans les cas de cancers (Trang & al., 2008; Weiland & al., 2012; Liu & al., 2013), mais également d'affections ostéo-articulaires (arthrose, etc.), cardio-vasculaires, musculaires, hépatiques, endocrines, immunitaires (Zhang, 2008; Weiland & al., 2012)... Un article de revue (Weiland & al., 2012) liste notamment les miARN circulants biomarqueurs de différentes conditions pathologiques (états cancéreux, affections cardiaques, hépatopathies, infections, troubles psychiatriques) et de certaines conditions physiologiques (gestation) humaines, recensés jusqu'à sa publication. La tendance actuelle, aussi bien en oncologie que dans les autres domaines médicaux, est à l'intégration de toutes ces données omiques, afin d'identifier des « signatures moléculaires complexes » (ARNm, miARN, protéines et métabolites) (Devaux & al., 2010; Liu & al., 2011; Cavill & al., 2011; Liu & al., 2013; Yeoman & al., 2013; Hong, Tong, 2014). En effet, l'application de la génomique à la biologie évolue désormais de la découverte de déterminants génétiques (génomique structurale) à ces « signatures biologiques » (génomique fonctionnelle et autres « omiques »), c'est-à-dire un motif typique d'expression génétique, ou même un profil

protéique ou de métabolites typique, caractérisant un état physiologique ou pathologique spécifique.

Des approches similaires se dessinent peu à peu chez le cheval, peut-être à la faveur du contexte sportif équin actuel où les contraintes du contrôle antidopage accroissent l'intérêt pour une thérapeutique plus ciblée et plus efficace. On met ainsi à profit les connaissances émergentes des microtranscriptome, protéome et métabolome équins dans des états physiologiques et pathologiques.

Actuellement, 715 séquences de précurseurs de miARN équins sont recensées, et on compte 690 miARN équins matures ([http://www.mirbase.org/cgi-bin/mirna_summary.pl?org=eca, \[EquCab2.0\]](http://www.mirbase.org/cgi-bin/mirna_summary.pl?org=eca, [EquCab2.0]))). Les séquences codant des miARN sont souvent regroupées en clusters sur les chromosomes (Zhou & al., 2009) ; on en trouve sur tous les chromosomes sauf ECA31. La base de données microtranscriptomiques équines est régulièrement mise à jour, notamment par l'apport des données issues des méthodes novatrices comme les NGS ; une étude relativement récente a ainsi permis d'enrichir cette base de données par la caractérisation de 292 miARN équins déjà connus et l'identification de 329 nouveaux miARN équins dans différents tissus (dont le muscle squelettique, le colon et le foie) en conditions physiologiques, à l'aide de séquenceurs haut débit de la société IlluminaTM : il s'agit de la première analyse expérimentale ayant identifié et caractérisé le profil d'expression global des miARN dans des tissus équins sains (Kim & al., 2014). L'afflux de telles connaissances commence à être mis à profit dans des domaines variés, notamment de santé équine :

i) en élevage : les polymorphismes affectant des séquences codant des miARN sont de nouveaux biomarqueurs de caractères phénotypiques ; leur analyse et leur comparaison entre différentes espèces, notamment celles de rente, a pour objectif de servir à la recherche dans le domaine des productions animales afin de sélectionner de manière efficace des *SNP* sur les séquences codant des miARN et/ou des *QTL* d'expression de miARN pour une évaluation génétique quantitative et moléculaire plus poussée de leurs effets phénotypiques et de leurs associations causales avec les caractères de production (Jevsinek Skok & al., 2013) ;

ii) en reproduction : en ce qui concerne la jument, l'étude de miARN présents à différents niveaux d'expression dans des microvésicules et exosomes du liquide folliculaire ovarien équin suggère leur implication dans la communication cellulaire lors du développement folliculaire chez les chevaux (Schauer & al., 2013) ; côté mâle, une étude d'association a révélé la présence de polymorphismes affectant des régions non-codantes (en d'autres termes, introns) dans le gène *PLCz1* (codant la protéine phospholipase C zeta 1) : ces *SNP* seraient impliqués dans la fertilité

de l'étalon, et *PLCz1* serait un locus candidat pour la fertilité mâle chez les chevaux Hanovriens (Schrimpf & al., 2014) ;

iii) en pathologie ostéo-articulaire : la caractérisation du miRNome (microtranscriptome, ensemble des miARN) du cartilage épiphysaire et de l'os sous-chondral équins par *NGS* a mis en évidence des différences d'expression des miARN entre l'os et le cartilage chez 3 chevaux sains et 3 poulains de race Anglo-Arabe atteints d'ostéochondrose, ce qui les place comme potentiels biomarqueurs de l'intégrité cartilagineuse et osseuse (via le contrôle du cycle et de la différenciation cellulaire, de la production énergétique, du métabolisme, et de la structure et dynamique de la matrice extracellulaire) (Desjardin & al., 2014) ; parmi ces miARN, le rôle prépondérant de miR-140 comme miARN spécifique du tissu cartilagineux est mis en lumière, avec quelques cibles impliquées dans le développement et l'homéostasie du cartilage (Buechli & al., 2012) ;

iv) dans l'étude du métabolisme associé à l'exercice physique, sur laquelle nous reviendrons plus tard ;

v) dans la pathologie musculaire : le rôle majeur des miARN dans la prolifération et la différenciation des myoblastes ayant été démontré chez l'Homme et la souris, on a suspecté que leur profil d'expression soit lié au statut musculaire dans l'espèce équine. Des approches protéomiques et microtranscriptomiques ont été menées de front dans la recherche de biomarqueurs en relation avec différentes myopathies comme les rhabdomyolyses récurrentes à l'effort (*RER*, *recurrent exertional rhabdomyolysis*, affection caractérisée par une augmentation de concentration des enzymes musculaires plasmatiques, la libération de myoglobine, et par une inflammation ou nécrose des myocytes de type II) et les myopathies par rétention/stockage de polysaccharides (*PSSM*, *polysaccharide storage myopathy*, maladie caractérisée en histopathologie par le stockage élevé de glycogène dans les myocytes au sein de nombreuses vacuoles et sous forme d'inclusions positives à la coloration à l'Acide Périodique de Schiff, situées sous le sarcolemme (membrane plasmique des myocytes striés squelettiques)). La recherche de biomarqueurs en relation avec la *RER* a ainsi été réalisée par électrophorèse bidimensionnelle (*2-DE*), identifiant 9 protéines surexprimées et 4 sous-exprimées dans les biopsies de muscle de chevaux atteints de *RER* (Bouwman & al., 2010). Une étude de 10 miARN, réalisée sur des biopsies musculaires de chevaux cobs normands (sains ou atteints de *PSSM*) et trotteurs français (sains ou atteints de *RER*), a, quant à elle, montré des différences d'expression de ces miARN entre races, ainsi que des profils d'expression distincts pour chacune de ces deux myopathies : ces profils d'expression spécifiques dans le muscle *gluteus medius* ont de fait été associés respectivement à la *PSSM* et à la *RER* (Barrey & al., 2010) ;

cette information pourrait être utilisée dans l'établissement du diagnostic différentiel et contribuer de cette manière à améliorer la définition d'entités nosologiques.

vi) en pathologie respiratoire : comme vu plus haut, la *RAO* est une maladie d'origine allergique, se rapprochant de l'asthme humain, qui affecte couramment les chevaux dans le monde entier (*Robinson, 2001*). Or, des différences d'expression des miARN ont déjà été montrées dans le sérum de patients humains asthmatiques (*Panganiban & al., 2012*), dans les exosomes isolés à partir de prélèvements par lavage broncho-alvéolaire réalisés sur ces individus (*Eissa, 2013*), et dans les lymphocytes T de leurs voies respiratoires (*Simpson & al., 2014*) ; on peut donc s'attendre à ce que les profils de miARN fassent évoluer notre compréhension du mécanisme de la *RAO* équine. La maladie se déclarant chez des chevaux d'âge mûr, l'identification de biomarqueurs précoces tels que les miARN pourrait permettre de la diagnostiquer à ses premiers stades, avant l'apparition des symptômes (*van der Kolk & al., 2015*). Il est intéressant de savoir que l'effet thérapeutique de l'antagomir (miARN antagoniste d'un autre miARN) antimir-126 a déjà été rapporté sur le modèle murin de l'asthme (*Mattes & al., 2009*).

vii) enfin, dans d'autres troubles médicaux équins, des profils plasmatiques obtenus par spectrométrie de masse pourraient être utilisés, afin de confirmer ou de préciser un diagnostic d'intoxication, par le sénéçon de Jacob par exemple (*Moore & al., 2008*), ou d'infection (*Chaussabel & al., 2010*).

Par ailleurs, les séquences nucléotidiques spécifiques de pathogènes constituent autant de biomarqueurs génomiques utilisables dans le diagnostic des maladies infectieuses ; les biopuces pourraient ainsi permettre de détecter par hybridation des séquences spécifiques de pathogènes et de conclure à une contamination ou une infection, même très faible. Le séquençage massivement parallèle devenant de plus en plus abordable, son utilisation en routine pourrait jouer dans l'amélioration du suivi global de l'état sanitaire des animaux par des approches de métagénomique (par le diagnostic des maladies infectieuses et des parasitoses, l'étude des écosystèmes bactériens de la flore intestinale en lien avec l'âge, l'alimentation ou la maladie...) (*Schibler, 2012*).

Enfin, la métabolomique, seule ou utilisée en complément aux autres « omiques », vient progressivement enrichir les études en santé équine. Bien que les travaux en métabolomique équine se soient jusqu'à présent beaucoup concentrés sur la détection de substances dopantes dans le domaine de la course (*Scarth & al., 2010; Liu & al., 2011*), on trouve des analyses métabolomiques utilisant le cheval comme modèle pour l'étude de la réponse à différents stimuli dans le cadre de maladies métaboliques. Pour ne citer qu'un exemple, une analyse métabolomique utilisant la RMN sur un modèle d'induction de fourbure par une surcharge en

oligofructose (sucre utilisé dans les aliments administrés aux chevaux et habituellement considéré comme indigestible dans l'intestin grêle) a confirmé l'augmentation de lactatémie plasmatique caractéristique des stades primaires de fourbure équine, et a surtout révélé des modifications biochimiques sanguines auparavant inconnues, notamment une augmentation de concentration plasmatique en oligofructose et en certains phospholipides, modifications métaboliques que les auteurs ont liées à la rupture de barrière muqueuse intestinale et à l'augmentation de perméabilité intestinale se manifestant au cours du développement d'une fourbure (Keller & al., 2011). Certains autres auteurs ont publié des analyses métaboliques de biofluides équins par rapport à l'âge (Ralston & al., 2011) ou à un état pathologique comme l'ostéochondrose (Ralston & al., 2009). Une étude a aussi exploré la variabilité métabolique selon la race du cheval et a révélé des différences significatives importantes, probablement liées à une vitesse de croissance et une utilisation protéique différentes (Pappalardo & al., 2013). Tous ces travaux métabolomiques alimentent le potentiel énorme d'exploration de variabilité physiologique et pathologique équine. Cependant, on constate encore un manque de données de base sur le phénotype métabolique des chevaux, manque que certains auteurs ont voulu combler en réalisant un atlas des métabolites de biofluides courants équins (Escalona & al., 2015) : cette étude a permis de dégager les composants dominants du métabolome de chevaux Pur Sang par un profilage métabolique en spectroscopie RMN-¹H de plasma, urine et liquide fécal de chevaux de course en bonne santé, fournissant ainsi un guide de référence préliminaire sur le milieu métabolique physiologique équin. Ainsi, 102 pics ont été attribués à des métabolites pour l'ensemble des biofluides et un « métabolome noyau » constitué de 14 métabolites communs aux trois matrices biologiques a été déterminé. Ce noyau regroupe des éléments du métabolisme énergétique tels que le glucose et le lactate, un nombre important d'acides aminés dont l'alanine, l'arginine, la glycine, la glutamine, la taurine, la thréonine et la valine, ainsi que des métabolites liés au microbiote comme l'acétate, le formate et le *p*-hydroxyphénylacétate.

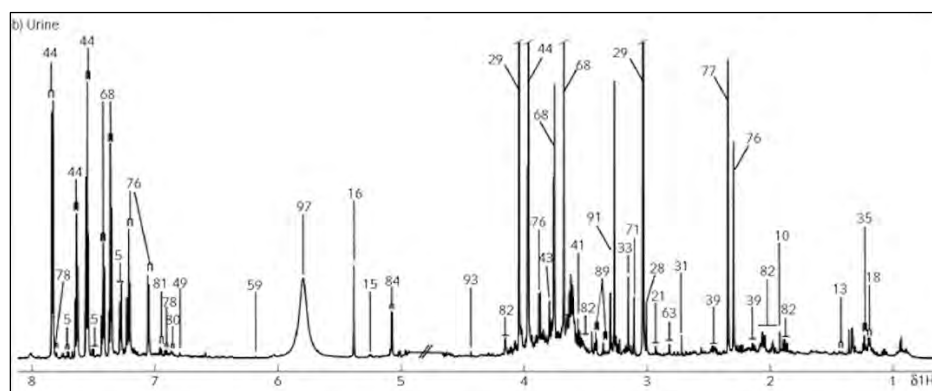


Figure XXVIII : Spectre métabolomique de l'urine équine en RMN 1H 700 MHz 1D (Escalona & al., 2015)

Chaque numéro correspond à un métabolite (se référer à l'article correspondant, Table 1).

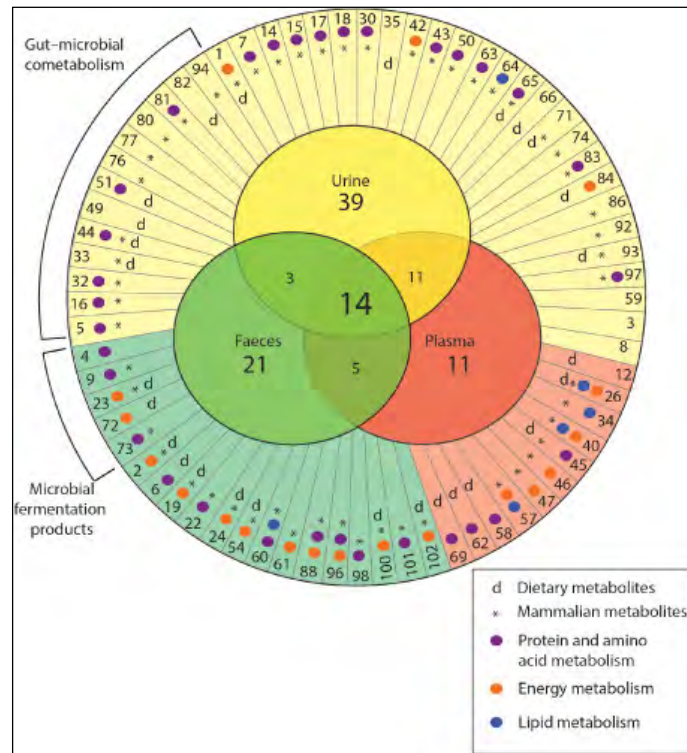


Figure XXIX : Diagramme de Venn mettant en évidence les similitudes et différences métaboliques entre compartiments biologiques (Escalona & al., 2015)

Section centrale : nombre de métabolites « noyaux » présents dans les trois matrices biologiques (14 métabolites communs).

Partie périphérique découpée en portions : métabolites spécifiques à chaque biofluide (rouge : plasma, jaune : urine, vert : fecès), ordonnés selon leur origine/fonction, numérotés selon la Table 1 de l'article correspondant).

Pois colorés selon les fonctions métaboliques : métabolisme protéique et des acides aminés en violet, métabolisme énergétique en orange, métabolisme lipidique en bleu.

** : métabolites pouvant être d'origine mammalienne.*

d : métabolites pouvant être d'origine alimentaire (dietary).

L'étude d'E. Escalona & al. dégage des similarités métaboliques entre chevaux et hommes : en effet, 32 métabolites plasmatiques, 23 métabolites urinaires et 27 métabolites fécaux détectés lors de cette étude sont retrouvés dans ces deux espèces (Nicholson & al., 1995; Holmes & al., 1997; Le Gall & al., 2011). Le fait qu'une partie plus restreinte des profils métaboliques fécal et urinaire soit partagée par ces deux espèces que le profil métabolique plasmatique est lié à leurs systèmes digestifs et voies métaboliques différents ; si l'on compare les ruminants aux chevaux, on remarque que de nombreux métabolites liés au microbiote intestinal sont communs à ces herbivores, qui présentent des stratégies digestives semblables aboutissant à l'obtention de nutriments à partir d'une alimentation riche en fibres (Ametaj & al., 2010; Nyberg & al., 2010; Li & al., 2011). La variabilité interindividuelle inhérente aux échantillons fécaux, la nature invasive des prélèvements sanguins ainsi que l'abondance relative de métabolites liés au co-métabolisme entre mammifère hôte et microbiote intestinal dans l'urine (65 métabolites identifiés, parmi lesquels 39 spécifiques de cette matrice biologique, avec la plupart de ceux-là liés au co-

métabolisme mammifère hôte/micro-organisme symbiote) placent cette dernière comme biofluide de choix pour les objectifs de l'analyse métabolomique et fenêtre biologique sur les interactions métaboliques hôte/symbiote, comme cela a été montré pour d'autres espèces que le cheval (*Merrifield & al., 2011*). Les profils métabolomiques sanguins, quant à eux, sont plus susceptibles d'être représentatifs de processus physiologiques chez l'hôte (*Escalona & al., 2015*).

o Les outils thérapeutiques du futur ?

L'observation de signatures -omiques d'un état pathologique peut aider à l'identification de nouvelles molécules thérapeutiques : effectivement, la mise en évidence, par un screening pharmacologique *in vitro*, de composés produisant des signatures inverses à celles observées dans le tissu atteint permet de pointer du doigt des molécules candidates, susceptibles de contrecarrer les effets négatifs de la maladie (*Schibler, 2012*).

Les miARN, en particulier, se révèlent prometteurs en thérapeutique. En effet, leur action au stade post-transcriptionnel comme inhibiteurs de la traduction ou inducteurs de dégradation d'ARNm cibles (*Bronze-da-Rocha, 2014*) serait intéressante à mettre à profit dans le traitement de certaines maladies. En ciblant les miARN liés à une maladie par l'utilisation d'oligonucléotides anti-miARN (inhibiteurs de ces miARN), on pourrait contrecarrer la surexpression de ces miARN spécifiques à la maladie, donc lutter contre cette dernière (*Zhang, 2008*). L'intérêt de cibler les miARN plutôt que les ARNm ou des protéines spécifiques est évident : de nombreux gènes étant impliqués dans une même voie métabolique, cibler le miARN qui régule tous ou une majeure partie de ces gènes (*Bronze-da-Rocha, 2014*) semble d'une part plus simple que de lutter contre l'expression de chacun de ces gènes, et d'autre part plus efficace que de cibler une ou quelques protéines spécifiques de la maladie étudiée (*Caroli & al., 2013*). Cela dit, le risque de ce type d'approche est de réguler, par la même occasion, des gènes sur lesquels on ne souhaite pas intervenir. Les miARN peuvent être régulés par des oligonucléotides composés de nucléotides mimétiques à forte affinité pour ces miARN, désignés *miR-mimics*, et des oligonucléotides simple-brin antisens, nommés *antimiRs* ou *antagomirs*. Les *miR-mimics* sont utilisés afin d'augmenter le niveau d'un miARN spécifique lorsqu'une certaine affection est à l'origine d'une sous-expression de ce miARN ; ce sont des petites séquences nucléotidiques de synthèse, semblables aux pre-miARN, reconnus par la machinerie de biogenèse des miARN ou pris en charge par le complexe RISC (présenté par la suite), se comportant comme le miARN endogène d'intérêt et bloquant l'expression génétique (*Galasso & al., 2010; Fasanaro & al., 2010; Oliveira-Carvalho & al., 2013; Caroli & al., 2013; Bronze-da-*

Rocha, 2014). À l'inverse, les *antagomirs* sont utilisés quand les conditions pathologiques résultent de la surexpression d'un miARN particulier ; ce sont des oligonucléotides antisens modifiés (*morpholinos*) de synthèse qui ciblent la séquence de miARN mature, augmentant par conséquent la quantité d'ARNm cible de ce miARN particulier (Weiler & al., 2006; Galasso & al., 2010; Fasanaro & al., 2010; Bronze-da-Rocha, 2014). Ils peuvent exercer leur fonction à différents stades de la biogenèse des miARN en interférant avec les pri-miARN (précurseurs des pre-miARN) et avec l'export de pre-miARN (transcrit intermédiaire) (Caroli & al., 2013). Enfin, on dispose d'autres moyens pour inhiber la fonction des miARN, par des mécanismes appelés *sponges*, *masking* et *erasing*. Les *sponges* (éponges) empêchent la fixation du miARN sur sa cible en agissant comme inhibiteurs compétitifs. Le *masking* (masquage) utilise des oligonucléotides qui empêchent l'action d'un miARN particulier grâce à leur complémentarité parfaite au miARN en question. Enfin, les *erasers* (effaceurs) sont aussi des oligonucléotides à séquences répétées en tandem spécifiques de la séquence antisens complémentaire du miARN d'intérêt (Galasso & al., 2010; Fasanaro & al., 2010; Oliveira-Carvalho & al., 2013; Caroli & al., 2013; Bronze-da-Rocha, 2014). Ces miARN font leur entrée dans le domaine de la thérapeutique humaine, il sera donc intéressant de suivre leur étude à des fins thérapeutiques dans l'espèce équine.

- La génomique au service de la gestion du cheval athlète (Schibler, 2012)
 - La caractérisation des aptitudes (Schibler, 2012)

L'aptitude sportive est un caractère phénotypique synthétique complexe, abordé par la mesure de plusieurs paramètres : le modèle, les allures et les performances en compétition. L'obtention d'une information fiable sur le potentiel génétique d'un cheval demande du temps, puisqu'il faut attendre de pouvoir observer ses performances, voire celles de sa descendance, afin de déterminer s'il possède une combinaison favorable d'allèles des gènes d'intérêt (Schibler, 2012). Les sciences omiques appliquées à l'analyse des performances et à la physiologie de la réponse à l'entraînement permettent d'identifier des éléments (gènes clés, transcrits, séquences nucléotidiques, protéines, métabolites) ayant un rôle majeur. La génomique, en particulier, permet en outre d'évaluer directement les allèles transmis par les parents pour obtenir une caractérisation du poulain dès sa naissance. L'analyse globale du génome permet de s'affranchir de la connaissance préalable des gènes clés en exploitant la totalité des 74 000 marqueurs génétiques disponibles sur la puce de génotypage *SNP* 74K équine. Cette caractérisation

génomique repose sur le génotypage et le phénotypage d'une population de référence, afin d'établir des équations de prédiction appliquées au reste de la population après génotypage, pour évaluer le potentiel des animaux non phénotypés.

La sélection sur des caractères de performance a été et est encore la plus étudiée chez les chevaux Warmblood (regroupant différentes races de chevaux de sang) (Stock, Distl, 2008; Viklund & al., 2010; Schröder & al., 2012), ainsi que chez les pur-sangs anglais de course (Gu & al., 2009; Tozaki & al., 2010; Hill & al., 2010).

Chez les Warmbloods, la sélection des reproducteurs était orientée sur des caractères de conformation, performance et santé radiographique des membres (Stock, Distl, 2008). Les héritabilités concernant le saut d'obstacle varient de 0,39 à 0,61 chez les Hanovriens (Stock, Distl, 2007) et de 0,12 à 0,28 chez les chevaux de selle suédois (Swedish Warmblood) (Viklund & al., 2010), ce qui est relativement faible. Une étude d'association concernant l'aptitude au saut d'obstacle chez les Hanovriens a récemment identifié 6 régions *QTL* contenant des gènes connus chez l'Homme comme liés à la performance (Schröder & al., 2012).

En ce qui concerne les pur-sangs anglais de course, ce sont les groupes de gènes associés au contrôle de l'utilisation des substrats, aux voies de signalisation de l'insuline, ainsi qu'à la force musculaire, qui semblent être essentiels à la performance (Gu & al., 2009; Tozaki & al., 2010; Hill & al., 2010a, 2010b). Des variants de séquence associés aux caractères de performance ont été détectés dans les gènes *MSTN* (codant la myostatine) (Tozaki & al., 2010; Hill & al., 2010a, 2010b; Binns & al., 2010), *CKM* (codant la créatine kinase), *COX4I2* (codant la sous-unité 4 de l'isoforme 2 de la protéine cytochrome C oxydase) (Gu & al., 2010), et *PDK4* (gène mitochondrial codant l'isoenzyme 4 de la pyruvate déshydrogénase kinase) (Hill & al., 2010c). Parmi ces gènes, le plus étudié reste celui codant la myostatine (*MSTN*, *GDF-8*). Cette protéine est un facteur de croissance sécrété qui inhibe la différenciation musculaire et la croissance lors de la myogenèse, régulant donc négativement la masse musculaire. Une étude transcriptomique comparée sur le muscle avant et après 10 mois d'entraînement a mis en évidence une diminution de l'expression de 58 gènes, dont le gène *MSTN* (McGivney & al., 2010). Les polymorphismes du gène *MSTN* chez les chevaux Quarter Horse ont déjà été évoqués plus haut ; des variations de séquence et de structure de ce même gène *MSTN* ont également été associées à des distances de courses optimales chez les pur-sangs anglais (Tozaki & al., 2010; Hill & al., 2010a; Binns & al., 2010; McGivney & al., 2012). Les variations détectées ici ne concernent pas les régions codantes du gène : tous les *SNP* sont introniques. Une recherche de polymorphisme par séquençage du génome équin et une analyse d'association pangénomique ont montré qu'un polymorphisme C>T dans

l'intron 1 du gène *MSTN* était associé aux aptitudes sportives (Hill & al., 2010a). Ainsi, les chevaux homozygotes pour l'allèle « C » (chevaux C/C) sont plus performants sur des courses rapides et courtes (chevaux sprinters), les hétérozygotes (C/T) seraient d'excellents « milers » (courses de distance intermédiaire) et les homozygotes pour l'allèle « T » (T/T) sont plus endurants (donc performants sur des courses longues), sachant que les distances ont été définies comme : courte pour une distance moyenne de $6,5 \pm 1,5$ furlongs ($1\ 300 \pm 300$ m), intermédiaire pour une distance moyenne de $9,1 \pm 2,3$ furlongs ($1\ 800 \pm 460$ m), et longue pour une distance moyenne de $11,0 \pm 2,1$ furlongs ($2\ 200 \pm 420$ m) (Hill & al., 2010b; Hill & al., 2012).

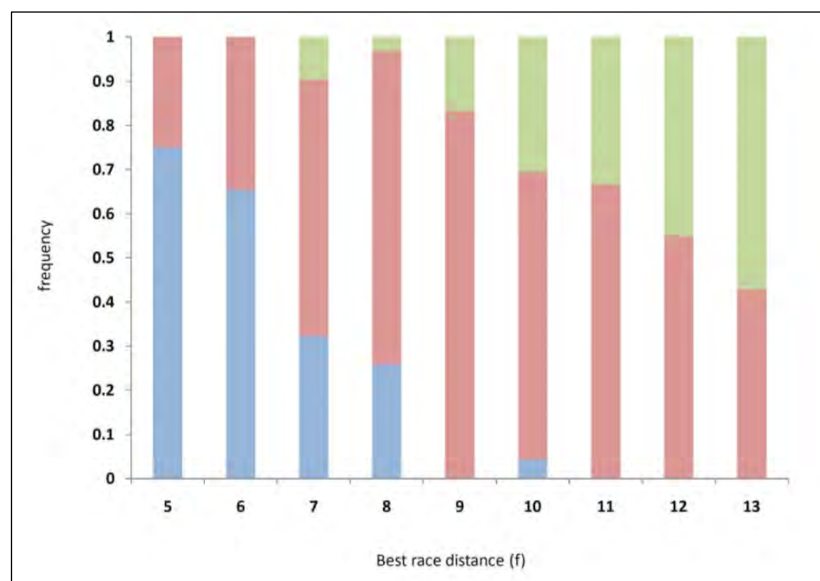


Figure XXX : Histogramme représentant le génotype *MSTN* optimal par rapport à la distance de course (Hill & al., 2010a) : distribution des génotypes C/C (bleu), C/T (rouge) et T/T (vert) parmi les 179 pur-sangs en fonction de la distance de course(s) remportée(s) par ces chevaux (en ordonnées : fréquences génotypiques, en abscisses : longueur de course optimale en furlongs).

Une firme de biotechnologie irlandaise, EquinomeTM, commercialise donc depuis 2010 un test génétique nommé *Speed Gene Test* permettant aux propriétaires et aux entraîneurs de déterminer la distance optimale de course de leurs poulains à partir du génotype de ces derniers au niveau de ce locus intronique, et d'orienter en conséquence leur entraînement (<http://www.equinome.com/>). Ce test donne aussi une idée de l'âge moyen du cheval à sa première victoire : il distingue des chevaux « précoces » (les chevaux C/C), au potentiel maximal à 32 mois, de chevaux « tardifs » (T/T), au potentiel maximal à 40 mois ; les hétérozygotes C/T sont considérés de précocité intermédiaire (potentiel maximal à 35 mois). La précision du test est évaluée à plus de 90%. L'*Equinome Speed Gene Test* a connu un réel succès : plus de 2 500 chevaux avaient déjà été analysés à la fin de sa première année de commercialisation. Un test complémentaire (*Equinome Elite Performance Test*) est proposé depuis fin 2011. Sa première

version (version 1.0) analyse 80 variants génétiques supplémentaires influant sur le potentiel de course, en fonction du génotype myostatine du cheval, c'est-à-dire en fonction de sa distance optimale de course (sprinter/miler/endurant). Sa deuxième version est disponible depuis Juin 2015 et vantée par la société comme « prédicteur génétique de course et de succès reproductif le plus complet et le plus sophistiqué ». Cette version 2.0 prend en compte plus de 70 000 marqueurs « majeurs dans le succès aussi bien à la piste qu'à l'écurie », dont plus d'un millier sont utilisés de manière spécifique pour la performance en course. Elle fournit une valeur génomique d'élevage (*Genomic (Estimated/Enhanced) Breeding Value, G(E)BV*) calquée sur le modèle des animaux de rente, qui représente le potentiel d'une poulinière ou d'un étalon à produire une descendance à la meilleure aptitude possible en course d'élite, estimée à partir du nombre de variants génétiques favorables qu'un individu peut transmettre à sa descendance. Un article de revue présente en détails l'application de ces *GEBV* chez les pur-sangs de course (*Webbon, 2012*). Elle donne également le niveau de consanguinité de l'individu (*Genomic Inbreeding Score*), ce que la version 1.0 ne proposait pas. Les prix affichés par la firme restent très élevés : 590€ pour le *Speed Gene Test* auxquels se rajoutent 940€ pour la version 1.0 ou 1450€ pour la version 2.0 de l'*Elite Performance Test* ! Les tarifs sont cependant réduits au tiers de ces valeurs dans le cadre des Ventes Equinome (*Equinome Sales*). La société a été rachetée récemment (fin 2015) par PlusvitalTM, compagnie de nutrition équine irlandaise.

Nous ne reviendrons pas sur le caractère d'allures qui a déjà été présenté auparavant avec les mutations du gène *DMRT3*.

Enfin, et pour finir sur la caractérisation génomique, des résultats encourageants ont été obtenus pour la discipline majeure qu'est le Concours de Saut d'Obstacles (CSO) à la faveur du projet *JumpSNP* financé par l'IFCE, l'INRA et le Fonds Eperon. Ces données peuvent être utilisées dans un but d'amélioration génétique des populations (indices génomiques), mais aussi individuellement, afin de mieux valoriser les chevaux par une meilleure gestion de leur carrière sportive (*Brard, 2015; Schibler, 2012*).

- Des biomarqueurs pour le suivi de l'état physiologique et le contrôle antidopage (*Schibler, 2012*)

La transcriptomique et la protéomique commencent à se mettre au service de la détection de biomarqueurs reflétant des états physiologiques particuliers chez le cheval, ainsi que du contrôle antidopage.

Les profils d'expression génique obtenus par transcriptomique permettent d'identifier des groupes de gènes présentant des signatures spécifiques d'un état physiologique. Chez le cheval de sport, on s'intéresse donc à l'identification de signatures en relation avec la réponse à l'entraînement, ou la récupération après un exercice physique. Dans ce cadre, une analyse transcriptomique comparée des leucocytes de 10 chevaux d'endurance de haut niveau, avant et après la course, a mis en évidence une réponse inflammatoire aiguë après l'effort (*Capomaccio & al., 2010*), phénomène déjà bien connu dans l'espèce humaine. Chez ces chevaux d'élite, l'inflammation diminue rapidement après 24 heures et non pas plusieurs jours comme c'est habituellement le cas ; reste à déterminer si ce retour rapide à l'équilibre cellulaire est une réponse adaptative à l'entraînement ou une particularité des animaux de haut niveau. L'identification par les « omiques » de biomarqueurs en relation avec la réponse à l'entraînement pourrait s'insérer dans le cadre de l'amélioration de la gestion sportive des chevaux, en limitant les risques d'affections liées au surentraînement ou à la fatigue.

Il est également possible d'identifier des signatures « omiques » de dopage, les molécules dopantes (telles que l'érythropoïétine, E.P.O.) agissant globalement sur les mêmes voies métaboliques que celles empruntées lors d'un effort physique. Le contrôle vise à mettre en évidence non pas la molécule, mais son effet global. Afin de découvrir des biomarqueurs candidats, des analyses ciblées sont effectuées par spectrométrie de masse afin de réaliser un suivi et de détecter indirectement les effets d'un éventuel dopage (*Barton & al., 2009; Teale & al., 2009*). Des signatures d'expression de gènes ou de miARN sont testées (*Bailly-Chouriberry & al., 2010*) dans le but final de coupler test antidopage et contrôle d'identité génétique à partir du même prélèvement, ce qui rendrait les contestations liées à une hypothétique inversion de tubes non recevables.

- La nutriginomique et l'optimisation de la nutrition (*Schibler, 2012*)

La nutriginomique est une science multidisciplinaire étudiant l'impact des nutriments sur l'expression du génome et l'impact des polymorphismes génomiques sur la réponse aux nutriments et sur les besoins nutritionnels (<http://nutrigenomics.ucdavis.edu/?page=information>). Un objectif de cette étude est la médecine personnalisée : en confrontant le génotype aux besoins nutritionnels et à l'état nutritionnel et de santé d'un individu, on espère comprendre les variations de réponse aux nutriments selon les individus, et donc optimiser la performance de chacun en lui fournissant les nutriments les plus adéquats par rapport à son génotype. Certaines sociétés utilisent ce concept chez l'Homme, en proposant d'établir des régimes alimentaires

personnalisés censés prévenir l'apparition de maladies métaboliques selon le profil génomique de chaque individu. Chez le cheval, il s'agirait d'identifier des éléments nutritifs à influence potentielle sur l'expression des gènes en lien avec les performances (sportives, de reproduction...) ou avec la santé. En expliquant en partie la variation de réponse aux nutriments par les différences de génotypes, on espère optimiser l'alimentation de chaque cheval afin de maximiser l'expression de son potentiel génétique tout au long de sa carrière sportive. Cela ouvre également la voie à la commercialisation de nouveaux compléments alimentaires élaborés à cet effet.

L'utilisation simultanée et harmonieuse des sciences omiques s'avère donc être un outil puissant et performant pour la gestion des chevaux : leur élevage, leur carrière sportive, leur santé et leur bien-être... Grâce aux innovations technologiques de plus en plus abordables, les connaissances sont générées par ces sciences en quantité et à une vitesse qui révolutionnent les approches scientifiques classiques, et les applications dans le domaine équin sont en plein essor. Une des conséquences en médecine vétérinaire équine est que les cliniciens ont désormais le devoir d'être conscients des nombreux phénomènes génétiques impliqués dans la santé, la performance et l'élevage des chevaux et d'en tenir compte dans leur pratique quotidienne.

d. Quelques applications (exemples) d'utilisation des données « omiques » dans d'autres espèces que l'espèce équine

Les innovations technologiques dans le domaine des sciences en « omiques » (génomique, épigénomique, transcriptomique, protéomique, métabolomique et métagénomique) au cours des dernières décennies ont permis d'étudier simultanément les variations génétiques, l'expression de milliers de gènes, et la présence de différentes protéines ou métabolites, et ainsi de commencer à mettre en lumière des molécules ou réseaux de molécules agissant en interaction dans le contrôle des fonctions biologiques. Alors que la biologie moléculaire ciblait quelques gènes, marque épigénétiques, ARN ou protéines choisis selon leur fonction et intérêt biologique dans les processus étudiés, le principe des approches en « omique » est d'identifier la totalité des gènes, ARN, protéines ou métabolites, dont les niveaux d'expression ou concentrations sont contrastés entre des échantillons différant pour un ou plusieurs paramètres (par exemple état clinique, régime alimentaire, fond génétique, effort,

stress), sans hypothèse sur les molécules mises en jeu, c'est-à-dire sans *a priori* sur leur fonction biologique, de manière à générer de nouvelles pistes de recherche.

De nombreuses illustrations de l'utilisation de données « omiques » intégrées existent ; il s'agit ici d'en présenter quelques exemples, notamment dans le domaine zootechnique (le domaine sportif sera longuement abordé dans les parties suivantes). Chez les animaux d'élevage, l'intégration de connaissances « omiques » à haut débit peut contribuer à l'émergence de systèmes de productions animales plus durables (sur les plans économique, social et environnemental).

Tout d'abord, les approches « omiques » à haut débit peuvent permettre d'améliorer l'**efficacité économique** des productions de diverses manières.

Une stratégie visant cet objectif est la détection de biomarqueurs de l'exposition aux mycotoxines (contaminants alimentaires dont l'ingestion est à l'origine d'effets néfastes en termes de santé et de performances zootechniques), afin de rendre plus aisé et plus rapide le diagnostic de mycotoxicose dans l'espèce ovine. En détectant les modifications précoces du profil des métabolites urinaires après une exposition à une des mycotoxines majeures (l'ochratoxine A), la métabolomique a permis, via l'utilisation de la spectrométrie de masse couplée à la chromatographie liquide, de mettre en évidence de nombreux marqueurs-candidats, qui sont en cours d'identification (Boudra & al., 2008). Dans le domaine de la production de viande, certaines études visent aussi à mieux caractériser d'un point de vue génétique la différenciation musculaire : l'analyse du transcriptome (Sudre & al., 2003) et du protéome (Chaze & al., 2008) chez les bovins a permis d'identifier des profils d'expression de gènes et de protéines régulés au cours de la myogenèse, et ces derniers ont été mis en relation avec le nombre total de fibres musculaires, les différentes générations de myoblastes associés aux isoformes protéiques qu'ils possèdent, et la masse musculaire (mutations du gène de la myostatine et hypertrophie musculaire, troponine T rapide et prédiction de la masse musculaire (Bouley & al., 2005), etc.). L'analyse du protéome chez le porc a permis d'évaluer l'influence de la race et du type de muscle sur l'expression des protéines musculaires (Mach & al., 2010) ainsi que leurs interactions avec la qualité de la viande (Marcos & al., 2013).

D'autre part, une meilleure connaissance de la régulation de l'expression des gènes par les nutriments joue dans l'amélioration de l'efficacité économique des productions. On peut à ce titre citer certaines études sur la régulation nutritionnelle de l'expression des gènes intervenant dans la lipogenèse mammaire chez la vache et la chèvre (Bernard & al., 2008; Mach &

al., 2011a; Mach & al., 2011b; Mobuchon & al., 2014; Mobuchon & al., 2015), ou d'autres études portant sur les différences de profil génique hépatique en réponse à des variations de la nutrition prepartum chez des vaches laitières (*Loor & al., 2006; Mach & al., 2013*) ou encore sur le différentiel d'expression du gène de la sélénoprotéine W (anti-oxdant) selon le régime diététique des animaux (*Cassar-Malek & al., 2005*).

Enfin, l'efficacité économique des productions peut aussi être améliorée par une meilleure efficacité métabolique et physiologique des herbivores. Des études sont conduites à cet effet sur le partage des nutriments entre tissus et organes, par une méthode de marquage radioactif des nutriments pour réaliser un suivi de leur devenir dans l'organisme ou dans les voies métaboliques, à l'aide de la spectroscopie de masse couplée ou non à la chromatographie liquide (*Hocquette & al., 2008*).

Les approches « omiques » vont également dans le sens des **attentes sociétales** en lien avec le développement durable (pilier social de la durabilité). Le respect du bien-être animal est abordé par la mise en évidence de marqueurs de stress : le stress émotionnel et physique, notamment à l'abattage, est à l'origine chez les bovins de l'expression de gènes et protéines de la famille des Heat Shock Proteins (HSP), dont le lien avec la tendreté de la viande a été mis en évidence par des analyses complémentaires protéome/transcriptome (*Bernard & al., 2007*) ; le stress augmente également la production d'espèces radicalaires réactives à l'oxygène instables et nocives lors de la conduite des animaux à l'abattoir (*Gobert & al., 2009*).

La **protection de l'environnement** est un autre pilier du développement durable dans le cadre duquel s'inscrivent les approches « omiques », par le biais de la maîtrise des rejets azotés par les animaux, le but étant de limiter l'excrétion d'azote dans l'urine et les fèces des animaux tout en maintenant les niveaux de production. Dans cette optique, les recherches sont tournées vers une optimisation de la fonction ruminale et une amélioration de l'efficacité d'utilisation des nutriments azotés par les tissus. Pour cela, les flux métaboliques de nutriments (fluxomique) au niveau du corps entier ou de tissus ou organes spécifiques sont mesurés, et les métabolites urinaires d'animaux présentant une efficacité d'utilisation de l'azote différente sont caractérisés pour trouver des biomarqueurs spécifiques d'animaux présentant une efficacité optimale d'utilisation de l'azote (*Hocquette & al., 2008*).

Pour finir, certaines études « multi-omiques » entrent dans le cadre « recherche de **qualité des produits** », qui pourrait être considéré comme un volet supplémentaire du

développement durable, en s'inscrivant dans une démarche de maîtrise de la composition en acides gras et de la qualité sensorielle des produits laitiers et carnés.

Des approches fluxomique, transcriptomique et protéomique sont utilisées pour mieux comprendre et améliorer les mécanismes de bioconversion et de synthèse des acides gras longs dans les tissus animaux, notamment des AGPI (acides gras polyinsaturés) à longue chaîne de la série des oméga 3 comme l'acide docosohexanoïque (DHA) et des acides gras mineurs comme l'acide linoléique conjugué (CLA), à propriétés potentiellement anticancéreuses ou antiathérogènes chez l'Homme (*De La Torre & al., 2006; Gruffat & al., 2008; Hocquette & al., 2008; Gruffat & al., 2011*).

Les techniques à haut débit sont aussi employées pour élaborer un diagnostic de qualité sensorielle et nutritionnelle du lait, du fromage et de la viande, via l'analyse des vitamines, pigments caroténoïdes, composés phénoliques et terpènes, micronutriments dont la présence dans les produits animaux est particulièrement liée à l'alimentation de ces animaux (*Prache & al., 2007*).

D'autres études utilisant de manière complémentaire protéomique et transcriptomique sont menées dans le cadre de la recherche biologique de prédicteurs de la tendreté de la viande : la surexpression de protéines impliquées dans une vitesse de contraction musculaire lente et un métabolisme oxydatif (*Hocquette & al., 2007*), l'abondance d'enzymes du métabolisme oxydatif comme la succinate déshydrogénase (*Morzel & al., 2008*) et la présence de certaines protéines dont des HSP (*Guillemin & al., 2009; Guillemin & al., 2011*) apparaissent ainsi corrélées à une tendreté supérieure de la viande.

Le développement des approches en « omique », notamment via le programme français AGENAE (Analyse du GENome des Animaux d'Élevage) conçu et conduit en partenariat entre l'INRA et les professionnels des filières animales, a donc permis de mieux connaître les génomes des espèces animales d'élevage et de rendre plus efficaces les recherches menées dans une perspective de développement durable des filières animales d'élevage, grâce aux innovations technologiques à haut débit (séquençage d'ADN, technologies de puces à ADN, approches protéomiques...).

Dans tous les cas, des données de natures multiples (phénomiques, génomiques, etc.) sont intégrées afin de prédire des phénotypes d'intérêts variés : de type zootechnique, médical, sportif... Cette intégration nécessite une innovation biostatistique conséquente, visant à traiter un nombre important de données.

e. Les méthodologies statistiques utilisées pour l'intégration de différentes sources de données « omiques »

Une fois que toutes les données « omiques » sont produites, il est nécessaire de les trier, de les organiser pour leur donner un sens pertinent dans tel ou tel contexte. Pour cela, et notamment dans le cadre de la prédiction phénotypique, on peut utiliser la biologie intégrative, qui fait le lien entre les informations moléculaires (ADN, ARN, protéines, métabolites...) et les informations sur les individus et les populations, analysées de manière bioinformatique, puis les modélise, de manière à appréhender et représenter des phénomènes biologiques complexes et leurs interrelations. Il convient donc ici de présenter et de définir les méthodes statistiques utilisées dans ce contexte.

Encart : La biologie intégrative

Ce concept se situant au cœur de ma thèse, il paraît important de le définir clairement et précisément, et de soulever les différentes notions qu'il implique. Mais au-delà d'être le cadre même de ce projet, c'est un concept en plein essor et qui révolutionne le monde des sciences : ainsi, F. Sigaux, rédacteur en chef de la revue *Hématologie*, affirmait en 2005 que l'on entrait dans l'ère de la biologie intégrative (Sigaux, 2005). Il soulignait donc, il y a une dizaine d'années déjà, les difficultés de traduction de la notion de *systems biology*, et de définition de ce concept traduit en français par les termes « biologie intégrative ». Parmi les objectifs de cette approche, on peut noter l'établissement de liens ou réseaux entre objets de même nature ou de nature différente, la simulation de leur fonctionnement, et l'exploitation de l'information présente globalement dans l'ensemble de la production scientifique. Restreindre la complexité et ordonner la masse d'informations générées nécessite le respect de plusieurs points cruciaux : i) l'utilisation de ces objets impose de pouvoir les désigner de façon univoque, de créer un langage qui puisse se prêter à des représentations symboliques et qui puisse permettre de manipuler les concepts ; ii) l'accumulation de données biologiques de masse ne suffit pas, il faut qu'elles soient ouvertes à tous (accessibles) et standardisées ; iii) des moyens de calcul massifs doivent

être affectés à la démarche, sans pour autant perdre le caractère raisonné et déterministe de cette dernière (Sigaux, 2005).

De nos jours et avec un peu de recul, on peut se tourner vers la présentation qu'en fait l'*Institute for Systems Biology (ISB)*, institut entièrement consacré à ce domaine : la biologie intégrative (*systems biology*) repose sur la compréhension qu'« un tout est plus grand que la somme des éléments qui le composent ». Ainsi, l'*ISB* souligne que ce concept est responsable de certains développements les plus importants dans les sciences de la santé humaine et de la durabilité de l'environnement. Si l'on se tient à la définition que cet institut propose, c'est une approche holistique visant à appréhender la complexité de systèmes biologiques, en commençant par comprendre que les réseaux formant l'ensemble des organismes vivants sont plus que la simple somme de leurs éléments. La biologie intégrative est collaborative, mettant en jeu de nombreuses disciplines scientifiques (biologie, informatique, bioinformatique, ingénierie, physique...) et leurs acteurs respectifs (qui doivent donc travailler ensemble de manière harmonieuse, et communiquer, ce qui est à l'origine d'une expérience sociale enrichissante), afin de prédire comment ces systèmes changent avec le temps et sous des conditions variables, et de développer des solutions répondant aux enjeux (notamment de santé et d'environnement) les plus pressants dans le monde. En effet, les modifications de l'ADN aussi bien d'origine génétique qu'environnementale, les niveaux d'expression de l'ARN, l'expression de gènes, la synthèse protéique, et la présence de certains métabolites, constituent autant de paramètres mesurables désormais, en l'espace de quelques jours, et à un coût de plus en plus abordable. Il revient donc aux scientifiques et aux analystes de données de se servir de cette variété de types de données et de construire des modèles intégratifs permettant une compréhension approfondie du système étudié. C'est justement ce pour quoi les modèles multi-échelles (intégration de modèles à différentes échelles en considérant le flux d'informations entre eux) sont utilisés : ils décrivent un système dans son intégralité et reposent donc sur les bases mêmes de la biologie intégrative. Cette capacité à concevoir des modèles prédictifs, multi-échelles, permet notamment aux scientifiques de découvrir de nouveaux biomarqueurs de maladie, de stratifier les patients sur la base de profils génétiques uniques, et de cibler des médicaments et d'autres traitements. Enfin, la biologie intégrative génère un potentiel d'exploration original et une innovation constante dans le domaine des biotechnologies.

La disponibilité croissante de données -omiques fournit aux chercheurs une vue sans précédent, à grande échelle, des systèmes biologiques. Au cours des dernières années, de nombreux chercheurs ont fait de grands progrès dans l'extraction de connaissances à partir de ces données, en consacrant une grande partie de leurs travaux à l'étude des propriétés et du contenu des ensembles de données -omiques, développant ainsi des méthodes générales pour intégrer différents types de données à l'échelle du génome. Le but est d'incorporer l'information de chacun des ensembles de données afin d'améliorer la prédiction, en appliquant au préalable certaines transformations adaptées à la nature des données.

L'essor récent et le développement considérable de la bioinformatique et de la biostatistique permettent de traiter de très grandes quantités d'information grâce à des approches statistiques comme l'analyse multivariée, regroupant de nombreuses méthodes telles que le partitionnement de données (« data clustering »), l'analyse factorielle, l'analyse en composante principale (A.C.P. ou P.C.A., Principal Component Analysis, ou encore K.L.T., « transformation de Karhunen-Loève »), les forêts d'arbres décisionnels (« random (decision) forests »), l'algorithme CART (Classification And Regression Tree), les Machines à Vecteur de Support (S.V.M., Support Vector Machines) et l'analyse canonique des corrélations (A.C.C. ou C.Cor.A., aussi dénommée analyse des corrélations canoniques).

Le partitionnement (ou regroupement) de données est une méthode statistique d'analyse de données qui consiste à scinder un ensemble de données en différents « clusters » (groupes, grappes) homogènes, c'est-à-dire qui partagent des caractéristiques communes, correspondant en général à des critères de similarité définis en introduisant des mesures et des classes de distance entre objets (*Jain & al., 1999*). On distingue différentes méthodes de regroupement : les méthodes « hiérarchiques » (« hierarchical clustering ») et « de partition » (« partition clustering »). Les trois applications principales du « clustering » sont la segmentation de bases de données, la classification (en sous-groupes, sous-populations au sein de la base de données), et l'extraction de connaissances (pour mettre en évidence des sous-ensembles et sous-concepts non distincts de manière évidente).

Les méthodes d'analyse factorielle et de P.C.A. sont aussi des méthodes de statistique multivariée d'analyse de données, et elles ont toutes deux pour but de réduire la dimension d'un ensemble de données. La première permet d'expliquer la structure de données en utilisant les corrélations entre variables ; elle résume les données en quelques dimensions en condensant un grand nombre de variables en un groupe plus petit de variables ou facteurs. La P.C.A. consiste

à transformer des variables corrélées en un nombre réduit de variables décorrélées les unes des autres (ces dernières sont appelées « composantes principales », ou axes principaux), sans perte importante d'information. Le but est d'expliquer la plus grande variabilité avec le moins de composantes principales. La P.C.A. est à la fois une méthode statistique et graphique ; elle permet d'exprimer les données de manière à mettre en évidence les similarités et les différences, et cela peut être visualisé en représentation graphique. Ces méthodes d'analyse peuvent ainsi permettre de regrouper en une seule variable l'information issue de nombreuses variables initiales décrivant le résultat de quantités observées mesurables. Concrètement, voici comment cela s'applique pour évaluer la complexité du phénomène : des milliers de mesures phénotypiques sont répertoriées pour chaque individu mais un nombre plus faible de variables pourrait expliquer la plupart des différences inter-individuelles. Les efforts de phénotypage peuvent donc se concentrer sur l'identification des mesures qui contribuent le plus à définir les variables cachées sous-jacentes, et ces dernières pourraient être corrélées à des données génotypiques mieux que n'importe quelle mesure phénotypique unique.

Un exemple d'utilisation du « hierarchical clustering » et de la P.C.A. se trouve dans la partie expérimentale de ma thèse, qui sera présentée par la suite : en effet, nous avons utilisé ces méthodes (entre autres) afin de mettre en évidence l'effet de l'exercice physique sur le miRNome. Sur la figure ci-dessous apparaissent clairement deux groupes (« clusters ») distincts (à savoir : *T0* : miARN sanguins exprimés chez les chevaux avant l'effort d'endurance, et *T1* : miARN sanguins exprimés chez les chevaux après l'effort d'endurance).

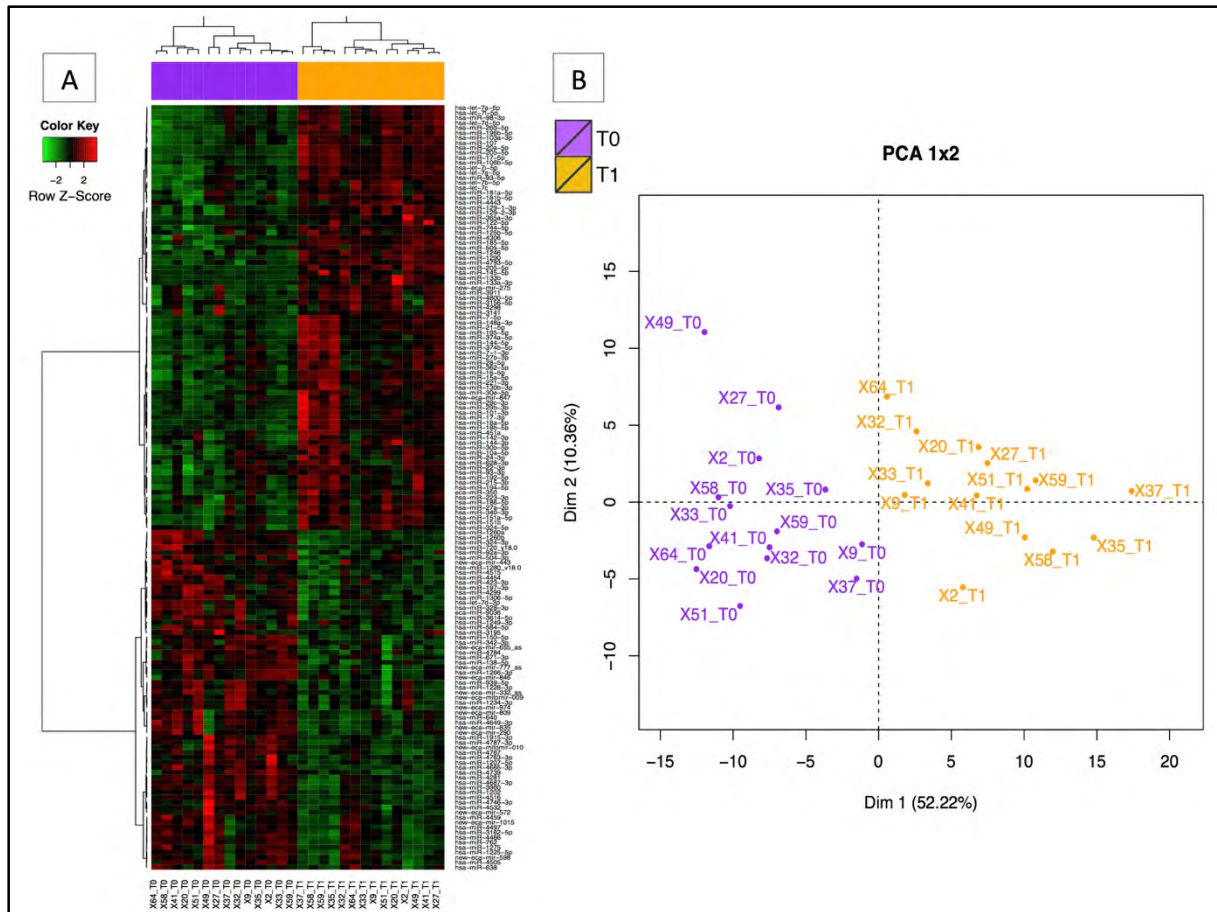


Figure XXXI : Différence de profil d'expression des miARN dans le sang en fonction de l'exercice

(A) Représentation « heat-map », c'est-à-dire en cluster hiérarchisé, des 167 miARN différentiellement exprimés en comparant T0 (avant la course d'endurance) et T1 (après la course d'endurance) ($FDR < 0.05$). Une représentation colorée de l'expression des miARN montre en rouge les miARN surexprimés et en vert les miARN sous-exprimés.

(B) P.C.A. des miARN différentiellement exprimés dans le sang en comparant T1 à T0 ($FDR < 0.05$). Le premier axe (vertical) représente 52.60% de la variation, et les deux axes représentent 62.58% de l'inertie totale (la notion d'inertie généralise celle de la variance en dimension quelconque, la variance étant relative à une seule variable).

L'algorithme CART (Classification And Regression Tree) permet de construire un arbre de décision binaire en classant un ensemble de données ou d'individus décrits par des variables qualitatives et quantitatives ; cet arbre décisionnel produit des classes les plus homogènes possibles et fournit un modèle pour classer de nouveaux échantillons. Les forêts d'arbres décisionnels (ou forêts aléatoires) sont des méthodes de classification et de régression par « apprentissage » (reconnaissance d'une règle de classement à partir d'un ensemble de données) reposant sur des algorithmes qui fonctionnent en construisant une multitude d'arbres de décision et en faisant ressortir la classe qu'est le mode de sortie des classes obtenues par les arbres individuels ; plus le nombre d'arbres de décision augmente, plus la précision de ce « classifieur » (prédicteur) augmente et plus l'erreur est faible (Ho, 1995; Breiman, 2001; Strickland, 2014).

Les Machines à Vecteur de Support sont des générateurs de modèles associés à des algorithmes analysant des données et reconnaissant des schémas, permettant une classification binaire de ces données ; ils appartiennent aux méthodes de classification supervisée, dont l'idée est de générer une règle de classement à partir d'un ensemble de données dont le classement est déjà connu, puis de l'appliquer pour catégoriser de nouvelles données, dont le classement est inconnu (Cristianini & Shawe-Taylor, 2000). Les S.V.M. sont particulièrement performantes dans le cadre du traitement de données de grande dimension (comme l'expression de gènes) et structurées (comme les séquences d'ADN). Concrètement, le but de cette méthode est de trouver un classifieur linéaire (« hyperplan ») séparant les données et maximisant la distance entre les deux classes obtenues ; les points les plus proches sont utilisés pour la détermination de l'hyperplan et appelés « vecteurs de support ».

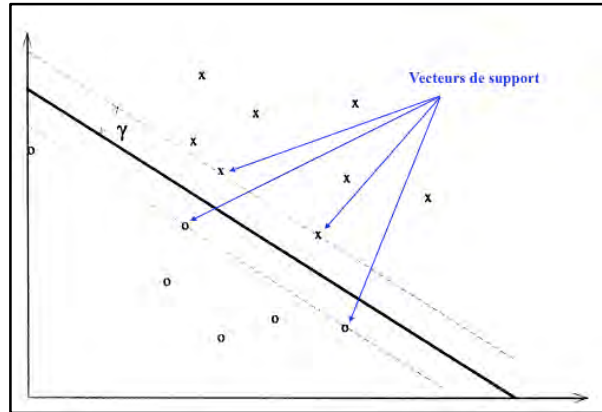


Figure XXXII : Principe de classification graphique par une méthode utilisant les machines à vecteur de support

Enfin, l'A.C.C. permet d'explorer les relations entre différents tableaux de données mesurées sur les mêmes individus, donc de comparer des groupes de variables quantitatives appliqués sur les mêmes individus. Il s'agit ici d'extraire des combinaisons linéaires de variables canoniques de chaque table telles que ces dernières maximisent la corrélation entre les jeux de données et la covariance au sein de chaque jeu de données. Le but de l'analyse canonique est de comparer les groupes de variables afin de savoir s'ils décrivent un même phénomène, auquel cas on pourra se passer d'un de ces groupes. Dans le cadre de données de grande dimension (ou de variables fortement corrélées), l'utilisation de méthodes de régularisation est nécessaire. L'A.C.C. peut être utilisée à la fois pour explorer les relations entre les différents types de variables omiques et méta-omiques, mais également pour extraire un faible nombre de méta-variables (réduction de dimension) à inclure dans un modèle de prédiction. Ci-dessous est présenté un exemple d'utilisation de la méthode d'A.C.C. pour mettre en évidence des corrélations entre la production d'immunoglobuline (immunoglobuline A luminale (sécrétée), sIgA) ou le poids vif (*body weight*, BW) chez les porcelets et la présence de certains genres bactériens du microbiote fécal porcin (Mach & al., 2015).

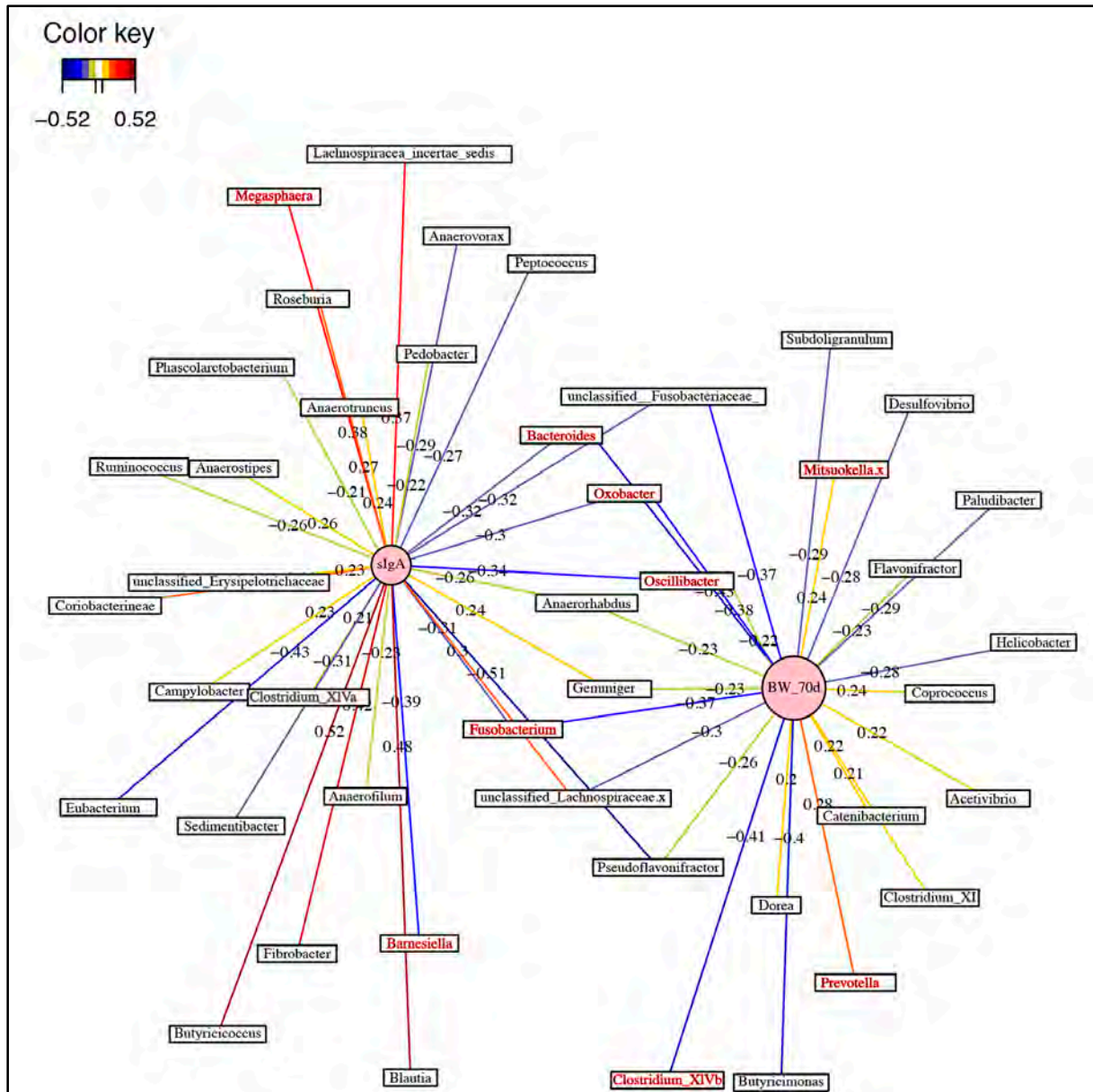


Figure XXXIII : Détermination de genres bactériens potentiellement clés pour le poids vif et la production d'*sIgA* chez les porcelets (Mach & al., 2015)

Covariances entre les abondances relatives des bactéries du microbiote fécal au jour 36 et différents phénotypes au jour 70 établies en utilisant une régression PLS (Partial Least Squares) parcimonieuse. Les genres bactériens les plus abondants ont été inclus dans l'analyse. Le réseau est représenté graphiquement par des nœuds (genres et phénotypes) et des liens (relation biologique entre nœuds). L'intensité de couleur des liens indique le niveau d'association : rouge = positif, bleu = négatif. La forme des nœuds indique s'il s'agit d'un phénotype (rond) ou d'un genre (rectangle). Le degré d'association est indiqué sous chaque lien. Les associations par paires avec des valeurs de plus de 0.20 sont représentées.

A ce stade, on dispose des données ainsi que des outils et méthodes nécessaires à leur traitement, et on connaît leurs applications principales dans différents domaines. Il s'agit donc maintenant de faire le lien entre ces -omiques et le sport d'endurance.

2. La prédiction phénotypique appliquée aux performances sportives d'endurance

Avant de s'intéresser au cheval, on peut se demander ce qui a été étudié et réalisé chez les athlètes humains. En effet, des moyens plus importants sont souvent mis en œuvre en recherche humaine, et certains résultats intéressants obtenus chez l'Homme peuvent potentiellement être transposés chez l'animal, notamment chez le cheval. Par exemple, la découverte de biomarqueurs de l'effort intense et/ou prolongé chez l'Homme peut motiver des études sur ces marqueurs chez le cheval dans le cadre de performances sportives de haut niveau.

a. Chez l'Homme : à la recherche de biomarqueurs de l'effort d'endurance

Comme évoqué dans l'introduction, le sport est à l'origine de perturbations physiologiques, biochimiques, hormonales et métaboliques, notamment d'une réponse de stress. Ces modifications peuvent même engendrer, à long terme, des ajustements micro et macroscopiques, repérables chez les athlètes de haut niveau (tels que des fréquences cardiaques remarquablement basses au repos). On rappellera donc les perturbations entraînées par ce paramètre extérieur qu'est l'exercice physique intense et prolongé de type « endurance ».

Avant de prétendre faire de la « prédiction » en intégrant diverses données -omiques dans le cadre des performances sportives, il est important de connaître ces données dans ce contexte : les études de phénotypique, génomique, transcriptomique, protéomique, métabolomique en lien avec l'exercice physique intense et la performance sportive (notamment d'endurance) sont nombreuses, mais il convient d'en dégager quelques grandes lignes, qui aident à comprendre le contexte de l'étude expérimentale qui suit. Il s'agit de considérer ces informations dans une optique de recherche de biomarqueurs : de maintes études proposent ainsi des biomarqueurs de l'effort, qu'il s'agisse de biomarqueurs physiologiques (capacités telles que la $V_{O_{2max}}$), moléculaires (paramètres de souffrance cardiaque ou musculaire par exemple) ou d'un autre type. La métabolomique est une méthode particulièrement efficace pour révéler ces biomarqueurs et voies biologiques induites par une telle intervention sur un système biologique.

Enfin, l'application des sciences –omiques dans le domaine de la performance sportive est même à l'origine d'un nouveau concept, la « sportomique » (*sportomics*), que nous présenterons brièvement.

i. Les effets d'un exercice physique intense de type « endurance »

Nombreuses sont les études qui s'accordent à dire que l'activité physique modérée est bénéfique pour la santé des individus, aussi bien sur le plan cardio-vasculaire, que pulmonaire, musculaire, métabolique, nerveux... Tout d'abord, elle exerce un effet protecteur sur les fonctions pulmonaire et cardio-vasculaire (*Shiroma, Lee, 2010*). De plus, elle intervient favorablement dans la neurogenèse adulte et contribue à améliorer les fonctions cognitives (*Lafenetre & al., 2011*). Elle joue en outre sur l'état mental et émotionnel et permet ainsi de lutter contre l'anxiété et la dépression dans divers contextes (*Yeung, 1996; Anderson, Shivakumar, 2013; Josefsson & al., 2014; Stanton, Reaburn, 2014; Levin, 2014*), notamment via l'activation de l'axe hypothalamo-hypophyso-surrénalien. Elle participe également à la prévention de certaines maladies métaboliques liées au mode de vie (*Sato & al., 2003*) et de certains cancers (*World Cancer Research Fund / American Institute for Cancer Research, 2007*). Elle améliore aussi la survie à certains cancers et la qualité de vie de patients ayant guéri d'un cancer (*West-Wright & al., 2009; Speck & al., 2010*). Rappelons qu'elle développe bien évidemment les capacités musculaires, et qu'elle possède d'autres effets bénéfiques analysés dans une pléthore d'études plus ou moins récentes. Pour toutes ces raisons, le jogging et la course à basse intensité sont devenus des activités sportives en plein essor chez des individus de tous genres, de tous âges, partout dans le monde, en tant qu'atout santé ou activité de loisir.

Cependant, la forme extrême de ce type d'activité, à savoir les sports d'endurance tels que le marathon ou l'ultra-marathon (caractérisé par des distances supérieures à celles du marathon), constitue un exercice intense, éprouvant et prolongé, qui peut nuire à la santé ou, du moins, générer des effets moins désirables. Ainsi, de multiples études soulèvent les dangers potentiels liés à un exercice physique de type « marathon ». Un tel effort induit des changements marqués aux niveaux métabolique et hormonal (*Hartley & al., 1972; Ahlborg & al., 1974*), plaçant de nombreux tissus et organes dans des conditions de stress métabolique, surtout si les besoins énergétiques ne sont pas remplis ; le contrôle de ce stress métabolique inclut des réponses hormonales et immunitaires importantes (*Coyle, 2000*). Concernant le métabolisme énergétique, un effort physique intense et prolongé entraîne un stress oxydatif, une augmentation de la production de chaleur, une activation de médiateurs de l'inflammation, des dysfonctions

cellulaires liées à l'apport d'ATP devenu insuffisant pour subvenir aux fonctions cellulaires de base comme l'alimentation des pompes à ions et le maintien du volume cellulaire (Branth & al., 2009). Un effort d'endurance tel que le marathon est responsable d'une déplétion sans équivoque de glycogène (liée à une utilisation prononcée du glycogène musculaire, surtout dans les fibres de type I) et de triglycérides (liée à une lipolyse considérablement accélérée) dans les muscles squelettiques (Jansson, Kaijser, 1987; Branth & al., 2009) et d'une augmentation de la concentration plasmatique en taurine (acide aminé soufré) reflétant les pertes intracellulaires des myocytes lors de fortes demandes énergétiques (Ward & al., 1999; Cuisinier & al., 2001). Une étude (Branth & al., 2009) a ainsi mis en évidence la condition de stress métabolique induite par l'exercice éprouvant chez les athlètes : une période prolongée d'effort intense continu à dépense énergétique très élevée est à l'origine d'une production accrue de dérivés réactifs de l'oxygène (*Reactive Oxygen Species, ROS*, dont le malondialdéhyde (ou aldéhyde malonique, MDA)), témoins de stress oxydant, accompagnée de signes de résistance à l'insuline et d'une disponibilité réduite en glucose intracellulaire, entraînant une alimentation compromise des pompes à ions responsable d'une perturbation de l'osmorégulation cellulaire, visible par l'efflux de taurine et la turgescence cellulaire des myocytes des jambes des individus ayant effectué cet exercice. A plus grande échelle, les dangers inhérents à la pratique d'un exercice éprouvant prolongé de manière répétée de type « marathon » rapportés par des études touchent aussi bien : i) le plan cardio-vasculaire, avec une augmentation de rigidité des parois des grosses artères et des anomalies ventriculaires gauche et droite (O'Keefe & al., 2012) ainsi qu'un risque de fibrillation (Mascia & al., 2012) ; ii) les plans musculaire (Suzuki & al., 2006; Bessa & al., 2008) et cartilagineux (Kim & al., 2009; Vuolteenaho & al., 2014), notamment chez les athlètes d'élite qui réalisent leurs activités avec des impacts forts et qui exercent donc des contraintes très importantes sur leurs articulations, à l'origine d'un risque accru de traumatismes articulaires et d'ostéoarthrose ; iii) ou encore le plan immunitaire, avec une réaction inflammatoire systémique (Tidball, 2004; Toumi & al., 2006; Butterfield & al., 2006; Wallberg & al., 2011)... Concernant le système immunitaire, des facteurs de risque de sensibilité accrue aux infections liés à un effort physique important d'endurance ont été mis en évidence au niveau transcriptionnel. Ainsi, une étude transcriptomique a établi des profils d'expression génétique des cytokines de type Th1 et Th2 et des gènes de facteurs de transcription liés, dans des cellules mononucléaires du sang périphérique (*Peripheral Blood Mononuclear Cells, PBMC*) d'individus ayant participé à un marathon (Xiang & al., 2014). Elle a permis de mettre en évidence un déséquilibre immunitaire prolongé (déplacement de l'équilibre réponse immunitaire de type Th1/réponse immunitaire de type Th2 vers un profil Th2 caractéristique dans les *PBMC*, persistant au minimum une semaine

après le marathon) et une affection de l'expression génétique des facteurs de transcription associés à la production de cytokines. Ce déséquilibre est responsable d'un risque augmenté d'infections respiratoires supérieures après ce genre d'exercice physique éprouvant et prolongé.

Bien que de nombreux paramètres, notamment métaboliques (tels que l'alcalose respiratoire liée à l'hypocapnie hyperventilatoire, la souffrance musculaire reflétée par l'augmentation de concentration des enzymes musculaires, la souffrance du myocarde visible par la libération excessive de certains biomarqueurs de dysfonction cardiaque, le déplacement net du profil lipidique sanguin lié à l'augmentation d'oxydation des acides gras, etc.), reviennent à des valeurs physiologiques pendant les 48 heures de récupération suite à une course de type ultra-marathon (Klapcińska & al., 2013), la répétition de tels efforts chez les athlètes est susceptible de générer un syndrome de surentrainement (Overtraining Syndrome, *OTS*) caractérisé par une baisse de performance, des signes de fatigue allant jusqu'à l'épuisement, et d'autres symptômes (Purvis & al., 2010).

ii. Les biomarqueurs de l'effort d'endurance, ou quand les – omiques se mettent au service (de la prédiction) des performances sportives d'endurance

Toutes les sciences en « omique » peuvent intervenir dans l'étude des effets d'un exercice intense et prolongé de type marathon, et la combinaison de ces données de différentes natures peut donner soit une vision globale, soit une perspective très ciblée, de certains paramètres « clés » de ce type d'effort.

Si l'on considère les paramètres physiologiques comme autant de phénotypes, alors on peut envisager les trois paramètres principaux apparaissant comme cruciaux en performance d'endurance, que M.J. Joyner et E.F. Coyle définissent dans leur article : la $V_{O_{2,max}}$ (consommation maximale en dioxygène), le seuil anaérobie (*lactate threshold*, seuil à partir duquel le métabolisme aérobie ne suffit plus à couvrir les besoins nécessaires à l'activité musculaire), et l'efficacité (coût en dioxygène pour générer une vitesse de course donnée ou une puissance de pédalage par exemple) (Joyner, Coyle, 2008). Ces trois marqueurs physiologiques sont utilisés comme prédicteurs de performance d'élite. Il s'agit donc d'établir une association entre ces paramètres et des marqueurs –omiques particuliers chez l'Homme, en gardant à l'esprit que les facteurs motivationnels et sociologiques, ainsi que les aspects cultureux et environnementaux ont des rôles déterminants dans le fait de devenir un champion ou non, ce

qui complexifie la recherche de facteurs génétiques pouvant contribuer au succès dans le domaine sportif chez l'Homme.

Avec des variations de $V_{O_2,max}$ en réponse à l'entraînement allant de 0 à plus de 100%, l'implication de facteurs génomiques sous-jacents a été proposée. L'amélioration de la capacité à l'exercice aérobie en réponse à un entraînement d'endurance a ainsi vu son héritabilité estimée à 45-50% (Bouchard & al., 2011). Quelques approches de pangénomique et des analyses transcriptomiques ont été utilisées pour caractériser des biomarqueurs génomiques de la capacité à l'exercice aérobie. La recherche de gènes et marqueurs candidats pertinents pouvant contribuer à la prédiction de performance chez des individus sédentaires ou actifs, en adaptation à de l'exercice intense, ou pour des modifications induites par l'entraînement, a ainsi abouti à l'établissement d'une carte génétique humaine de phénotypes de performance et de forme (*fitness*) liée à la santé (*The human gene map for performance and health-related fitness phenotypes*) (Rankinen & al., 2006) : la dernière version de cette carte, datant de l'année 2005, liste 165 gènes et *QTL* autosomiques, 5 marqueurs sur le chromosome X, ainsi que 17 gènes mitochondriaux, associés à des phénotypes pertinents de forme et de performance. Certains d'entre eux ont été caractérisés comme prédicteurs génomiques de la capacité aérobie maximale en utilisant différentes méthodes (Keller & al., 2011; Bouchard & al., 2011; Thomaes & al., 2011).

Une étude combinant approches transcriptomique et phénotypique évalue de manière concomitante les performances cardiaques, les biomarqueurs cardio-vasculaires de l'effort, et l'expression génétique dans les *PMBC* chez des sujets participant à un entraînement de semi-marathon (Vance & al., 2014), le but étant d'analyser les effets cardio-vasculaires de ce type d'activité sportive et les changements d'expression génétique correspondants ; autrement dit, d'identifier une base génomique pour la réponse à l'exercice d'endurance, en corrélant des modifications spécifiques dans l'expression génétique avec des changements physiologiques suivant un entraînement au marathon. Les résultats concernant l'augmentation de performance cardiaque et hémodynamique par l'entraînement intense ne sont certes pas surprenants, mais cette étude prospective apporte une approche novatrice en y associant des modifications significatives de l'expression de gènes impliqués dans la réponse immunitaire et inflammatoire. Autre élément intéressant ressortant de cette étude : la subdivision du groupe d'individus entraînés en un groupe de « forts répondeurs » et un groupe de « faibles répondeurs » sur la base du pourcentage de variation de la $V_{O_2,max}$ met en évidence des loci permettant de différencier ces deux groupes au repos et après l'entraînement.

La recherche de biomarqueurs se poursuit avec des analyses microtranscriptomiques. L'une d'elle a permis l'identification de trois microARN plasmatiques (miR-1, mir-133a et mir-206) spécifiquement corrélés aux paramètres de performance aérobie que sont la $V_{O_2,max}$ et la vitesse de course au seuil anaérobie (Mooren & al., 2014), dans le cadre d'une course de marathon. Cette étude place ces trois myomiRs particuliers (miARN spécifiques des muscles squelettiques et cardiaque, dont les niveaux d'expression sont modifiés en réponse à l'exercice) comme potentiels biomarqueurs de capacité aérobie (bien que leur expression ne soit pas corrélée à des marqueurs conventionnels de souffrance cardiaque tels que la troponine T, la troponine I ou le peptide natriurétique de type B (*Brain Natriuretic Peptide, BNP*)). D'autres études traitent également du rôle de ces trois myomiRs ainsi que d'autres miARN dans la biogénèse mitochondriale et les métabolismes glucidique et lipidique dans les muscles squelettiques et cardiaque à la suite d'un effort d'endurance, soulevant leur caractère de potentiels biomarqueurs de ce type d'exercice physique (Baggish & al., 2011; Russell & al., 2013).

Enfin, la métabolomique étant une stratégie établie d'exploration des effets de stimuli variés sur l'empreinte métabolique des biofluides, elle se révèle particulièrement efficace pour mettre en avant des biomarqueurs et voies métaboliques induites par une certaine intervention sur un système biologique, comme l'exercice physique par exemple (Pechlivanis & al., 2013). Ainsi, A. Pechlivanis et al. ont réalisé une étude métabolomique en RMN 1H sur l'impact à court et long terme de deux programmes d'entraînement à la course de sprint (caractérisés par 3 sessions par semaine sur un total de 18 semaines, avec au sein de chaque session, des intervalles de récupération différents entre sprints selon le protocole d'entraînement : soit 10 secondes, soit une minute) sur l'empreinte métabolique du sérum humain, et ont montré que l'on pouvait discriminer des échantillons sur la base de l'exercice réalisé (Pechlivanis & al., 2013) : ils ont constaté une modification des niveaux de 18 métabolites (dont 17 identifiés) sur 33 contribuant à la distinction entre l'état avant effort et l'état post-effort, et/ou entre l'état entraîné (dernière session d'exercice) et l'état non entraîné (1^{ère} session). La classification selon l'exercice (distinction avant/après effort), établie par un modèle validé, repose principalement sur le lactate, et dans une moindre mesure sur le pyruvate, l'alanine, les trois acides aminés branchés (leucine, valine, isoleucine), l'arginine/lysine, les acétyles de glycoprotéines, et un métabolite non identifié ; la classification selon l'entraînement (distinction 1^{ère} session/dernière session), quant à elle, repose sur le lactate, le pyruvate, la méthylguanidine (biomarqueur de stress oxydant), le citrate, le glucose, la valine, la taurine, le N-oxyde de triméthylamine, les

composants à choline, les histidines, l'acétoacétate/acétone, les acétyles de glycoprotéines, et les lipides. Cependant, les échantillons ne pouvaient pas être classés selon l'intervalle de repos entre sprints, c'est-à-dire s'ils avaient réalisé le 1^{er} ou le 2^e protocole d'entraînement. Une autre étude de métabolomique en RMN a été réalisée afin d'évaluer les effets de l'exercice, en utilisant des biofluides murins pour modéliser les biofluides humains (les métabolites détectés dans l'urine murine étant très similaires à ceux détectés dans l'urine humaine, bien que les modulations de leurs niveaux par l'exercice soient différentes selon l'espèce), et a démontré que : i) les profils métabolomiques RMN étaient très dépendants de l'exercice réalisé, avec comme métabolites discriminants : l'allantoïne (produit du métabolisme oxydatif), l'inosine (remplacée par l'hypoxanthine, plus sensible à l'exercice que l'inosine, chez l'Homme) et les acides aminés branchés ; ii) les échantillons d'urine étaient plus informatifs que les prélèvements sanguins ; iii) la durée de l'exercice était un paramètre qui influait plus sur l'état métabolique que la vitesse pendant l'exercice (*Le Moyec & al., 2012*). Ce dernier point est particulièrement intéressant dans le cadre de l'exercice d'endurance, qui est caractérisé par un effort de longue durée. De plus, les métabolites discriminants représentent de potentiels biomarqueurs : des études plus ciblées seraient nécessaires pour tirer des conclusions plus poussées.

De nombreuses études mettent l'accent sur une cytokine particulière, l'interleukine-6 : l'IL-6 est proposée comme myokine (donc biomarqueur de l'effort), c'est-à-dire cytokine produite et libérée par les muscles squelettiques exerçant des effets sur d'autres parties du corps (*Pedersen, Hoffman-Goetz, 2000; Pedersen & al., 2001; Febbraio, Pedersen, 2002; Pedersen & al., 2003; Pedersen & al., 2004; Petersen, Pedersen, 2006; Pedersen, Fischer, 2007; Pedersen, Febbraio, 2008; Pedersen, 2009; Walsh & al., 2011; Klapcińska & al., 2013*). Le gène de ce « facteur d'exercice » (plutôt facteur métabolique qu'immunitaire dans le muscle à l'effort (*Walsh & al., 2011; Klapcińska & al., 2013*)) est activé par la contraction musculaire (sa transcription est induite par l'exercice), et l'IL-6 fonctionne comme détecteur d'énergie glycogène-dépendant dans le muscle. Elle est libérée en grandes quantités dans le sang par les muscles en contraction, agit sur le tissu adipeux en induisant la lipolyse et la transcription de gènes dans la graisse abdominale sous-cutanée, et présente des effets anti-inflammatoires puissants. L'IL-6 aurait ainsi un rôle dans la médiation d'effets bénéfiques de l'exercice sur les troubles liés à l'inactivité et à l'obésité (*Pedersen & al., 2004*). Dans le cadre d'un exercice d'ultra-endurance, l'augmentation de concentration plasmatique en IL-6 serait plus liée à l'intensité de l'effort qu'à la durée de ce dernier (*Wallberg & al., 2011*). Il serait intéressant d'aller plus loin dans les investigations sur l'IL-6 liées à l'effort en utilisant les sciences -omiques intégrées.

iii. Un nouveau concept : la sportomique

L'intégration de diverses données –omiques dans le domaine du sport est à l'origine de la création d'un concept novateur : la sportomique (*sportomics*). Il ne s'agit pas forcément de prédiction ici, mais plutôt de l'utilisation de l'exercice comme outil de choix afin d'étudier les interactions entre stress métabolique et système immunitaire (*Bassini-Cameron & al., 2007; Bessa & al., 2008; Gonçalves & al., 2012*).

Un article présente ce nouveau concept comme recherche sans *a priori* sur les modifications métaboliques touchant un individu pendant le sport et l'exercice : il s'agit d'une approche similaire à la métabolomique et aux autres concepts en –omique, mais la sportomique a ceci de différent qu'elle se concentre sur le sport comme challenge métabolique (*Bassini, Cameron, 2014*).

Pendant plus d'une décennie, des approches alternatives ont été utilisées pour comprendre les réponses métaboliques à un stress physique : on a utilisé des protocoles expérimentaux et des régimes particuliers pour induire des changements de métabolisme et étudier ces modifications. En plus des études classiques de laboratoire (modèles cellulaires et animaux), des athlètes d'élite dans divers sports ont été mis à contribution pour analyser le stress métabolique. L'idée de sportomique est née d'une étude de protéolyse dans un neurone, ayant conduit les chercheurs à étudier l'hyperammoniémie et à rechercher un modèle satisfaisant d'analyse des effets de l'ammoniémie sur le système nerveux central. La sportomique s'ancre donc dans un effort de compréhension du métabolisme cellulaire comme système intégré, et des relations entre nombreuses cellules et espèces moléculaires quantifiées. Comme les autres approches en –omique, il s'agit d'un concept à haut-débit, cette discipline est holistique et la démarche est descendante (*top-down*), avec un traitement de données systématique associé à la génération d'une vaste base de données, accessible et consultable ; le suffixe « ome » indique là aussi que tous les constituants sont considérés ensemble dans un état donné.

Les auteurs avancent que les analyses métaboliques de terrain sont essentielles à la compréhension, au soutien et à l'entraînement des athlètes d'élite, et placent leurs espoirs en cette approche afin de combler la lacune entre biologie intégrative et médecine traditionnelle.

Cela dit, la sportomique se heurte à quelques limites : de nombreuses études sur l'exercice physique ont été réalisées sur des sportifs amateurs, de loisir, ou simplement des personnes actives (et non pas de véritables athlètes), et peu de données sont disponibles sur les athlètes s'illustrant en compétition à un niveau national ou international (elles sont effectivement souvent gardées secrètes, dans la mesure où les compétiteurs pourraient s'en servir). A. Bassini et L.C. Cameron soulignent donc l'intérêt pour la communauté scientifique d'avoir accès aux informations métaboliques des athlètes de haut niveau et de les utiliser dans la compréhension du métabolisme de ces individus via la méthodologie à la pointe dont on dispose désormais.

Forts de ces informations issues d'études menées sur l'Homme, nous pouvons désormais nous tourner vers l'apport des sciences –omiques dans le contexte sportif de l'endurance équine.

b. Chez l'animal athlète : exemple du cheval

Depuis une dizaine d'années maintenant, la génomique équine a bénéficié d'un saut technologique inédit en matière de séquençage et de génotypage, notamment avec le séquençage complet du génome équin en 2009 (*Wade & al., 2009*), à l'origine d'applications et d'idées novatrices, comme présentées plus haut dans le chapitre sur les –omiques équins.

Les données –omiques sont tout particulièrement mises au service de l'étude de phénotypes sportifs équins variés, en lien avec des disciplines tout aussi variées (saut d'obstacles, courses de vitesse, endurance...), souvent dans une optique d'optimisation de l'entraînement des chevaux. Il s'agit, entre autres, de choisir le cheval le plus susceptible de se révéler performant dans telle ou telle discipline (par sa morphologie, etc.), de prévoir son entraînement de manière harmonieuse avec son rythme biologique afin de maximiser son potentiel de performances, de lui fournir une alimentation adaptée à la fois à ses capacités de digestion (gouvernées en partie par ses propres informations –omiques) et à son mode de vie...

Afin de prédire le niveau de performance que l'on peut espérer atteindre chez un jeune cheval, on peut mesurer des caractéristiques physiologiques dans un contexte sportif (types de fibres musculaires, fréquence cardiaque, lactatémie, allures, style de saut, conformation...) et on calcule l'héritabilité de chaque paramètre (*Barrey, 2010*), c'est-à-dire la part de variation

observée au niveau d'un caractère phénotypique (comme des performances mesurées par exemple) due à des différences de valeur génétique additive entre les individus d'une population (autrement dit, l'héritabilité correspond à la variance des valeurs génétiques additives divisée par la variance phénotypique). La valeur génétique additive représente la partie transmissible (pour moitié, en espérance) de la valeur d'un génome. La notion d'héritabilité est utilisée en sélection classique (ou génomique) et dans les modèles de prédiction chez les chevaux de sport. Elle est un indicateur de la facilité avec laquelle un caractère pourra être sélectionné (ou du coût qu'il faudra consentir pour sélectionner un caractère avec une efficacité donnée) au sein d'une population. Ainsi, des héritabilités ont été estimées pour les performances dans différentes disciplines : la course de galop ($h^2 = 0,15$ à $0,55$), la course de trot ou d'amble ($h^2 = 0,17$ à $0,26$), le saut d'obstacles ($h^2 = 0,05$ à $0,28$), le dressage ($h^2 = 0,11$) (*Hintz, 1980; Langlois, 1980; Ricard & al., 2000*)... Les estimations d'héritabilité sont bien sûr race-dépendantes. La thèse de Sophie Brard présente de manière exhaustive les principes, méthodes et applications de la sélection génomique des chevaux athlètes de sport et de course (*Brard, 2015*).

En ce qui concerne la gestion de l'entraînement par rapport à la physiologie équine, l'objectif est de déterminer les meilleures conditions, le contexte le plus propice à l'entraînement du cheval : des études ont par exemple analysé à cet effet la régulation circadienne de l'activité locomotrice et de l'expression génétique dans le muscle squelettique, et ont démontré que certains moments de la journée étaient plus propices à une forte performance que d'autres, en raison des variations d'expression génétique dans le muscle squelettique au cours de la journée (*Martin & al., 2010*).

L'alimentation et le concept de « nutrigenomique » ayant déjà été abordés dans l'encart sur les -omiques équins, nous ne reviendrons pas dessus.

Le phénotype sportif particulier qui nous intéresse dans cette thèse est celui associé à l'endurance équine. Un projet spécifiquement orienté sur ce thème a vu le jour il y a quelques années, et sa finalisation récente est à l'origine de découvertes intéressantes et d'études toujours en cours : il s'agit du projet GenEndurance. Après en avoir présenté les grandes lignes, et en gardant en tête les études réalisées sur le phénotype sportif associé à des efforts prolongés intenses chez l'Homme, nous examinerons plus en détails les résultats des différents travaux -omiques appliqués à ce phénotype chez le cheval, et terminerons cette partie par une présentation schématique expliquant le principe d'analyse intégrée utilisée dans la partie expérimentale de la thèse qui suivra.



- i. La détermination d'un phénotype sportif associé à l'endurance équine : les premiers pas du projet GenEndurance (2011-2014 ; C. Robert, E. Barrey)

- Présentation des objectifs et synthèse du projet

Le projet GenEndurance (<http://genendurance.over-blog.com/>) met en lumière l'utilisation de données phénotypiques et génomiques dans l'objectif de prédire une aptitude sportive chez le cheval. Comme évoqué dans l'introduction, ce projet stratégique pour l'élevage et la performance du cheval d'endurance français, initié en 2011, financé conjointement par l'Institut Français du Cheval et de l'Équitation (I.F.C.E.), l'I.N.R.A. (unité mixte de recherches G.A.B.I. : Génétique Animale et Biologie Intégrative) et le Fonds Eperon, a pour but de rechercher des biomarqueurs et des déterminants génétiques permettant de caractériser l'aptitude sportive de ces chevaux à l'exercice aérobie en course d'endurance. La détermination de critères phénotypiques associés aux aptitudes (en particulier à la performance), et de causes de contre-performance, chez le cheval d'endurance, permettrait de proposer aux éleveurs, cavaliers et entraîneurs, des outils pratiques et peu invasifs d'aide à la caractérisation de leurs chevaux. Ce projet se propose donc de fournir des éléments utiles au maintien de la production de chevaux d'endurance au meilleur niveau, et de contribuer à une bonne gestion des jeunes chevaux afin d'amener les meilleurs au plus haut niveau par une sélection sur des critères objectifs.

La performance en endurance résulte de multiples composantes (squelettique, musculaire, métabolique aérobie/anaérobie, comportementale...) et constitue un caractère complexe : une part est liée au milieu (élevage, alimentation, environnement), une autre est liée au génotype, avec intervention d'un grand nombre de gènes hérités des parents. Afin de préciser la part de chaque composante et leurs interactions, le projet GenEndurance a donc déployé une large campagne de phénotypage des caractères morphologiques, locomoteurs, physiologiques et métaboliques sur plusieurs générations de chevaux d'endurance, et des analyses de corrélation entre ces mesures et la performance ont été réalisées afin d'identifier d'éventuels biomarqueurs. Compte tenu de la complexité du caractère « performance en endurance », on s'attend à ce que ces corrélations directes soient faibles et que le caractère prédictif de ces biomarqueurs soit limité.

Le projet GenEndurance se démarque des autres projets de phénotypage par l'analyse des mesures phénotypiques comme autant de phénotypes « fins » afin d'identifier les régions du génome qui en contrôlent la variabilité. Pour cela, les mêmes chevaux, ainsi que leurs géniteurs, ont été génotypés à partir d'un simple prélèvement sanguin. Les analyses des marqueurs (*SNP*) identifiés par le génotypage des chevaux permettent de détecter des régions génomiques polymorphes, dont le polymorphisme explique une part de la variabilité des caractères phénotypiques (*QTL*). La comparaison des régions ainsi détectées à une référence doit permettre de préciser la part des différentes composantes dans la performance globale, d'identifier des biomarqueurs tout en affinant leur potentiel prédictif, et faciliter le cas échéant l'identification des gènes candidats impliqués dans des fonctions essentielles à la possession d'un bon potentiel de performance à l'exercice d'endurance. Pour résumer, le projet GenEndurance se base sur l'hypothèse que des facteurs génétiques contribuent à la performance en course d'endurance chez le cheval Arabe et croisé Arabe, et tente de valider ou d'infirmer cette hypothèse en réalisant une recherche d'association entre 74 000 *SNP* répartis sur l'ensemble du génome et les performances en course d'endurance (de 90 à 160 km) ainsi que les mensurations corporelles de ces races de chevaux.

Cette approche intégrée, à la fois génétique, génomique et phénotypique, doit permettre à terme d'évaluer les aptitudes d'un jeune cheval (autrement dit, de profiler son potentiel) d'après un génotypage sur prise de sang et d'après une série de mesures validées comme prédictives de la performance et/ou du risque de contre-performance.

- Pertinence du cheval d'endurance comme modèle

Le cheval d'endurance est un modèle remarquable d'exercice prolongé faisant appel presque exclusivement au métabolisme aérobie. Outre une bonne aptitude énergétique à utiliser cette voie métabolique, la durée de l'effort demandé à ces animaux implique une capacité à endurer la fatigue (notamment musculaire), évacuer la chaleur, éviter la déshydratation, et maintenir l'intégrité de leurs tissus malgré un stress métabolique. L'utilisation du modèle « cheval d'endurance » semble donc être judicieux pour détecter des marqueurs génétiques de type *SNP* proches de gènes ou de sites régulateurs à rôle important dans les voies métaboliques impliquées lors de l'effort d'endurance. Pour atteindre cet objectif, on étudie la variabilité génétique nucléaire et mitochondriale sur des chevaux ayant couru sur des épreuves de 90 à

160 km afin de mettre en relation les variants génétiques avec les différences de performances en course, quantifiées par 3 indices : vitesse moyenne, distance parcourue et classement final.

- Déroulement des 3 étapes successives et complémentaires du projet

La première étape consiste en une **caractérisation génomique et phénotypique de l'aptitude à l'endurance** sur des **chevaux sortis en compétition** (donc de 7 ans d'âge, ou plus). Des prélèvements sanguins sont réalisés sur un effectif de 645 chevaux adultes performeurs (ayant couru sur 90 km et plus en vitesse libre), déjà informatifs sur le plan des performances globales. Quelques mesures phénotypiques simples (mensurations, composition corporelle...) sont également répertoriées lors des prélèvements. L'échantillonnage choisi privilégie au maximum la reconstitution des grandes familles, en recherchant si nécessaire des ascendants (étalons ou juments) dans les élevages ; ces familles permettent d'estimer l'héritabilité des caractéristiques phénotypiques mesurées. Un génotypage de ces chevaux est réalisé à l'aide de la dernière version de puce *SNP* équine, comportant 74 000 marqueurs (puce *SNP* 74K), révélant les régions du génome associées aux performances. Cela permet d'établir des équations de prédiction et de développer un index génomique pouvant être appliqué précocement aux jeunes chevaux utilisés à l'étape suivante. Une étude de gènes candidats au sein des régions du génome mises en évidence est ensuite réalisée afin de rechercher les polymorphismes fonctionnels et d'en mesurer l'association avec les performances.

La deuxième étape est de **caractériser le phénotype et le génotype de jeunes chevaux d'endurance** (691 chevaux de 4, 5 et 6 ans). D'une part, on collecte des mesures phénotypiques à potentiel de prédiction d'aptitude afin d'étudier : i) les corrélations entre ces mesures et les résultats en compétition d'une année sur l'autre, ii) les héritabilités des mesures effectuées sur les chevaux apparentés. D'autre part, on génotype les animaux afin : i) d'identifier des marqueurs génomiques contrôlant ces paramètres, ii) d'étudier la co-localisation de ces marqueurs avec ceux associés à la performance globale, iii) d'établir leur index génomique, iv) de valider les régions mises en évidence dans la 1^{ère} étape. Cette analyse permet de décomposer la performance globale en différentes composantes physiologiques, d'identifier des biomarqueurs, d'améliorer la précision de localisation des régions génomiques d'intérêt, et de contribuer à l'identification de gènes candidats.

La troisième et dernière étape consiste en un **phénotypage plus complet** d'un **petit effectif de chevaux** (79 chevaux de 4, 5 et 6 ans, répartis sur les années 2012, 2013 et 2014). Chacune de ces trois années, une trentaine de ces chevaux est soumise à un vaste protocole de mesures (échographie cardiaque, tests d'efforts, biopsies musculaires, dosages de marqueurs biochimiques et endocrinologiques...). Ils sont réexaminés les années suivantes dans les mêmes conditions, ou lors des finales à Uzès. Cela permet de collecter des données phénotypiques d'intérêt potentiel pour la prédiction de l'aptitude qui ne peuvent pas être obtenues dans la 2^e étape en raison du temps nécessaire à leur recueil.

La partie « terrain » du projet devait prendre fin avec les finales d'Uzès du 9 au 11 Octobre 2014, mais les intempéries n'ont pas permis à cet événement d'avoir lieu comme prévu ; cette absence de prélèvements a été compensée en Décembre 2014 par un génotypage des chevaux arabes de 5 et 6 ans engagés sur les finales, ce qui a permis d'atteindre l'effectif prévu initialement et d'achever la collecte de données sur le terrain fin 2014.

Désormais, un total d'environ 1200 chevaux d'endurance est donc génotypé et une grande partie phénotypée pour de nombreux caractères, ce qui représente une base de données unique et très riche pour détecter des marqueurs génétiques d'intérêt et des biomarqueurs de la performance et de caractères secondaires. Les jeunes chevaux collectés constituent un échantillon de validation des marqueurs génétiques identifiés sur les adultes comme facteurs de performance.

Tableau III : Tableau bilan des effectifs en fin d'année 2014 (Compte Rendu GenEndurance 2014)

Types de chevaux	Années	Lieux	Nombre de sujets
<i>Compétiteurs adultes</i>	2011-2012	9 courses + 5 écuries	367 avec phénotype
<i>Jeunes chevaux (4 à 6 ans)</i>	2011 à 2014	Uzès + 3 élevages	431 + 58
<i>Jeunes chevaux (test d'effort)</i>	2012 à 2014	4 sites	79
<i>Compétiteurs adultes</i>	2011	Génotypage seul	159
<i>Jeunes chevaux</i>	2014	Génotypage seul	121

- Résultats préliminaires du projet

○ Première étape

Les résultats du phénotypage (mensurations, morphologie, allures, biochimie sanguine) ont permis :

- De décrire le modèle des chevaux arabes d'endurance par des mensurations et analyses morphométriques par analyse d'image. Sur les 54 paramètres morphologiques et morphométriques, seuls 4 critères apparaissent significativement liés à la performance (la longueur corporelle, l'épaisseur du pli de peau et quelques angles d'inclinaison du bassin et du fémur présentent une corrélation significative avec les performances ($r=0.10-0.15$; $p<0.05$)). Ainsi, un indice de distance plus élevé est associé à une longueur corporelle plus grande, un pli de peau moins épais et un angle du grasset plus fermé ; un indice de vitesse élevé est associé à un fémur plus horizontal (Tribout, 2013). Ces résultats ne permettent toutefois pas de faire de la prédiction sur ces critères. En revanche, les mensurations moyennes seront utiles aux juges de concours d'élevage : la hauteur au garrot (HG) moyenne est de 154 ± 4 cm, le périmètre thoracique (PT) moyen de $175 \pm 6,4$ cm, et la longueur pointe de l'épaule – pointe de la fesse (LC) moyenne de 154 ± 7 cm ; le poids moyen est de 410 ± 40 kg (selon la formule Poids (kg) = $2.70 \text{ HG} + 0.74 \text{ LC} + 3,1 \text{ PT} - 670.82$, avec $r^2 = 0.67$).



Figure XXXIV : Mensurations moyennes des chevaux arabes d'endurance (Robert & al., 2012)

Les pur-sang arabes sont de plus petit format que les demi-sang arabes ou les anglo-arabes. Comparé au cheval Selle Français, réputé pour ses performances en saut d'obstacles, le cheval d'endurance de haut niveau est plus petit, son encolure est proportionnellement plus longue, sa scapula est plus horizontale, son angle du jarret est plus fermé et son canon est plus court.

- De dégager, grâce aux tests accélérométriques des allures en main sur 251 chevaux la veille d'une course, une corrélation significative entre la puissance dorso-ventrale au trot et les indices de performance IRE (*De Verdière, 2014*).
- De mettre en évidence, par des tests d'allures avant la course sur trois groupes de chevaux différents, des différences significatives de la régularité des foulées au pas, ce qui place cette dernière comme prédicteur de l'abandon et de l'élimination pour raison métabolique (elle est plus élevée chez les chevaux éliminés pour causes métaboliques que chez les classés ou éliminés pour boiteries, elle est en revanche plus faible chez les chevaux qui abandonnent que chez les classés ou éliminés pour boiteries) (*Barrey & al., 2012*).
- De montrer, via des profils métabolomiques (méthode de spectroscopie RMN, détectant toutes les molécules de poids moléculaire inférieur à 200 Da dans le plasma), de nettes différences entre les profils sanguins des 69 chevaux testés avant le départ et à l'arrivée d'une course nationale de 160 km (*Le Moyec & al., 2014*), mais aussi entre 2 sites de course aux conditions climatiques et de dénivelé très différentes. La partie expérimentale de ma thèse concerne directement cet aspect du projet GenEndurance, en participant à l'analyse des relations entre profils métabolomiques, autres -omiques, et performances d'après le classement. Certaines molécules détectées sont de potentiels biomarqueurs de performance ou de contre-performance.

La comparaison des profils métabolomiques par spectroscopie RMN avant et après la course (*Le Moyec & al., 2014*) montre que l'effort d'endurance sollicite simultanément le métabolisme énergétique (avec notamment une **corrélation positive entre marqueurs lipidiques plasmatiques et vitesse moyenne chez les chevaux classés à l'arrivée**), le catabolisme protéique (ce qui s'explique par les dérèglements métaboliques musculaires, hépatiques et rénaux majeurs engendrés par le stress de l'exercice) et le métabolisme des glycoprotéines (avec une diminution de certains marqueurs glycoprotéiques dans le sérum avec l'effort). Des hypothèses ont été émises pour expliquer cette diminution de biomarqueurs glycoprotéiques (encore peu connus, même chez l'Homme), comme celle d'une dégradation de tissu conjonctif ou de cartilage articulaire liée aux contraintes mécaniques engendrées par la course, mais elles restent à valider.

Cette étude a permis, par l'analyse du profil métabolomique en spectroscopie RMN de tous les chevaux avant et après la course, de proposer un bon modèle de description et de prédiction de performance individuelle à partir de l'analyse de profils métabolomiques avant ou après la course (*Figure XXXV* ci-dessous), modèle testé et validé sur un échantillon d'autres chevaux.

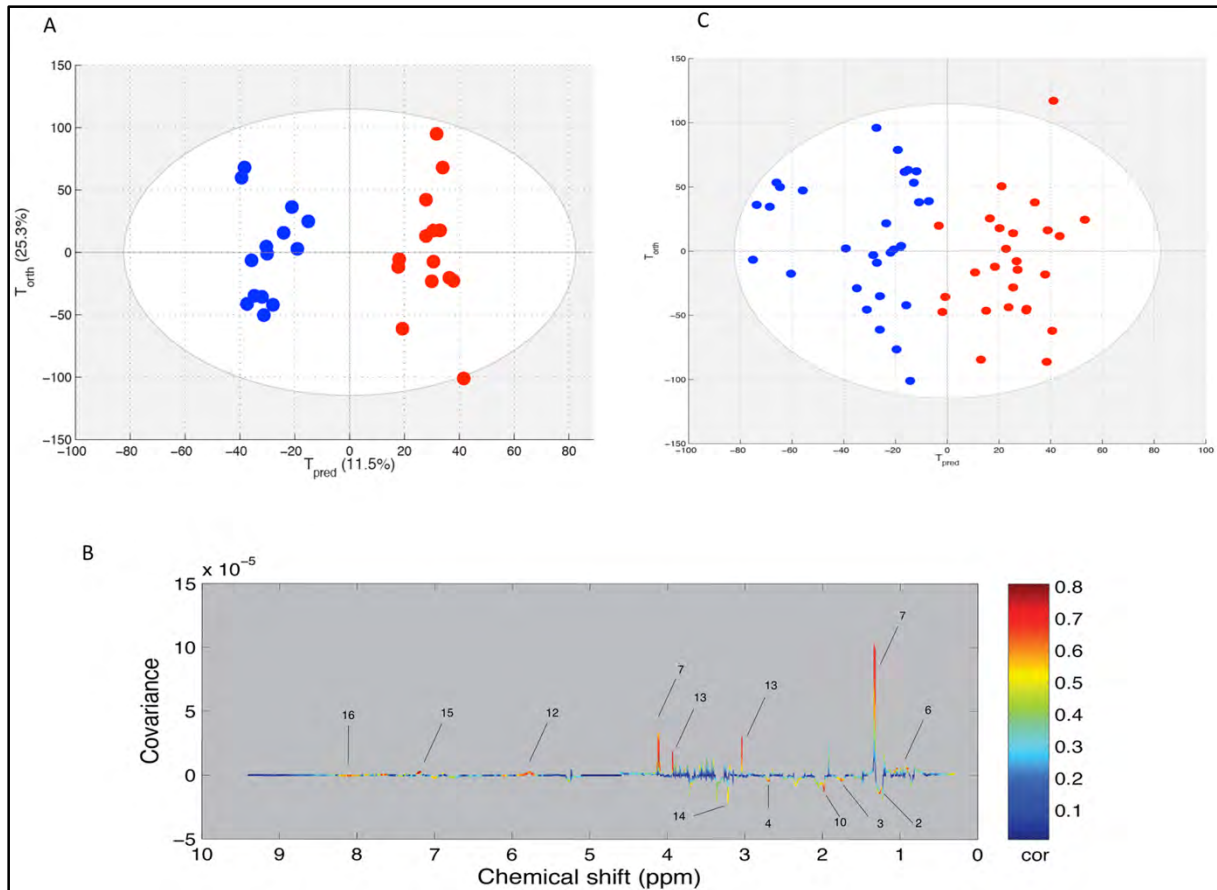


Figure XXXV : Mise en évidence via la méthode de spectroscopie RMN de différences entre les profils métabolomiques sanguins avant l'exercice (BE) et après l'exercice (PE) (Le Moyec & al., 2014)

A) Score plot du modèle O-PLS calculé avec des échantillons BE et PE du même cheval (distribution des échantillons BE et PE d'un même cheval). T_{pred} : axe prédictif, T_{orth} : axe orthogonal. Chaque point correspond à un spectre, coloré en bleu pour BE, en rouge pour PE.

B) Spectre RMN sur sérum après course : « Loading plot » de l'axe prédictif du score plot. Les corrélations des métabolites sont représentées par l'échelle colorimétrique. Les signaux positifs correspondent aux métabolites présents à des concentrations augmentées PE par rapport à BE. Inversement, les signaux négatifs correspondent aux métabolites présents à des concentrations augmentées BE. Les pics sont légendés selon les affectations de métabolites suivantes : 2, 3, et 4 sont des fractions lipidiques à chaînes d'acides gras (respectivement : méthylène, méthylène β ester, méthylène α ester), 6 représente des acides aminés ramifiés (valine, leucine, isoleucine), 7 correspond au lactate, 10 à des groupements N-acétylés (glycoprotéines), 12 à l'urée, 13 à la créatine, 14 à la phosphocholine, 15 à la tyrosine, et 16 à des protons échangeables de groupements carboxyliques.

Figure XXXV A) : On voit clairement apparaître deux groupes distincts : les échantillons individuels pré et post exercice ont donc pu être discriminés par ce modèle via leur profilage métabolomique.

Figure XXXV B) : Quelques variables physiologiques sont très corrélées au moment d'échantillonnage (avant ou après la course) : le graphique laisse clairement apparaître que l'effort prolongé entraîne ici une **augmentation** de la concentration sanguine en **lactate**, **créatine**, **urée** et quelques acides aminés ramifiés comme la **valine** et la **leucine** ainsi que des acides aminés aromatiques comme la **tyrosine**, et une **diminution** de la concentration sanguine des **lipides** à chaîne d'acides gras.

Figure XXXV C : Il s'agit d'un test du modèle déterminé avec l'analyse ; on remarque qu'il classe correctement les échantillons pré et post exercice.

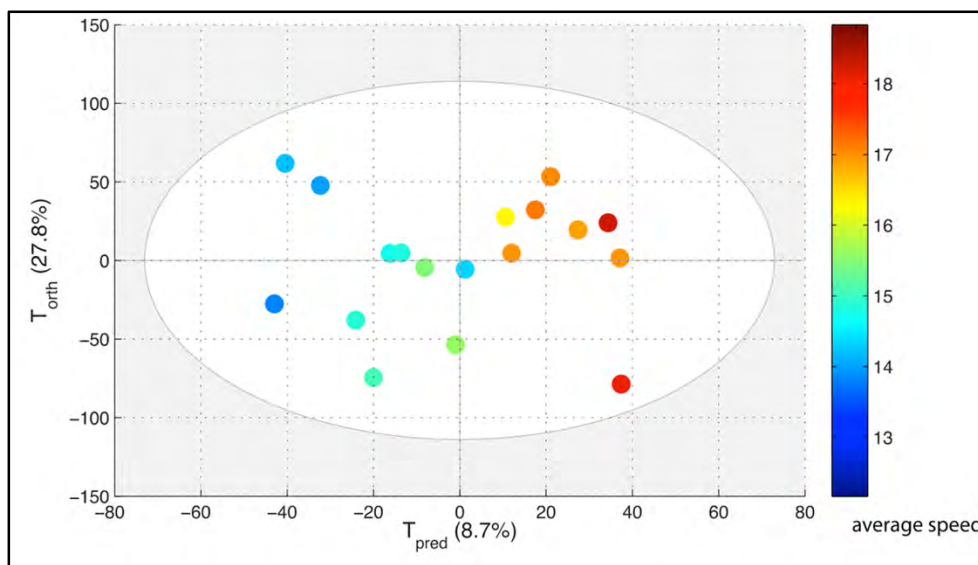


Figure XXXVI : Score plot des échantillons avant course calculé par le modèle sur des chevaux classés en fonction de leur vitesse moyenne pendant la course (Le Moyec & al., 2014)

T_{pred} : axe prédictif, T_{orth} : axe orthogonal. Chaque point correspond à un spectre, coloré en fonction de l'échelle de vitesse (« average speed », à droite du graphique).

Les spectres obtenus BE sur les chevaux classés peuvent être discriminés selon la vitesse pendant la course sur la base de leur contenu lipidique métabolomique, comme le montre la figure ci-dessus (les chevaux classés les plus rapides pendant la course présentent des concentrations plus élevées en certains lipides avant la course) ; les spectres obtenus PE sur les chevaux classés peuvent être discriminés selon le classement final des chevaux à l'issue de la course à partir de leur lactatémie plasmatique. Concernant cette dernière observation, cela concorde avec une étude de spectroscopie RMN humaine montrant que le sprint entraîne un pic de lactatémie que l'on impute au passage à un métabolisme anaérobie (Pechlivanis & al., 2013) ; ainsi, les chevaux les mieux classés sont sans doute ceux qui ont les capacités à effectuer un sprint final de quelques centaines de mètres au galop en utilisant leur métabolisme anaérobie, ce qui s'accorde bien avec une lactatémie plus élevée parmi les échantillons post exercice.

L'étude de L. Le Moyec & al. montre que l'analyse de profils métabolomiques associés à l'exercice et au stress généré apparaît informative quant à la grande demande énergétique et au catabolisme protéique sévère résultant d'un effort extrême comme une course d'endurance de 160 km, exposant potentiellement le cheval à des désordres métaboliques. La méthode RMN plasmatique / métabolomique s'avère prometteuse pour évaluer les ajustements à l'entraînement et en course d'endurance. Après cette première approche sur

les chevaux d'endurance adultes performeurs, elle a donc été utilisée sur les jeunes chevaux aux étapes suivantes du projet ; les analyses sont actuellement en cours.

Les chevaux adultes performeurs en endurance prélevés et génotypés dans le cadre du projet sont en moyenne un peu plus performants que l'ensemble de la population des chevaux indicés en endurance. L'analyse a été menée sur une population de 597 chevaux (performers et collatéraux) de génotype exploitable et dont les indices de performance sur la période 2002-2011 sont référencés (IRE_{distance} , IRE_{vitesse} , $IRE_{\text{classement}}$, IRE_{global}). Le génotypage de ces chevaux sur puce *SNP* 74K (parmi les 74 000 marqueurs, 56 200 *SNP* étaient utilisables) a permis de détecter quelques marqueurs génétiques significatifs ; ces QTL regroupent l'ensemble des *SNP* proches et en fort déséquilibre de liaison pour lesquels le test passe le seuil de significativité. Ainsi 4 QTL ont été détectés par 11 *SNP* significativement associés : i) à l'aptitude à l'endurance ($p.\text{value} \leq 10^{-5}$) : localisés sur les chromosomes 6 et 16 pour la performance de distance et la performance de classement (c'est-à-dire le fait d'être classé ou éliminé) d'une part, et sur les chromosomes 7 et 29 pour la performance de classement seule d'autre part ; ii) à la conformation ($p.\text{value} \leq 10^{-5}$) : localisés sur le chromosome 1 pour le rapport périmètre thoracique/longueur corporelle, et sur le chromosome 3 pour la hauteur au garrot chez les chevaux croisés Arabe uniquement.

L'effectif de chevaux utilisé (environ 1 200) représente des efforts et coûts déjà considérables, en particulier au vu de la faible disponibilité de chevaux d'endurance. La taille du dispositif disponible reste néanmoins relativement modérée (en comparaison avec d'autres travaux de génétique d'association conduits chez l'Homme par exemple, qui mobilisent classiquement des cohortes de plusieurs dizaines voire centaines de milliers d'individus), limitant la puissance des analyses statistiques, et expliquant le faible nombre de régions QTL significatives détectées.

Des essais de séquençage du génome mitochondrial d'environ 350 chevaux ont permis d'entamer des études d'association entre *SNP* mitochondriaux et performances.

- Deuxième étape

Le génotypage sur puce 74K et le phénotypage (prélèvements sanguins, mesures morphométriques, enregistrement accélérométrique des allures par Equimétrie®, évaluation

comportementale, échographie cardiaque) de chevaux de 4, 5 et 6 ans de 2011 à 2013 ont, pour l'instant, permis :

- De montrer que les mensurations des jeunes chevaux pur-sang arabes évoluent significativement entre 4 et 6 ans (en fin de croissance), ce qui n'est pas le cas chez les anglo-arabes et demi-sang arabes.
- De relever des modifications cardiaques significatives entre 4 et 6 ans d'âge (notamment une augmentation de taille du ventricule gauche entre 5 et 6 ans, et une augmentation des dimensions de l'aorte et du volume d'éjection systolique avec l'âge), sachant que la part de ces modifications liée à l'entraînement et celle liée à la croissance doivent encore être différenciées (*Trachsel & al., 2014*).
- De mettre en évidence une corrélation entre la capacité anti-oxydante et la concentration en enzymes musculaires dans le plasma de chevaux de 6 ans avant et après la course ($p < 0.05$). La résistance au stress oxydatif avant la course est plus élevée chez les chevaux les plus rapides sur la course ($p < 0.05$) (*Robert & al., 2014*).
- De constater, via la méthode RMN / métabolomique, une modification forte des métabolismes protéique, lipidique et glucidique par une course de 90 km, chez les chevaux de 6 ans (*Luck & al., 2015*). Cependant, les métabolites utilisés sont différents entre ces jeunes chevaux sur 90 km et le groupe de chevaux performeurs adultes sur 160 km utilisé dans l'étude de L. Le Moyec. Le profil métabolomique déterminé avec le modèle OPLS montre une glycémie plus basse après la course qu'avant la course chez les jeunes chevaux, et cela malgré le recours au catabolisme lipidique et protéique. Les chevaux expérimentés maintiennent leur glycémie avec une lactatémie plus élevée et une lipémie plus basse après la course en comparaison aux jeunes chevaux. Il reste à déterminer dans quelle mesure ces différences sont imputables à l'entraînement, à la durée de l'effort ou à la capacité intrinsèque du cheval.

- Troisième étape

Suite au phénotypage plus complet d'une centaine de jeunes chevaux d'endurance (prélèvements et mesures réalisés avant, pendant et après un test d'effort standardisé sur piste), des vitesses pour lesquelles la fréquence cardiaque atteint 160 et 180 battements par minute (respectivement V160 et V180) ont pu être déterminées pour chaque cheval. Les analyses révèlent que ces vitesses augmentent significativement avec l'âge du cheval (4, 5 ou 6 ans). L'analyse de l'ensemble de l'effectif et le suivi des performances de ces chevaux à l'âge adulte

sont en cours, afin de déterminer si les valeurs de V160 et V180 par classe d'âge peuvent être prédictives des capacités sportives du cheval testé.

En parallèle aux résultats des trois étapes, des articles analysant les données enregistrées sur diverses courses d'endurance sont progressivement publiés. Par exemple, M. Younes & al. ont montré en 2015 que l'on pouvait prédire la probabilité d'élimination des chevaux à un point de contrôle vétérinaire en course d'endurance (*vet gate*) à l'aide de régressions logistiques prenant en compte les données de fréquence cardiaque, de temps de récupération cardiaque (délai entre l'arrivée du cheval et la présentation de ce dernier par son cavalier à l'examen vétérinaire, à la fin d'une phase de la course) et de vitesse moyenne enregistrées au *vet gate* précédent (Younes & al., 2015). Ces régressions logistiques prédictives appliquées aux points de contrôle 2 à 5 ont correctement classé de 62 à 86% des chevaux éliminés. Parmi les trois variables explicatives, le temps de récupération cardiaque apparaît particulièrement pertinent pour détecter les chevaux n'étant pas aptes à continuer la course. En effet, cette étude a permis de montrer qu'un cheval a 70% de chance d'être éliminé au *vet gate* suivant si son temps de récupération cardiaque est supérieur à 11 minutes pour les *vet gates* 1 et 2, ou supérieur à 13 minutes pour les *vet gates* 3 et 4. Les auteurs proposent donc de vérifier ce temps à chaque point de contrôle vétérinaire afin d'identifier au plus tôt et de disqualifier les chevaux inaptes à continuer la course, et ainsi de contribuer à leur bien-être. Cette étude a notamment fait l'objet d'une présentation à l'occasion de la 42^e Journée de la Recherche Equine le 18 Mars 2016 à Paris.

Tous ces résultats préliminaires du projet GenEndurance devront être complétés, précisés et expliqués lorsque les données de l'ensemble de l'effectif évalué auront été entièrement analysées. Le passage à l'année 2016 a marqué la fin de l'analyse « simple » des données, c'est-à-dire l'exploitation isolée de chaque type d'information phénotypique (locomotion, morphologie, activité cardiaque...). Cette étape était indispensable pour déterminer la distribution des critères, l'influence des facteurs fixes (âge, race, sexe, lieu de mesure) et les liens entre paramètres. Il s'agit désormais d'aller plus loin dans la détermination des corrélations entre phénotype et performance d'une part, et phénotype et génotype d'autre part.

La partie expérimentale de ma thèse, qui établit un lien entre certaines données – omiques et les performances de chevaux d'endurance, intervient directement dans la suite des

analyses du projet GenEndurance. Pour cela, nous avons réalisé un tour d’horizon des travaux déjà publiés en termes d’analyses –omiques dans un cadre d’endurance équine.

ii. Etude sur les transcriptome (ARNm), microtranscriptome (miARN), métabolome (métabolites) et leurs liens lors d’un effort d’endurance chez le cheval

Les efforts réalisés en génomique équine dans l’optique de l’étude et de la prédiction d’un phénotype équin d’endurance sont notamment illustrés par la thèse de Sophie Brard (*Brard, 2015*) et de multiples études telles que celle de A. Ricard & al., montrant que les valeurs de sélection des chevaux d’endurance devraient être estimées sur des critères de vitesse et de classement et uniquement sur des courses de chevaux adultes, d’une distance supérieure à 90 km (*Ricard, Touvais, 2007*).

Il s’agit maintenant de s’intéresser aux études transcriptomiques, microtranscriptomiques, protéomiques et métabolomiques dans le cadre de l’endurance équine, et d’établir, si possible, des liens entre ces données de nature différente. En effet, l’ajustement cellulaire à un nouvel état d’homéostasie implique la modification de certains aspects de la physiologie cellulaire, et les réponses de stress telles que celle induite par un effort d’endurance sont caractérisées par des modifications d’expression génétique, de métabolisme, de progression dans le cycle cellulaire et d’homéostasie protéique (*Capomaccio & al., 2013*). Ces réponses agissent à des échelles de temps diverses, allant des effets post-transcriptionnels qui peuvent être à l’origine de réponses immédiates, à la régulation de l’expression génétique qui est essentielle pour les ajustements plus lents et à plus long-terme, et pour les phases de récupération, avec dans tous les cas une réversibilité contrôlée.

Une des premières études transcriptomiques dans le cadre d’efforts d’endurance équine a été menée afin de montrer que l’expression de certains gènes est modulée dans les leucocytes de chevaux, en lien avec leur performance en endurance et leur statut clinique (cette étude fait partie de celles qui ont motivé la construction de projets comme GenEndurance) (*Barrey & al., 2006*). Les globules blancs sont des cellules de choix dans l’étude de phénomènes inflammatoires pendant l’exercice. En effet, ils circulent via le flux sanguin dans tous les organes impliqués dans l’effort physique et sont impliqués dans la réponse inflammatoire, notamment musculaire et articulaire. On s’attend donc à ce que ce stress physiologique soit

observable à l'échelle du matériel génétique de ces cellules, puisque toutes les synthèses protéiques et régulations métaboliques sont gouvernées par des régions du génome plus ou moins exprimées. Ainsi, E. Barrey & al. ont comparé les profils d'expression génétique dans les leucocytes et les paramètres hématologiques et biologiques de chevaux qualifiés et éliminés sur critères métaboliques à l'issue d'une course d'endurance, et ont constaté des différences significatives entre ces deux groupes de chevaux, notamment 62 gènes différentiellement exprimés. Cette étude a mis en évidence les modulations génétiques induites par un effort prolongé dans les leucocytes de chevaux d'endurance, et a montré que certains gènes sont exprimés en lien avec le phénotype clinique observé chez les chevaux disqualifiés (rhabdomyolyse confirmée par les concentrations plasmatiques des enzymes musculaires créatinine kinase et aspartate aminotransférase, hémolyse, et inflammation), les plaçant comme potentiels marqueurs (et/ou candidats) d'un phénotype de performance médiocre en endurance. Suite à l'étude d'E. Barrey & al., d'autres analyses transcriptomiques ont donc été réalisées chez le cheval d'endurance, confirmant que dans cette espèce comme chez l'Homme, des modifications substantielles des profils d'expression génétique de leucocytes circulants prennent rapidement place au cours de l'exercice (*Barrey & al., 2006; Cappelli & al., 2007; Fehrenbach, 2006; Cappelli & al., 2009*). Qu'il s'agisse de l'espèce humaine ou équine, un effort physique prolongé est à l'origine d'une immunodépression temporaire, induisant un état inflammatoire dépendant de l'intensité et de la durée de l'exercice. S. Capomaccio & al. se sont appuyés sur l'hypothèse que l'état physiopathologique développé chez les athlètes sujets à un entraînement intensif est causé par un dérèglement de l'immunité cellulaire (*Purvis & al., 2010*), et ont cherché à obtenir des informations sur les modulations du transcriptome de chevaux d'endurance soumis à un effort physique éprouvant tel qu'une course d'endurance, et à identifier des gènes candidats à l'origine du dérèglement immunitaire. Ils ont, à cet effet, réalisé un scan d'expression génétique par la technologie des biopuces, sur des *PBMC* de 10 chevaux d'endurance d'élite, à trois temps différents : au repos, (T_0), immédiatement à l'issue de la course (T_1), puis 24 heures après la course (T_2) (*Capomaccio & al., 2010*). Leur étude n'a pas révélé de différence significative entre T_0 et T_2 , suggérant une restauration complète de l'homéostasie 24 heures après la course, mais a mis en lumière une forte modulation d'expression, affectant 132 gènes (dont 97 surexprimés, 35 sous-exprimés) entre l'état de repos et l'arrivée de la course. Une analyse *Ingenuity Pathway Analysis* (logiciel permettant de modéliser, analyser, et comprendre des systèmes biologiques et chimiques, en intégrant des données -omiques complexes) a permis d'observer que les principaux mécanismes et fonctions biologiques liés à ces gènes affectés par l'exercice sont clairement impliqués dans la réponse immunitaire et

l'inflammation et a mis l'accent sur les gènes codant l'IL-8 (interleukine-8) et deux autres cytokines (CCL5, CCL3), fortement surexprimés à l'issue de la course dans cette étude. Enfin, une étude transcriptomique récente a utilisé la méthode d'ARN-seq afin de séquencer le transcriptome d'athlètes équins dans un cadre expérimental de stress induit par l'exercice pendant les courses d'endurance (Capomaccio & al., 2013). Cette analyse de réseau souligne surtout l'implication de clusters de gènes dans le stress induit par un exercice physique tel que la course d'endurance, incluant ceux ayant un rôle avéré dans l'inflammation, le signalage cellulaire, et les interactions immunitaires. En effet, le suivi du « paysage transcriptionnel » par profilage d'expression génétique avant et après une course d'endurance chez des chevaux sportifs de haut niveau a révélé des différences d'expression significatives de nombreux gènes liés à l'inflammation et à l'activation du système immunitaire (codant des cytokines de type chémokines pro-inflammatoires, des récepteurs Toll-like (*Toll-Like Receptors*, *TLR*), des kinases...), avec un déplacement franc d'expression de transcrits codants à des transcrits non-codants. Ce déplacement suggère que la réponse de stress implique la régulation de régions non annotées, nouvellement actives d'un point de vue transcriptionnel : ce phénomène d'« exonisation » ou de « rétention d'introns » a été étudié dans un article (Schmitz, Brosius, 2011). Les augmentations significatives des niveaux d'expression post-course de séquences répétées (comme les rétrotransposons L1 et L2) concordent avec l'hypothèse que des éléments transposables et séquences introniques puissent servir comme unités transcriptionnelles capables d'enrichir les transcriptomes à ressources génomiques limitées quand cela est nécessaire, comme en situation de stress par exemple (Schmitz, Brosius, 2011). L'identification d'un grand nombre de transcrits correspondant à des régions intergéniques et introniques associés de manière causale à de nouveaux éléments transcriptionnels comme par exemple les lncRNA suggère le rôle important de ces derniers dans la régulation de l'expression génomique en réponse au stress. Enfin, avec les augmentations significatives des niveaux d'expression de séquences répétées (L1 et L2) et de brins antisens comparés aux brins sens, à la fois dans des régions introniques et régulatrices, les résultats de cette étude suggèrent que l'« exonisation » liée aux séquences répétées pourrait être une réponse de stress chez le cheval, et que la transcription antisens pourrait être un des mécanismes principaux de régulation d'éléments transposables chez le cheval sous des conditions de stress.

La réponse de stress dans le contexte d'endurance équine est également abordée à l'échelle protéomique chez le cheval. Des études se sont d'abord concentrées sur le protéome musculaire de chevaux à l'entraînement. L'une d'entre elles a analysé le protéome équin à

différents stades d'entraînement normal et intensif, notamment pour rechercher des biomarqueurs du surentraînement, en utilisant des petites biopsies musculaires de ces animaux (Bouwman & al., 2010). Elle a révélé des différences quantitatives de protéines musculaires liées à l'entraînement (avec une concentration modifiée de 16 protéines suite à l'exercice) associées à un changement structural du muscle squelettique après un entraînement normal vers une augmentation des capacités oxydative, de dégradation des acides gras à longue chaîne, et de stockage d'énergie sous forme de glycogène. Cette étude a, de plus, mis en évidence une protéine (l' α 1-antitrypsine, glycoprotéine sécrétée pouvant protéger les tissus des effets d'enzymes de cellules inflammatoires comme l'élastase) dont la concentration est augmentée suite à un entraînement intense, mais reste inchangée après un entraînement normal, la plaçant ainsi comme marqueur potentiel de surentraînement chez les chevaux. F. Scoppetta & al. ont, quant à eux, caractérisé les modifications protéiques plasmatiques chez les athlètes équins suite à une course, dans l'optique d'une identification de biomarqueurs : cette étude présente l'originalité d'être basée sur la protéomique plasmatique et non pas musculaire dans le cadre de l'exercice physique d'endurance (Scoppetta & al., 2012). Des variations quantitatives de composants spécifiques (concentrations modifiées de 13 protéines) suite à la course traduisent une activation parallèle de la phase aiguë de la réaction inflammatoire, de la coagulation, de l'immunomodulation, de l'activité oxydante/anti-oxydante, et des voies de dégradation cellulaires et vasculaires, résultant en des effets sur les systèmes métabolique, immunitaire et neuroendocrinien équins. Le retour de la majorité de ces paramètres à l'état initial en 24 à 48 heures suggère encore une fois que la réponse de l'animal à un effort éprouvant peut être considérée comme un ajustement rapide à une condition de stress aiguë. Ces approches protéomiques, par leur nature intégrative (prenant en compte l'ensemble des protéines et des voies biologiques associées), fournissent une vue globale des adaptations moléculaires à certains facteurs (ici, l'effort d'endurance), ce que les techniques conventionnelles antérieures ne permettaient pas.

Si l'on s'intéresse maintenant aux analyses métabolomiques en lien avec des performances d'endurance équine, on se rend compte que peu de données sont pour l'instant disponibles. Les études les plus pertinentes à ce propos ont été présentées plus haut, dans le cadre du projet GenEndurance (Le Moyec & al., 2014; Luck & al., 2015). Avant l'utilisation d'approches métabolomiques, l'aspect métabolique lié aux performances équines d'endurance était abordé d'une manière moins globale (les études ne traitaient pas l'ensemble des métabolites d'un biofluide), mais on peut tout de même mentionner quelques analyses

antérieures ayant étudié certains métabolites de chevaux d'endurance dans le cadre d'une course. L'étude conduite par L. Adamu & al. (Adamu & al., 2012) a notamment mis en évidence des différences de concentration en acide urique, pro-oxydant fort pendant l'exercice d'endurance équine (biomarqueur de stress oxydatif), avant et après une course, entre les chevaux qualifiés et les chevaux disqualifiés pour troubles métaboliques à l'issue de cette course, et a montré que ce biomarqueur de stress oxydatif était corrélé de manière significative à la fréquence cardiaque (paramètre physiologique), à la concentration en triglycérides et à la lactatémie plasmatiques (paramètres biochimiques) ainsi qu'à l'hématocrite (paramètre cellulaire sanguin), plaçant l'acide urique comme potentiel paramètre d'évaluation de performance et d'état de santé chez les chevaux d'endurance lors de l'entraînement et pendant une course d'endurance. Si l'on confronte ce type d'approche à celui de l'étude de L. Le Moyec & al. par exemple, on voit bien la richesse qu'apporte l'analyse métabolomique, utilisant un profil de tous les métabolites et non pas un seul paramètre métabolique tel que l'acide urique.

Pour finir et pour établir un lien avec la partie expérimentale de cette thèse, on peut se concentrer sur une approche mêlant transcriptomique et microtranscriptomique ayant permis de mettre en évidence des modifications d'expression de gènes du métabolisme glucidique associées à des changements microtranscriptomiques après l'exercice physique chez des chevaux Pur-Sang (Gim & al., 2014). Cette étude intégrative novatrice a dégagé des notions importantes : la sous-expression de gènes du métabolisme glucidique *LDHA* (codant la lactate deshydrogénase) et *GYS1* (codant la glycogène synthase 1) associée à la surexpression de deux miARN, eca-miR-33a et eca-miR-17, suite à l'exercice physique, montre le rôle physiologique de ces miARN dans la régulation du métabolisme glucidique dans un contexte sportif. eca-miR-33a et eca-miR-17 inhiberaient respectivement les gènes *LDHA* et *GYS1* en se fixant sur des sites cibles spécifiques dans la région 3'UTR de chaque gène. Cette analyse suggère que suite à un effort physique, la machinerie physiologique équine utilise une voie biologique se rajoutant à celle du métabolisme glucidique « simple », permettant une supplémentation énergétique efficace, par la régulation de l'expression génétique via les miARN. L'approche utilisée dans l'étude de J. Gim & al., établissant un lien transcriptome/microtranscriptome dans le cadre de l'effort physique équin, est très similaire à celle que nous avons utilisée dans la partie expérimentale de ma thèse, à la différence que la nôtre a été réalisée sur des chevaux d'endurance et non des Pur-Sang de course, et que nous espérons compléter notre travail d'une analyse métabolomique, permettant d'avoir une vision intégrée et globale des modifications liées à l'effort d'endurance chez le cheval athlète. Le schéma bilan qui suit montre le principe

de fonctionnement d'une telle approche intégrée, et permet de comprendre sur quelles bases notre étude a été construite.

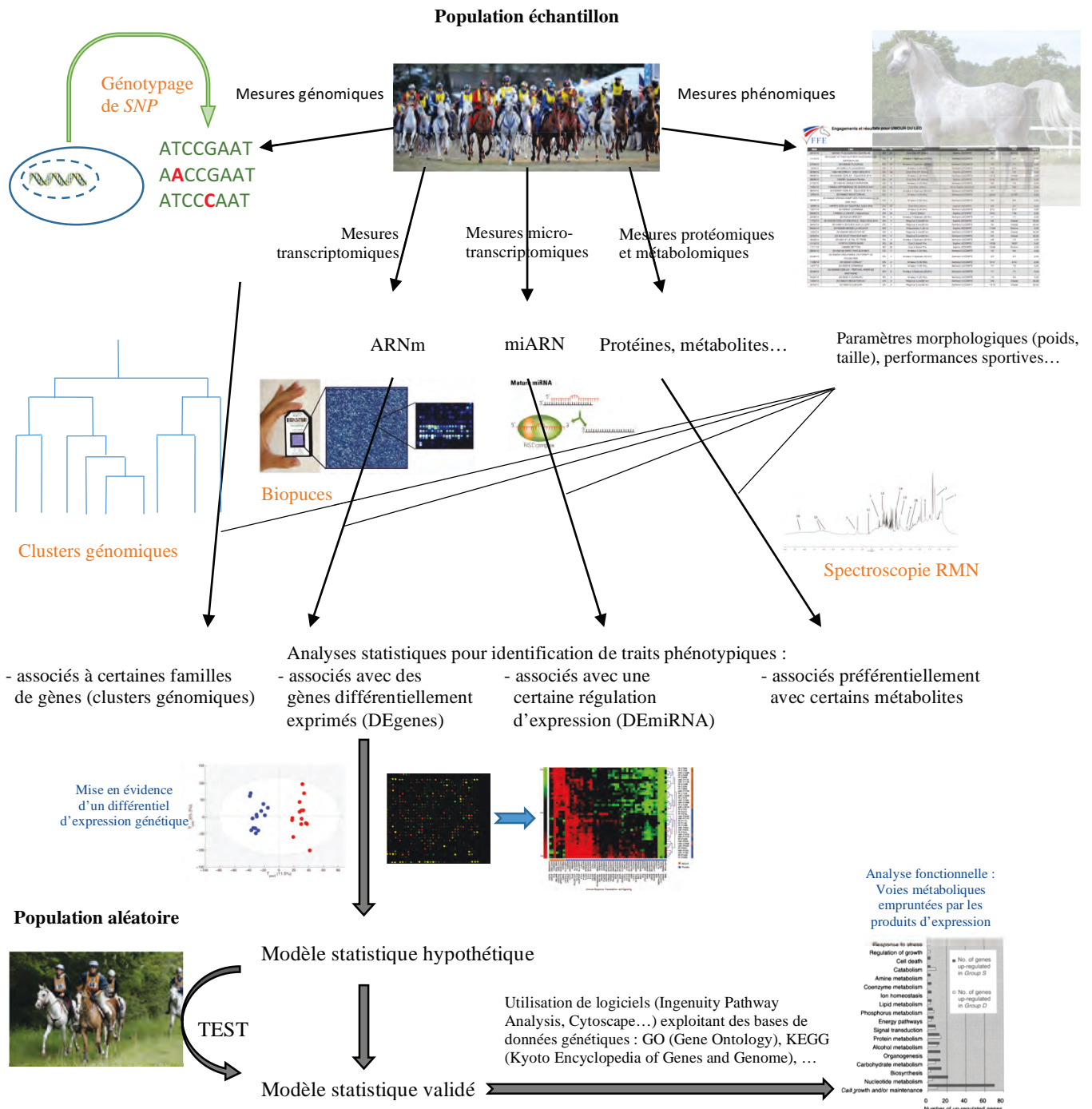


Figure XXXVII : Schéma bilan explicatif d'une analyse intégrée de données -omiques dans le cadre de l'étude et de la prédiction de statuts métabolique et de santé de chevaux d'endurance

DEgenes = gènes présentant des différences significatives d'expression entre deux états (par exemple : avant la course et après la course), DEmiRNA = miARN présentant des différences significatives d'expression entre deux états.

DEUXIEME PARTIE :

Etude expérimentale

II. ETUDE EXPERIMENTALE

1. Introduction et objectifs de l'article

Comme nous l'avons vu dans la première partie de cette thèse, la réponse musculaire et systémique à l'exercice d'endurance engendre des ajustements physiologiques et biochimiques conséquents, tels que la modification des fonctions musculaires, la correction du déséquilibre électrolytique, la prise en charge de la déplétion en glycogène, la modulation du stress oxydatif, de la perméabilité intestinale, des réponses inflammatoire et immunitaire... Ces ajustements sont influencés par la régulation transcriptionnelle et traductionnelle des gènes qui codent les protéines contrôlant ces processus. Parmi les éléments clés intervenant dans cette régulation, se trouvent les miARN : certains d'entre eux sont directement en lien avec les ajustements micro et macroscopiques observés en réponse à l'effort d'endurance chez l'Homme.

Nous avons donc émis l'hypothèse que le sang total (contenant des cellules qui circulent dans le corps, notamment dans les organes perfusés de manière accrue lors de l'effort d'endurance) pourrait fournir un marqueur précoce mesurable de la fonction immunitaire d'un athlète équin, et serait assez sensible pour détecter un stress et des troubles métaboliques induits par l'effort. Ainsi, nous nous sommes proposés de vérifier si l'identification de relations entre miARN et ARNm sanguins régulés de manière spécifique par l'exercice d'endurance chez les chevaux pouvait révéler des biomarqueurs uniques de la réponse de stress à l'endurance et fournir un aperçu pertinent du contrôle moléculaire de cette réponse.

Pour répondre à cette question, nous avons réalisé une analyse intégrée du transcriptome et du microtranscriptome de 61 athlètes équins avant et après une course d'endurance d'une distance de 160 km.

SCIENTIFIC REPORTS

OPEN

Integrated mRNA and miRNA expression profiling in blood reveals candidate biomarkers associated with endurance exercise in the horse

Received: 02 November 2015

Accepted: 25 February 2016

Published: 10 March 2016

Núria Mach¹, Sandra Plancade², Alicja Pacholewska³, Jérôme Lecardonnel¹, Julie Rivière¹, Marco Moroldo¹, Anne Vaiman¹, Caroline Morgenthaler¹, Marine Beinat¹, Alizée Nevot¹, Céline Robert^{1,4} & Eric Barrey^{1,5}

The adaptive response to extreme endurance exercise might involve transcriptional and translational regulation by microRNAs (miRNAs). Therefore, the objective of the present study was to perform an integrated analysis of the blood transcriptome and miRNome (using microarrays) in the horse before and after a 160 km endurance competition. A total of 2,453 differentially expressed genes and 167 differentially expressed microRNAs were identified when comparing pre- and post-ride samples. We used a hypergeometric test and its generalization to gain a better understanding of the biological functions regulated by the differentially expressed microRNA. In particular, 44 differentially expressed microRNAs putatively regulated a total of 351 depleted differentially expressed genes involved variously in glucose metabolism, fatty acid oxidation, mitochondrion biogenesis, and immune response pathways. In an independent validation set of animals, graphical Gaussian models confirmed that miR-21-5p, miR-181b-5p and miR-505-5p are candidate regulatory molecules for the adaptation to endurance exercise in the horse. To the best of our knowledge, the present study is the first to provide a comprehensive, integrated overview of the microRNA-mRNA co-regulation networks that may have a key role in controlling post-transcriptomic regulation during endurance exercise.

The physiological and biochemical demands of endurance exercise elicit both muscle-based and systemic responses. The main adaptations to endurance exercise include improvement of mechanical, metabolic, neuromuscular and contractile functions in muscle¹, correction of electrolyte imbalance², a decrease in glycogen storage³ and an increase in mitochondrial biogenesis in muscle tissue⁴, and the modulation of oxidative stress⁵, intestinal permeability, muscle damage, systemic inflammation and immune responses⁵. Consequently, adaptations to endurance exercise are influenced by the transcriptional and translational regulation of genes that encode the proteins controlling these processes⁵. Over the past decade, microRNAs (miRNAs) have emerged as novel elements in the rapid, reversible regulation of transcription and translation⁶. MiRNAs are small non-coding RNAs molecules (~19–24 bp in length) that are synthesized from short hairpin precursors and that reportedly degrade or inhibit the translation of their target genes by binding to the 3' untranslated region (UTR) of coding mRNAs⁷. In fact, miRNAs molecules may regulate up to one-third of the mammalian transcriptome⁸ and appear to be stable outside the cell (e.g. when incorporated into exosomes⁹, microvesicles¹⁰, lipoproteins¹¹ or Argonaute2 protein complexes¹²).

¹Animal Genetics and Integrative Biology unit (GABI), INRA, AgroParis Tech, Université Paris-Saclay, 78350, Jouy-en-Josas, France. ²INRA, MaLAGE, Jouy-en-Josas, France. ³Swiss Institute of Equine Medicine, Institute of Genetics, University of Bern and Agroscope, Bern, Switzerland. ⁴Université Paris-Est, Ecole Vétérinaire d'Alfort, Maisons-Alfort, France. ⁵Integrative Biology and Exercise Adaptation unit (UBIAE), EA7362, Université d'Evry Val d'Essonne, Evry, France. Correspondence and requests for materials should be addressed to N.M. (email: nuria.mach@jouy.inra.fr)

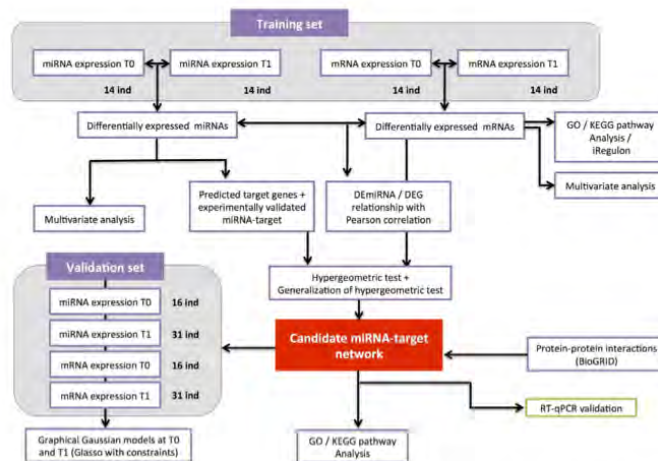


Figure 1. Overview of the data analysis. Step 1: a linear model analysis of DEmiRNAs and DEGs. Step 2: determination of the corresponding GO terms, KEGG pathways and TFs regulating the DEGs. Step 3: selection of experimentally validated target genes. Step 4: generation of the inverse correlation matrix for DEGs and miRNAs (in the hypergeometric test only). Step 5: the enrichment test (using the hypergeometric test or its generalization) used to select candidate miRNAs. Step 6: the functional network analysis; Step 7: validation of the functional network, using a validation set and GGMs; Step 8: Technical validation of the functional network, using RT-qPCR.

It has been shown that some miRNAs can modulate mitochondrial biogenesis and glucose and fatty acid metabolism in skeletal and cardiac muscle after endurance exercise in humans^{1,13}. Furthermore, circulating miRNAs induced by exhaustive exercise in humans may have an important role in the regulation of angiogenesis, inflammation, skeletal and cardiac muscle contractility and damage, and adaptation to hypoxia/ischaemia^{13–19}, whereas miRNA expression in peripheral blood mononuclear cells (PBMCs) may regulate inflammatory processes^{20,21}. Whole blood-derived miRNAs induced in trained athletes after a 30-minute treadmill test may regulate immune function, apoptosis, membrane traffic of proteins and transcription regulation²². Whole blood-derived miRNAs that control the stress response to endurance exercise have never previously been studied. However, we hypothesized that whole blood (which contains cells that circulate throughout the body) might be able to provide a rapidly measurable, effective marker of an athlete's immune function and might also be sensitive enough to detect exercise-induced stress and metabolic disorders.

Given the miRNAs' fundamental role in transcription and transcriptional regulation during physiological adaptation to endurance exercise, we hypothesized that the identification of whole blood miRNA-mRNA relationships specifically regulated by endurance exercise in horses could (i) reveal unique biomarkers of the stress response to endurance exercise and (ii) provide significant insights into the molecular control of this response. We therefore performed an integrated analysis of the blood transcriptome and miRNome in the horse before and after a long (160 km) endurance competition.

Results

The morphological and physiological parameters of the 61 equine athletes' included in the study are summarized in Supplementary Table S1, whereas the biochemical parameters average obtained from blood samples collected after the ride are depicted in the Supplementary Table S2. All horses had an above-normal total bilirubin, creatine kinase, aspartate transaminase and serum amyloid A concentrations, reflecting haemolysis and muscular membrane permeability or inflammation.

On the basis of equine transcriptome and miRNome profiles before and after a 160 km ride, we used a multi-step approach to identify and characterize a dynamically co-regulated miRNA-mRNA network related to endurance exercise in the horse (Fig. 1).

Differentially expressed genes (DEGs) and miRNAs. We used custom equine mRNA and miRNA microarrays to study the effect of endurance exercise on the blood transcriptome and miRNome, respectively, using an experimental set of 14 animals. We focused on the identification of genes whose expression was significantly altered at T1 (post-ride) relative to T0 (baseline, pre-ride), with an adjusted p -value < 0.05 after Bonferroni correction for multiple testing (Supplementary Table S3). Application of this threshold led to the identification of 2,453 DEGs, of which 1,165 were over-expressed and 1,288 were under-expressed at T1 (relative to T0).

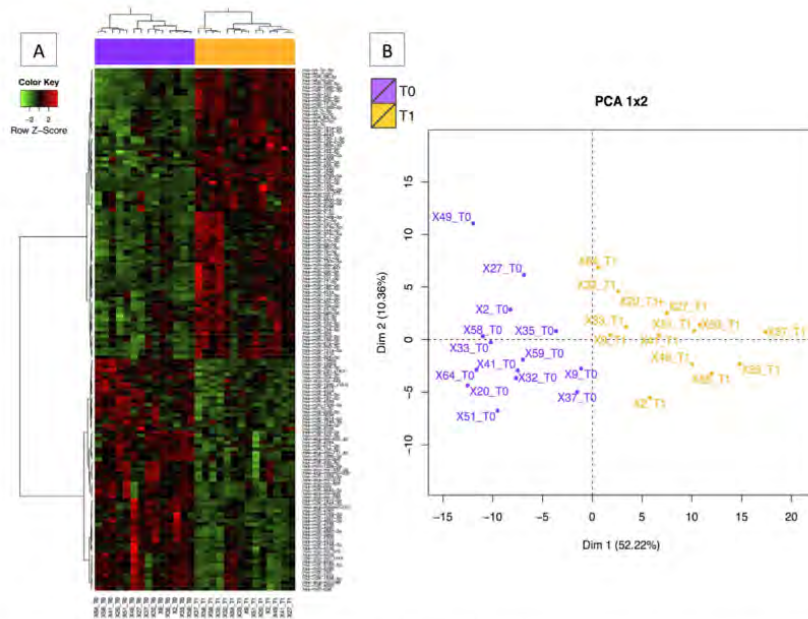


Figure 2. The differential miRNA expression profile in blood. In all cases, individual horses are represented as violet dots (for T0) and orange dots (for T1). (A) A heat-map representation of the 167 DE miRNAs when comparing T0 and T1 (FDR < 0.05). (B) PCA of DE miRNAs in blood when comparing T1 with T0 (FDR < 0.05). The first axis accounted for 52.22% of the variation, and the two axes accounted for 62.58% of the total inertia. The two groups were found to differ significantly in a Monte Carlo test with 999 replicates ($p < 0.001$).

A total of 362 miRNAs were found in the whole blood sample (Supplementary Table S4). We identified differentially expressed 167 miRNAs (DE miRNAs) when comparing pre- and post-ride samples (Supplementary Table S5). Along with orthologous human-equine DE miRNAs, we detected 12 equine-specific DE miRNAs and 19 putative novel miRNAs. Two of the putative novel miRNAs (mitomiR-009 and mitomiR-010) were probably encoded by the mitochondrial genome and thus are referred to here as mitomiRs (Supplementary Fig. S1). The effect of exercise on the miRNome was also evidenced by a cluster analysis and a principle component analysis (PCA) (Fig. 2). The first principal component accounted for 52.22% of the total variance, and the first two components accounted for 62.58% of the total variance.

To gain a better understanding of the functional implications of these DEGs, we performed gene set enrichment analysis in order to identify over-represented “biological process” gene ontology (GO) terms and we characterized the transcription factors (TFs) that regulate DEGs. The 1,165 DEGs over-expressed after exercise were strongly associated with the inflammatory response, intestinal permeability and regulation of the response to stress and bacterium (Fig. 3A). In contrast, the 1,288 under-expressed DEGs were related to macromolecule catabolism, cellular respiration, mitochondrial transport, and transcriptional and translational activity (Fig. 3B). The full list of significantly enriched pathways (false discovery rate (FDR) < 0.001) is given in Supplementary Table S6. The main putative regulatory TF in the network of DEGs was *ZFP42*, followed by the cooperatively transcriptional cofactors *SPI1*, *FOXO3*, *IRF3* and *NRF1* (Fig. 4). TFs act in regulatory networks and can drive or repress the expression of the miRNAs in a feed-forward and feedback manner²³. Accordingly, we found that these TFs might control the spatiotemporal expression patterns of some of the DE miRNAs (Fig. 4B).

The Supplementary Information section contains further details of (i) the characterization of transcription factors that regulate differentially expressed genes and (ii) the validation of mRNA and miRNA expression experiments using RT-qPCR.

Target DEGs were inversely correlated with functionally enriched miRNAs. We used two complementary statistical approaches (detailed in the Methods section) to identify the target-enriched miRNAs and assess their role in regulating the exercise response. Firstly, for each of the 167 DE miRNAs, we used the multiMiR package in R to build a comprehensive list of all the putative validated target genes. We found that only 91 DE miRNAs (70 of which were up-regulated and 21 of which were down-regulated in post-ride samples, compared with pre-ride samples) presented an experimentally annotated targetome. In total, these DE miRNAs regulated

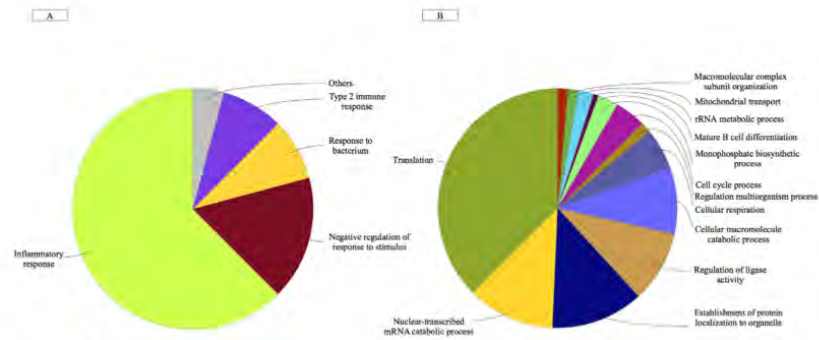


Figure 3. A functional map of DEGs, showing the top categories of GO biological processes associated with significantly over-expressed (A) and under-expressed (B) genes following endurance exercise. The chart fragments represent the number of genes associated with the various terms.

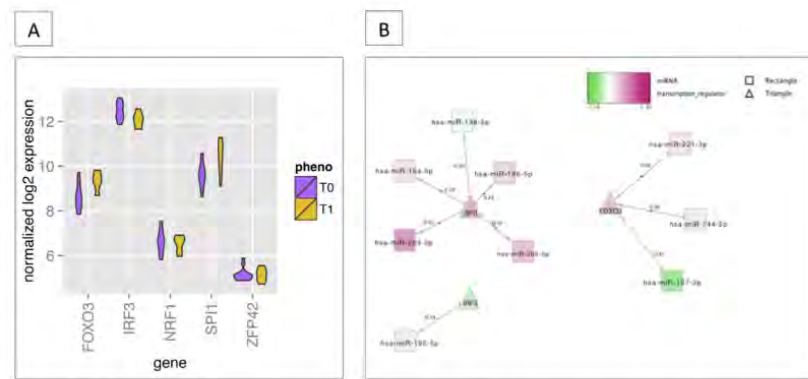


Figure 4. Relationship between TFs, miRNAs and DEGs. (A) The boxplot graph represents the expression levels (\log_2) of the main candidate TFs (discovered using iRegulon) within the set of DEGs. *ZFP42* was the top regulator, followed by *SPI1*, *FOXO3*, *IRF3* and *NRF1*. (B) A schematic view of the common correlations between TFs and miRNAs. The correlation score for expression of the two interactants is indicated under each edge. The network is displayed graphically as nodes (genes, TFs and miRNAs) and edges (biological relationships). The edge colour intensity indicates the expression level of the association: red = over-expression at T1 and green = under-expression at T1. The node shape indicates whether the node is a TF (triangles) or a miRNA (squares).

7,150 putative target genes—most of which were down-regulated after the ride (Supplementary Table S7). For each of the 91 DE miRNAs presenting at least one experimentally validated target, we looked at whether the DEGs that were significantly and negatively correlated with DE miRNAs were in fact miRNA targets. The hypergeometric test revealed two DE miRNAs (miR-138-5p and miR-26b-5p) with significantly more predicted targets among the negatively correlated DEGs than among other genes in the transcriptome ($FDR < 0.10$; Supplementary Table S8). Next, to address the hypergeometric test's lack of power (given that there were few anticorrelated target genes for a given DE miRNA), we applied a generalization of the hypergeometric test to each of the 91 DE miRNAs. The objective was to establish whether the experimentally validated target genes displayed smaller Limma p -values for differential expression (when comparing pre- and post-ride samples) than other genes in the transcriptome (according to a Wilcoxon sum rank test). This generalization of the hypergeometric test also encompassed targets affected by translation inhibition, i.e. targets for which changes in miRNA levels do not reduce mRNA levels ($r > 0.5$) but do reduce the protein levels²⁴. We observed 45 enriched DE miRNAs ($FDR < 0.1$; Supplementary Table S8) and then selected those for which the target genes were DEGs and were inversely correlated ($r^2 < -0.5$). This calculation removed the hsa-miR-374a-5p and hsa-miR-374b-5p from the list, since their targets were DEGs but were not greatly anticorrelated ($r > -0.5$).

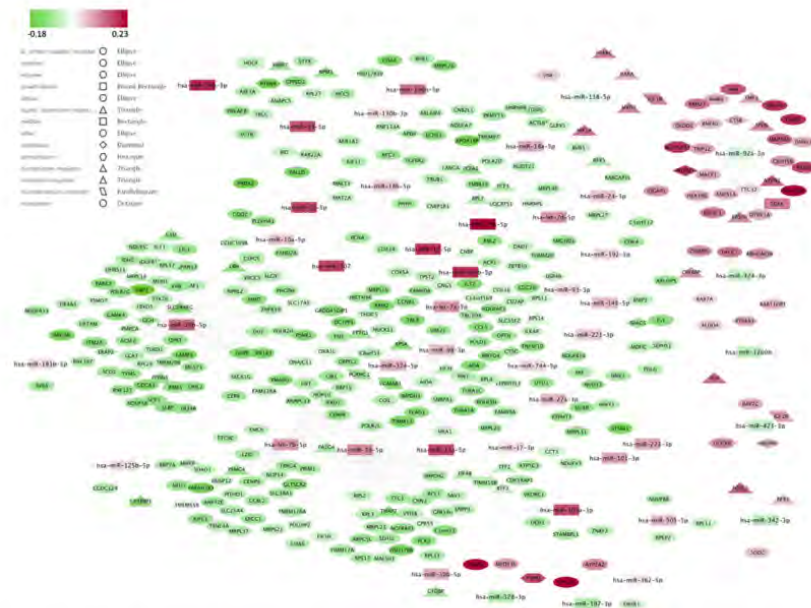


Figure 5. Regulatory network linking the 44 enriched miRNA and their respective inversely correlated target DEGs. We identified a total of 44 enriched DE miRNAs for which the miRNA targets defined by multi miR were significantly enriched (according to a hypergeometric test or its generalization). The 44 DE miRNAs were inversely correlated with a total of 351 unique target DEGs. The network is displayed graphically as nodes (genes and miRNAs) and edges (biological relationships). The edge colour intensity indicates the expression level (\log_2) of the association: red = over-expression at T1 and green = under-expression at T1. The node shape indicates whether the node is a transcription regulator (triangles), growth factor (round rectangles), peptidase (diamonds), phosphatase (hexagons), transmembrane receptor (parallelograms), transporter (octagons), a miRNA (squares) or other type of genes (ellipses).

Ultimately, we considered a total of 44 enriched miRNAs (Supplementary Fig. S2; Supplementary Table S9) inversely correlated with the expression of 351 target DEGs during exercise (Supplementary Table S10; Fig. 5). We next examined the protein-protein interaction (PPI) sub-network associated with the 44 enriched miRNAs and their inversely correlated target DEGs (see Supplementary Information Section).

To identify the most relevant cellular activities controlled by these 351 anticorrelated target DEGs, we analysed overrepresented GO biological process terms (using clueGO). The most significant GO terms were related to glucose metabolism, fatty acid oxidation, mitochondrial biogenesis through the peroxisome proliferator-activated receptor (PPAR γ) signalling pathway, oxidation stress, proteolysis and immune response pathways (Fig. 6). To provide further insight into proteolysis and immune response pathways, we also analysed the correlations between changes in the enriched DE miRNA levels and biochemical blood parameters (see Supplementary Information Section).

Furthermore, we found that the regulatory network based on the 351 target genes included 31 TFs (Supplementary Table S10). Our results suggest that the *EP300*, *RFX5* and *FOSL2* TFs regulated the expression of six different miRNAs within the regulatory network (Fig. 7). For example, the EP300 protein stimulated transcription of miR-92a, which in turn suppressed *EP300* expression. More sophisticated regulation was provided by dual negative-feedback loops, such as the one involving miR-138-5p and the TFs *FOSL2* and *EP300* (Fig. 7).

Validation of the enriched miRNAs and their inversely correlated target DEGs by applying graphical Gaussian models (GGMs) to a validation set. We next used an independent cohort of animals to validate the inversely correlated target-miRNA regulatory network by applying GGMs in a regression framework. This cohort included 16 horses sampled solely at T0 and 31 other horses sampled solely at T1. The models were built with $p = 44$ miRNAs and $j = 351$ inversely correlated DEGs targeted by at least one of these miRNAs. In order to increase the degree of precision, we specified a list of edges constrained to zero: only (i) interactions between miRNAs and their experimentally validated target genes, and (ii) gene-gene interactions predicted by PPI were considered (Supplementary Fig. S3). For each time point (T0 and T1) and each enriched DE miRNA, the coefficient of determination (R^2) and the mean squared error (MSE) were computed

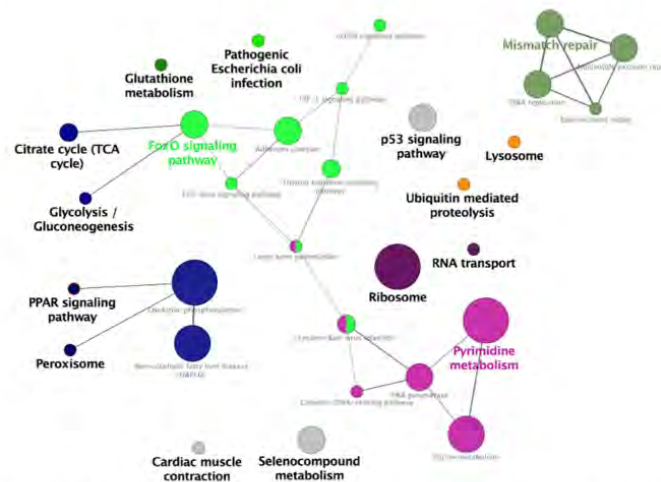


Figure 6. A functional map of the 351 inversely correlated target genes regulated by the 44 enriched miRNAs. Within the network, the GO biological terms were identified as nodes and then linked according to their kappa value (>0.4) and FDR (<0.001). The size of the node corresponds to the statistical significance of the enrichment term. Functionally related groups partially overlap. Similar GO terms are labelled in the same colour. Non-grouped terms are shown in grey. The colour gradient shows the proportion of genes in each cluster associated with the term.

and compared with the values obtained with 351 repeatedly randomly selected genes ($n = 500$ times). Moreover, for each criterion, the extent to which each DE miRNA was predicted by its inversely correlated target genes (as a p -value) was calculated as the proportion of repetitions with more statistically significant results than the original data. None of the enriched DE miRNAs had an adjusted p -value of below 0.05 in the FDR test (Supplementary Table S11). Although the adjusted p -values were not statistically significant, these analyses allowed us to rank the enriched DE miRNAs according to the strength of the direct interaction with their inversely correlated target genes (Fig. 8). At T0, the MSE and R^2 criteria were concordant, displaying p -values < 0.05 for miR-181b-5p and miR-505-5p. At T1, miR-21-5p displayed a p -value < 0.05 for both R^2 and MSE; this miRNA was not predicted by either R^2 or MSE at T0, suggesting that it might be involved in the regulation of exercise-related physiological processes (Fig. 8).

Discussion

Endurance exercise has a profound impact on metabolism in tissues other than skeletal muscle (including heart, brain, adipose tissue and liver)²⁵. The level of exercise performed by horses during an endurance competition is similar to that of a human marathon runner²⁶ or ultramarathon runner²⁷. By using microarrays to analyse the blood transcriptome and miRNome, our objective was to identify and characterize candidate miRNAs and the interplay that ultimately resulted in the coordinated response to endurance exercise.

Peripheral blood is now widely used as a surrogate tissue for monitoring whole-body status²⁸. Indeed, blood transcriptome analysis is undoubtedly capable of evidencing responses to exercise in the horse^{26,29,30}. By comparing the respective blood transcriptome profiles before and after exercise, we were able to identify a large number of DEGs that fit neatly into the well-characterized context of adaptive regulation to exercise in endurance horses (including energy metabolism, the inflammatory response, stress resistance, oxidative stress, cell death and proteolysis^{26,29,31}). These changes may help to supply the working muscles with energy or control excessive inflammatory reactions. They may also be involved in “staleness” and the transient immunosuppression that can occur during and/or after endurance exercise.

We found a total of 167 DE miRNAs when comparing T1 with T0. These included two mitomiRs (i.e. DE miRNAs encoded by the mtDNA), which might belong to the new family of miRNAs recently discovered in human vascular epithelium cells and matured from the mtDNA-transcribed long, non-coding RNA³². Endurance exercise is the most potent physiological inducer of mitochondrial biogenesis in skeletal muscle³³. We hypothesize that the increase in mitochondrial biogenesis and cell energy regulation in endurance exercise is boosted by these mitomiRs. Despite the relatively small number of animals studied, these findings are valuable.

In order to assess the main effects of DE miRNA regulation on exercise transcriptome patterns, we used the hypergeometric test and its generalization to search for DE miRNAs with significant target site enrichment within DEGs. In fact, 44 DE miRNAs of the 167 DE miRNAs had significantly more targets among DEGs than among the transcriptome as a whole. Twenty-one of these 44 enriched miRNAs were already known to be related to

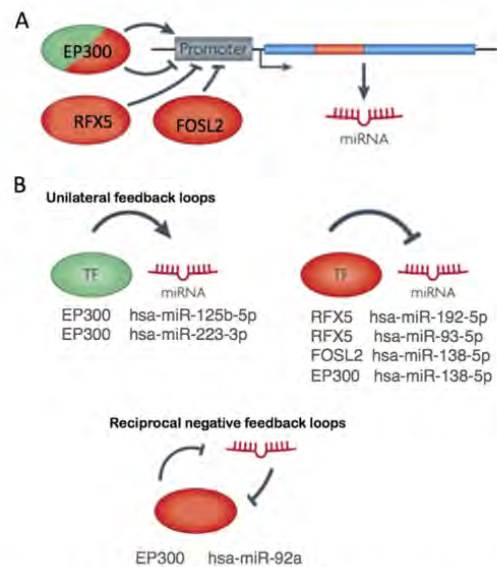


Figure 7. Activators and repressors of the 44 enriched miRNAs. When screening the negatively correlated target genes of the 44 enriched miRNAs, we found three TFs able to up- or downregulate miRNA expression (*EP300*, *RFX5* and *FOSL2*). The promoter regions of the expressed miRNAs are very similar to those of protein-coding genes. The presence of CpG islands, TATA box sequences, initiation elements and certain histone modifications indicate that the promoters of miRNA genes might be controlled by TFs. (A) *RFX5* and *FOSL2* both might inhibit expression of miR-192-5p, miR-93-5p and miR-138-5p in blood, although *EP300* might stimulate expression of miRNAs such as miR-125-5p and miR-223-3p. (B) Unilateral or reciprocal-negative feedback loops result in oscillatory or stable mutually exclusive expressions of the TFs and miRNA components²³. For instance, *EP300* might repress miR-92a expression. miR-92a targets the *EP300* TF and might block its expression. Figure adapted from Krol *et al.*²³.

exercise adaptation in humans, whereas all the others were identified for the first time in the present study. Hence, our results have added to the list of miRNAs expressed in blood following exercise. Future studies will have to determine whether this novel list of orthologous equine miRNAs expressed in blood is horse-specific or reflects a more general response to exercise in mammals (e.g. with counterparts in humans). More precisely, the observed differential expression of miR-15a, miR-16, miR-17, miR-18a, miR-20ab, miR-21-5p, miR-27a, miR-30b, miR-93, miR-101, miR-106, miR-107, miR-125b, miR-130, miR-138-5p, miR-145, miR-181ab-5p, miR-221, miR-223, miR-342-3p and miR-505 was consistent with previously reports on exhaustive exercise in humans^{14,15,20,21,34}. Many of the enriched DE miRNAs observed in our study are known to originate from blood cells and to regulate genes that are involved in immune processes, apoptosis and transcription regulation. For example, miR-21, miR-27a and miR-181 were found to be DE in the whole blood of highly trained human athletes after a 30-minute treadmill test²². It was then suggested that these miRNAs regulate physiological processes that are essential for the response to exercise (such as immune function, apoptosis, membrane traffic of proteins and transcriptional regulation)²². Furthermore, miR-17, miR-18a, miR-20a (all members of the miR-17-92 cluster), miR-106 and miR-93 (paralogs of the miR-17-92 cluster) were found to be DE in circulating neutrophils of 11 men after a brief bout of exercise; the miRNAs regulated the ubiquitin-mediated proteolysis pathway²¹. Lastly, we observed differential expression of miR-181, which is known to be affected in different types of leukocyte³⁵ and likely to be involved in the regulation of some fundamental adaptive changes in the immune system during endurance exercise³⁴. Similarly, Makarova *et al.*³⁴ suggested that the exercise-induced expression of miRNA-181 in blood cells may be a compensatory, anti-inflammatory adaptation to the primary, systemic inflammatory response caused by exercise.

The ability to exercise for extended periods requires not only adequate endocrine and immune machinery but also energy sources³. Studies on the role of miRNAs in these adaptive processes have just started. Some preliminary results in humans suggest that the post-exercise increase in miR-107 expression regulates glucose homeostasis in liver and adipose tissue¹⁵, whereas in the horse the increased expression of miR-17 after exercise might be linked to the regulation of glucose metabolism in different tissues³⁰. Investigation of the adaptive significance of each individual miRNA expressed in blood is still in its infancy, and the function of many miRNAs remains unknown.

The earliest studies in this field also have demonstrated that endurance exercise-induced changes in plasma/serum levels of circulating miRNAs (c-miRNAs)^{13,15-19}. In most reports, plasma levels of human miR-1³⁶,

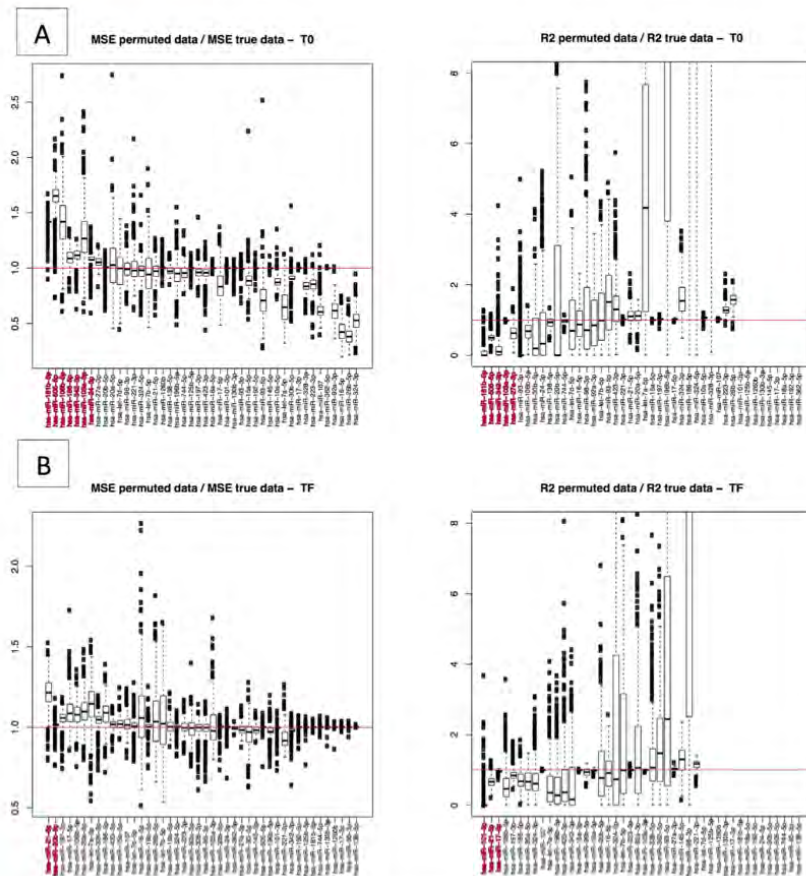


Figure 8. Validation of the 44 enriched miRNAs by applying GGMs to an independent validation set. For each selected enriched miRNA, MSE and R^2 were computed through a GGM implemented on the 44 enriched miRNAs and their 351 inversely correlated target genes. (A) The upper row of boxplots shows the ratio between MSE values for the randomly selected genes from among the 24,415 genes and candidate gene set at T0, and the values of the R^2 for predicted vs. observed values at T0, respectively. (B) The lower row of boxplots shows the same boxplots at T1. In all cases, a total of 500 repetitions were performed with 351 genes randomly selected from among the 24,415 genes for each miRNA. For both criteria, we computed permutation p -values, i.e. the proportion of repetitions leading to a smaller MSE (or a larger R^2) with respect to the values found with the originally selected 351 genes. MicroRNAs with $p < 0.10$ are shown in red.

miR-20^{13,18}, miR-21^{13,18}, miR-103¹⁵, miR-107¹⁵, miR-126¹⁹, miRNA-133^{19,36}, miR-146a^{13,18}, miR-206³⁶, miR-221^{13,18} and miR-222^{13,18} were found to be higher after endurance exercise – suggesting a possible role in exercise adaptations. Similarly, we observed an enrichment of miR-20ab, miR-21, miR-103a-3p, miR-107 and miR-221 when comparing pre- and post-ride blood samples. However, it is probable that most of the miRNAs in the present study came from blood cells, since the microarray-based estimation of circulating miRNA abundance from total blood RNA levels is particularly difficult³⁷. In the present study, the c-miRNAs may have been partly masked (or at least diluted) by greater amounts of cellular miRNAs³⁸. Furthermore, we only considered miRNAs that displayed a signal intensity greater than the mean + 1.5 SD for the negative control; as a result, we probably missed some potentially detectable c-miRNAs. Baggish *et al.*¹³ suggested that human miR-20a, miR-21 and miR-221 (i) are released after exercise into the bloodstream by tissues other than blood cells and (ii) may regulate key pathways in angiogenesis, inflammation, muscle contractibility and adaptation to hypoxia. However, we believe that blood cells may have significantly contributed to our present results. Accordingly, levels of miR-20a, miR-21 and miR-221 were not correlated with changes in plasma markers of muscle inflammation and liver damage. In

contrast, however, we found some significant, positive correlations between miR-133 levels and changes in plasma creatine kinase and aspartate transaminase levels (which reportedly reflect cell damage in human athletes after a marathon race¹⁹). In view of this result, the relationship between blood cell levels and plasma/serum levels of miR-133 also warrant further examination.

A further observation of interest relates to the fact that the 44 enriched miRNAs regulated a total of 351 inversely correlated target DEGs. All target genes were involved in processes highly relevant to the response to exercise, including immune function, apoptosis, protein degradation, transcription regulation, and mitochondrial biogenesis and energy metabolism. We hypothesized that these 351 target genes were driven not only by the miRNAs within the regulatory network but also by some of the TFs. As further confirmation of this hypothesis, we found that the regulatory network based on the 351 target genes and 44 enriched miRNAs featured a high proportion of TFs ($n = 31$) – including *PPAR γ* , one of the master regulators of mitochondrial biogenesis and energy metabolism³⁹. Furthermore, the assembly of miRNAs and TF co-regulators within the regulatory network revealed several interesting feed-forward and feedback loops.

After we had examined the structural and functional aspects of the co-regulatory networks, we further confirmed the regulatory effects of the 44 enriched miRNAs (identified by computational predictions on an experimental set of 14 animals) using an independent cohort of animals and a GMM approach. Our data confirmed a seldom-measured relationship between target DEGs and their miRNA regulators. Interestingly, we confirmed the post-exercise expression of miR-21-5p and their target DEGs. The expression of miR-21-5p is known to be stress-responsive; this miRNA has an important role in heart failure⁴⁰, renal ischemia reperfusion injury⁴¹, and in the self-protective anti-inflammatory reaction to exercise²². Aforesaid, the up-regulation of circulating miR-21-5p in humans was reported to occur in plasma of endurance athletes^{13–18} and in circulating PBMCs²⁰ and blood cells²² upon exercise. At baseline (T0), we confirmed the regulatory roles of miR-181a-5p, miR-505-5p and their target DEGs. As mentioned above, miR-181 in PBMCs has already been associated with a T-cell response to exercise in humans²⁰, and Makarova *et al.*³⁴ have suggested that miR-181 has several key roles in the adaptation to exercise. In contrast, miR-505-5p has not previously been linked to the regulation of exercise.

Our study had several limitations. Firstly, our main analysis was based on human orthologous miRNAs, for which we had a comprehensive list of validated miRNA–target interactions. We therefore had to ignore the predicted miRNA–target interactions that were specific to the horse. Moreover, the identification of putative equine miRNAs using sRNAseq was performed using samples derived from six tissues other than blood and with small number of biological replicates; this may have prevented us from detecting certain blood-specific or exercise-induced miRNAs. Secondly, there are several modes of gene regulation other than down-regulation by miRNAs: these include up-regulation by miRNAs, DNA methylation and chromatin remodelling—all of which may be involved in the responses to exercise. Furthermore, chemical perturbations (such as changes in pH and local temperature), systemic increases in cytokine and growth factor levels, and other factors such as the exercise dose, gender, age, genetic background and stochastic factors can all potentially modify transcriptome expression. Therefore, future research will have to focus on how these various regulatory interactions are synchronized to maintain homeostasis during and following endurance exercise in horses. Thirdly, we analysed miRNAs from whole blood. Thus, it remains to be determined whether the individual components of this heterogeneous fluid (i.e. plasma, platelets, erythrocytes, nucleated blood cells and exosomes) reflect the overall microRNA response to exercise in horses. With regard to the requirement for highly specific c-miRNA biomarkers of metabolic demand, muscle damage, myocardial injury or endotoxaemia during equine endurance events, it may make sense to focus on c-miRNAs not expressed by blood cells. However, it must be borne in mind that intravascular haemolysis occurs very frequently during and after intense exercise in horses^{42,43}, which can significantly affect levels of circulating miRNAs.

Despite these limitations, our results extend earlier observations on gene expression in equine athletes^{26,29}, and we showed that the main changes in the blood transcriptome appear to be driven by changes in the expression of *trans*-acting regulators in general and 44 enriched DE miRNAs in particular. Characterization of these miRNAs that lead to endurance-exercise-mediated systemic adaptation might also lead to the development of novel nutritional, pharmacological, and exercise-based training interventions. The goal would be to fine-tune levels of miRNA expression and thus improve the athlete's energy metabolism and inflammatory response. Moreover, our results also demonstrated the sensitivity and specificity of whole-blood derived mRNA and miRNAs as markers for the response to endurance exercise. In view of the availability of whole blood sample tubes that protect the RNA content and the minimal handling required during sample collection, the use of whole blood has some clear advantages for monitoring the post-exercise expression of miRNAs under field conditions. Lastly, the specialized nature of equine musculoskeletal tissues (effective high-speed locomotion and strength), the large size of both human and equine species, the good degree of access to tissue specimens, and the high overall degree of similarity between equine and human blood transcriptome responses to exercise make the horse a good natural model for basic exercise research.

Conclusions

The microRNA and gene co-regulatory network profiled in blood reflects the physiological regulation of the transcriptome following endurance exercise in the horse. Our present results suggest that 44 enriched miRNAs and their 351 inversely correlated target genes regulate lipid metabolism, carbohydrate metabolism, mitochondrial biogenesis, proteolysis and the immune response. By using various computational methods and an independent cohort of animals, we confirmed that miR-21-5p, miR-181b-5p and miR-505-5p are putative regulators of the response to endurance exercise. Our results also suggest that (i) a variety of mechanisms underlie adaptive changes to the transcriptome and (ii) changes in the expression of a few key regulators may improve the metabolic and inflammatory response to exercise in athletes.

Methods

Animals. Sixty-one pure-breed or half-breed Arabian horses (20 females and 41 geldings; mean \pm SEM age: 10 ± 2) competing in three 160 km endurance competitions were recruited with the owner's agreement. The weather conditions, terrain difficulty and altitude for the three endurance competitions were similar. The 61 horses were checked carefully for the population-genetic structure and covariates such as age, gender, and environmental influences (disease states, medical treatment, etc.). To ensure sample homogeneity, the participating horses were subject to the same management practices throughout the ride and passed the International Equestrian Federation (FEI)'s compulsory examination before the start. Animals were fed 2–3 hours before the start of the endurance competition with *ad libitum* hay and 1 kg of concentrate pellets. During the endurance competition, all the animals underwent veterinary checks every 20 to 40 km, followed by recovery periods of 40 to 50 minutes (in accordance with the FEI rules on endurance riding). After each vet gate check, animals were provided with *ad libitum* water and hay and a small amount of concentrate pellets. The median winning speed over the entire ride was 15.7 ± 1.04 km/h. Information on training, nutrition and medical examinations was obtained but it was impossible to control for the feed composition and levels of consumption. The study design was approved by the independent ethics committee at Alfort Veterinary School and the University of Paris Est (reference: 12/07/11-1). All methods were carried out in accordance with the approved guidelines. In all cases, the owner provided his/her informed consent prior to the initiation of study procedures with the animals. Further details have been published by Le Moyec *et al.*⁴⁴.

The horses were divided into two non-overlapping sets: the experimental set and the validation set. The validation set came from an independent cohort of animals, to ensure that the observed effects were reproducible in a broader context. The experimental set included 14 horses sampled before the 160 km ride (T0, baseline) and then again afterwards (T1). Samples were collected within 30 minutes of the end of the ride. Two horses in the experimental set failed a vet gate check (poor metabolic condition after 80 km and lameness at the end of the ride, respectively).

The validation set included 16 horses sampled solely at T0 and 31 other horses sampled solely at T1. Of the horses sampled at T1, four were eliminated for lameness after 106 km ($n = 1$), 133 km ($n = 2$) or 160 km ($n = 1$), two were eliminated after 30 km for metabolic disorders and one was eliminated for tiredness after 130 km.

Blood biochemical assays. Blood samples were collected in dry tubes at the end of the endurance event. After clotting, the tubes were centrifuged and the harvested serum was stored at 4 °C until analysis (no more than 48 later, in all cases). Sera were assayed for total bilirubin, conjugated bilirubin, total protein, creatinine, creatine kinase, aspartate transaminase, gamma glutamyltransferase and serum amyloid A levels on a RX Imola analyser (Randox, UK). Blood collected in EDTA tubes was used to measure the packed cell volume after centrifugation.

Blood samples, RNA isolation and microarray experiments. A single blood sample for both transcriptome and miRNome profiling was obtained in order to study direct relationships between mRNA expression and miRNA expression in each animal. Whole blood samples from each horse were collected in PAXgene Blood RNA tubes (Qiagen, Germany) and the total RNA was isolated using a PAXgene Blood RNA Kit (Qiagen), according to the manufacturer's instructions.

The RNA's purity and concentration were determined using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, USA) and RNA integrity was assessed using the Bioanalyzer 2100 (Agilent Technologies, USA). Each of the 61 RNA samples was then divided into two aliquots for use in the miRNA microarray or gene expression microarray experiments, respectively.

The transcriptome was profiled with a custom equine 4×44 K microarray (ID: 044466, Agilent Technologies). Cyanine-3 (Cy3)-labelled cRNAs were prepared from 100 ng of total RNA using the One-Color Low Input Quick Amp Labeling kit (Agilent Technologies). Specific activities and cRNA yields were measured using the NanoDrop ND-1000 (Thermo Scientific, USA). For each sample, 1.65 μ g of Cy3-labeled cRNA (specific activity > 9.0 pmol Cy3/ μ g of cRNA) were fragmented at 60 °C for 30 minutes in a volume of 55 μ l containing 25 \times Agilent Fragmentation Buffer and 10 \times Agilent Blocking Agent. Subsequently, 55 μ l of 2 \times Agilent Hybridization Buffer were added to the fragmentation mixture and hybridized to the array, in accordance with the recommended protocol. After hybridization, the microarrays were washed 1 minute at room temperature with GE Wash Buffer 1 (Agilent Technologies) and 1 minute at 37 °C using GE Wash Buffer 2 (Agilent Technologies).

The slides were then scanned using a G2565CA Scanner System (Agilent Technologies), using a scan protocol with a resolution of 3 μ m and a dynamic range of 20 bits. The resulting TIFF images were analysed with the Feature Extraction Software v10.7.3.1 (Agilent Technologies), using the GE1_107_Sep09 protocol. The microarray data can be obtained from the Gene Expression Omnibus (GEO) database⁴⁵ with the accession number GSE72973 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72973>).

For the miRNome profiling, an Agilent custom equine miRNA 8×60 K microarray was used (ID: 060464). For each sample ($n = 61$), a total amount of 100 ng of total RNA was dephosphorylated, 3' end-labelled with Cy3-pCp, purified on Micro Bio-Spin 6 columns (Bio-Rad, USA), dried, and hybridized to the arrays using the miRNA Complete Labeling and Hybridization Kit (Agilent Technologies) according to the manufacturer's instructions. After a washing step, hybridized microarray slides were scanned using a G2565CA Scanner System (Agilent Technologies) and a scan protocol with a resolution of 3 μ m and a dynamic range of 20 bits. The resulting TIFF images were analysed with the Feature Extraction Software v10.7.3.1 (Agilent Technologies), using the miRNA_107_Sep09 protocol. The microarray data are available at GEO (GSE73102: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73102>). The whole reference series (mRNA and miRNA) is available at GEO (GSE73104: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73104>).

Microarray dataset was deposited in GEO, in accordance with the MIAME guidelines.

Description of the Agilent custom equine miRNA 8 × 60 K microarray. The custom equine array contained 1,886 *Homo sapiens* (*hsa*) miRNA sequences from the miRBase database (V9.2). *Hsa* miRNAs were included in the custom equine microarray because Buza *et al.*⁴⁶ reported that over 60% of known mature equine miRNAs had perfect matches with human-associated miRNAs. Furthermore, the microarray included 84 known equine (*eca*) miRNAs (with no known human homologs), 876 putative novel miRNAs and 20 putative small RNAs encoded by the mitochondrial DNA. The putative novel sequences were identified in an independent sRNAseq experiment during which we used 16 small RNA libraries from six different equine tissues (heart, liver, cartilage and three muscles: the platysma, the *gluteus medius* and the masseter). miRNA from the platysma and masseter muscles, heart and liver were extracted from 15 healthy horses of different ages and breeds (collected at the slaughterhouse). MicroRNAs from the *gluteus medius* were extracted from 10 healthy and 8 Cob Normand horses affected by polysaccharide storage myopathy. Total RNA was purified from tissues using the miRNeasy Mini kit (Qiagen) and the RNeasy MinElute Cleanup kit (Qiagen) for enrichment of small RNA (<200 bp), according to the manufacturer's protocol. Total RNA quality and concentration were assessed as described above. For each tissue, three pools of three horses were formed (i.e. yielding 16 libraries). Approximately 500 ng of total RNA were used for library constructions, according to the manufacturer's protocol for the SOLiD[®] Total RNA-Seq Kit (Applied Biosystems, USA). The small RNA libraries were sequenced on a SOLiD 5500XL Series Genetic Analysis System at the Metaquant core facility (INRA, Jouy-en-Josas, France). Solid sequencing generated a total of 350 million reads of ~50 bp. The overall flow of the sequencing and bioinformatics analysis for small RNA has been described previously⁴⁷. In brief, after removal of low-quality reads, the flanking linker and primer sequences, all remaining small RNAs were aligned with the *Equus caballus* genome (Equ Cab 2; GCA_000002305.1) with BowTie⁴⁸ (v0.12.7). Fewer than 2 mismatches were required. Reads with >6 alignments were removed. Mapped reads overlapping with known non-coding RNAs obtained from RFAM (<http://rfam.sanger.ac.uk/>) were also discarded. To detect putative miRNAs within the small RNAs, reads were filtered by length (17–27 bp). Sequences that mapped solely to the mitochondrion were designed as mitomiRs. Subsequently, all potential novel miRNAs and mitomiRs were sorted and assigned to the putative ncRNA locus (i.e. clusters). Clusters with fewer than 10 reads were discarded because of their low information content. Potential pre-miRNAs were excised. Excision was initiated when a stack of reads was encountered. If there was a higher read stack within 20 bp downstream of the current read stack, the former stack was selected. Hence, the highest local read stack was identified. Next, sequences covered by the highest local read stack were excised twice: once including the 50 bp upstream and 10 bp downstream flanking sequences, and once including 10 bp upstream and 50 bp downstream flanking sequences (to simulate the 5p and 3p positions of the potential miRNA). These putative pre-miRNAs were then screened for microRNA-like hairpin structures with RNAfold (using the latter's default settings). For further details on sequencing analysis and prediction of novel miRNAs, see Desjardins *et al.*⁴⁷. Our analysis yielded a total of 876 novel miRNAs and 20 putative mitomiRs (Supplementary Table S12). Three novel miRNAs were re-annotated as *eca*-miRNAs because they have been already included in the new update of the miRNA database (with new-*eca*-mir-824 updated to *eca*-miR-1388; new-*eca*-mir-1047 to *eca*-miR-676; and new-*eca*-mir-1072 to *eca*-miR-2114). Among the plausible mitomiRs, two mapped to the 16S rRNA, nine mapped to various tRNAs, two mapped to genes for complex I subunits, two mapped to genes from complex IV subunits, and two mapped to genes from complex V subunits. Lastly, two potential mitoRNAs mapped to the D-loop locus (Supplementary Fig. S4). None of the predicted pre-miRNA sequences encoded by the mitochondrial genome were significantly folded into a duplex structure.

Statistical analysis of the microarray results. All the transcriptome pre-processing, normalization and statistical analysis steps were carried out using several Bioconductor packages in R programming language (version 3.02). Firstly, the data's quality was checked with the arrayQualityMetrics package⁴⁹. Subsequently, probes showing any of the following Agilent flags (*glsFeatPopnOL* = 1, *glsBGNNonUnifOL* = 1, *glsFeatNonUnifOL* = 1) were removed from the analysis. The raw intensity values were then background-corrected using the "normexp" function, and the expression data were quantile-normalized (using the Limma package⁵⁰; version 3.14.4). Probe summarization was based on the median expression value for the replicated probes. A total of 24,415 genes were selected in this analysis. Given that many of the original annotations for the custom equine Agilent microarray (AMAMID 021322, enriched with 384 equine transcripts) have been found to be incomplete, the annotation available in June 2014 at (<http://www.genomics.agilent.com/en/Custom-Design-Tools/eArray/?cid=AG-PT-122&tabId=AG-PR-1047>) was used to re-assign the probes to new probe sets.

PCA was performed with FactoMineR and Ade4 packages (version 1.23), in order to establish whether a particular array contributed markedly to variability in the gene expression data (i.e. whether it retained most of the information). Differences revealed by the PCA were assessed using the Monte Carlo Permutation Procedure (999 replicates; "randtest" function) in the Ade4 package in R. DEGs were identified by using the Limma package. The statistical linear model included the time point (T0/T1) as a fixed effect. To make the analysis more robust and control more strictly for the false discovery rate (FDR), the *p*-values were corrected for multiple testing using Bonferroni's method with a threshold of adjusted *p* < 0.05⁵¹. An unsupervised analysis of DEGs was carried out to identify clusters of samples or genes on the basis of their variance-covariance structure. Thus, a two-way hierarchical cluster analysis was performed using the *hclust* function with "1-cor (x)" as the distance and different aggregation criteria. The *heatmap* function was used to generate images.

For miRNome profiling, the raw data was pre-processed using a variant of the robust multi-array average (RMA) algorithm that has been specifically implemented for Agilent miRNA microarrays by Lopez-Romero *et al.*⁵². The RMA implementation in the *AgiMicroRna* package was used without the initial background correction step, as recommended⁵². A miRNA transcript was considered to be detectable only if it met the following two conditions: (i) expression in at least 33% of the experimental samples; (ii) a signal intensity above the mean value of the negative control +1.5 standard deviations. DE miRNAs were identified using the Limma package. In

this case, correction for multiple testing (FDR < 0.05) was performed using the Benjamini-Hochberg method (as a compromise between the unadjusted analysis and the Bonferroni procedures). Multivariate analysis was performed as explained above.

Generation of a functional annotation map from a list of genes. The GO term and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) enrichment analyses of DEGs and target DEGs were performed using Cytoscape V2.7 (<http://cytoscape.org/>) with the ClueGo V1.3 plug-in⁵³. ClueGO determines the distribution of the gene list for the various GO terms and pathways. The *p*-value was calculated using right-sided hypergeometric tests and the Benjamini-Hochberg correction for multiple testing (FDR < 0.001). Together with this very stringent FDR threshold, a high kappa value (0.4) enabled us to precisely select GO terms enriched in highly connected genes. The size of the nodes reflected the term's degree of enrichment. The network was automatically laid out using the organic layout algorithm in Cytoscape. Functional groups were created by iterative merging of the initially defined groups, according to the predefined kappa threshold. Only functional groups represented by their most significant term were visualized in the network⁵³. As a complementary approach, the list of selected genes was also fed into an Ingenuity Pathway Analysis (IPA; version 5.5, Ingenuity Systems, USA) to identify relevant categories of molecular functions, cellular components and biological processes. The IPA enabled us to identify (i) significantly overrepresented functional GO annotations, (ii) their over- or under-expression, and (iii) group-specific transcriptional networks. All listed or reconstructed cellular pathways were derived from the expert-annotated Ingenuity Knowledge Base of over 10⁶ PPIs. The IPA output a statistical assessment (based on Fisher's exact test) of the significance of representation for biological functions and signalling pathways. The IPA computed networks and ranked them according to a statistical likelihood approach.

Use of iRegulon to identify TFs involved in biological processes. The iRegulon computational method was used to identify TFs within the set of DEGs and target DEGs⁵⁴. iRegulon searches the regulatory sequences around each gene in order to detect enriched TF motifs. It uses a database of nearly 10,000 TF motifs. iRegulon links enriched motifs to candidate TFs and determines the optimal subset of direct target genes.

DEmiRNAs and their inversely correlated target DEGs: enrichment analysis. For each DEmiRNA, we assembled a comprehensive list of all experimentally validated target genes (using the multiMiR package in R). multiMiR is a comprehensive collection of predicted and validated miRNA-target interactions and their associations with diseases and drugs⁵⁵. It contains human and murine data from 14 external databases that are categorized into three components, including three validated miRNA-target databases (miRecords, miRTar-Base and TarBase), eight predicted miRNA-target databases (DIANA-microT, EIMMO, MicroCosm, miRanda, miRDB, PicTar, PITA and TargetScan), and three disease- or drug-related miRNA databases (miR2Disease, Pharmaco-miR and PhenomiR)⁵⁵.

To investigate the biological relevance of the DEmiRNAs, we used two complementary statistical approaches. For each of the DEmiRNAs presenting at least one target gene, we performed: (i) the hypergeometric test for enrichment in targets DEGs negatively correlated with the DEmiRNA; and (ii) the generalization of the hypergeometric test. Both methods were implemented on the experimental set. For the hypergeometric test, we performed a pair-wise Pearson correlation analysis of the expression levels of DEGs and the expression levels of DEmiRNAs presenting at least one experimentally validated target. Subsequently, we subtracted the set of correlated DEmiRNAs-DEGs with $r < -0.5$ and $p < 0.05$. One should note the existence of two types of negative regulatory interactions: (i) over-expressed mRNA/under-expressed miRNA and (ii) under-expressed mRNA/over-expressed miRNA. We next looked at whether the DEmiRNA targets were over-represented among the subtracted set of genes, when compared with the other genes in the transcriptome. Benjamini and Hochberg correction⁵⁶ for multiple testing was applied to the 91 *p*-values obtained.

Moreover, in order to address the hypergeometric test's lack of power when the number of inversely correlated target genes for a given DEmiRNA was very small, we also implemented a variant of the hypergeometric test. In contrast to the hypergeometric test (which is focused on the number of genes that were both DEGs and inversely correlated with a DEmiRNA at fixed significance thresholds), the generalization of the hypergeometric test considers all possible thresholds. In this variant, we tested for enrichment in target genes by selecting DEGs regardless of the sign of their correlation with the miRNA. When considering a given DEmiRNA with at least one experimentally validated target gene, we looked at whether their target genes presented smaller *p*-values than the other genes in the transcriptome using a one-sided Wilcoxon rank sum test (implemented with the "wilcox.test()" function in R). The FDR was determined to correct for multiple testing. Only the enriched DEmiRNAs for which at least one target gene was inversely correlated ($r < -0.5$) were selected. Lastly, only those DEmiRNAs that were significant either in the hypergeometric test or its generalization after correction for multiple testing (FDR < 0.1) were analyzed further. The FDR was set to 0.1 in order to retain as many biological functions as possible.

To find possible associations between the enriched DEmiRNA expression and biochemical blood parameters, a Pearson and a non-parametric Spearman rank correlation were applied using the `cor()` function in R.

Construction of a PPI sub-network on the basis of miRNA-mRNA interactions. The latest available version of the human PPI datasets from BioGRID (<http://thebiogrid.org/>) (release 3.2.105) was used for our analysis. The curated PPI data (containing 15,352 unique proteins and 281,862 interactions) constituted the parental PPI network. The Cytoscape program was used to visualize and analyze PPI networks. Firstly, the BioGRID parent PPI network was imported into Cytoscape. The target DEGs (human orthologous gene symbols) were listed in a text file (down regulated and up regulated DEGs, respectively) and mapped to the parental PPI network using the command "Select R Nodes R From ID List File". To confine the analysis to interactions close

to the target DEGs and to achieve maximum significance, only first-level interactions between target DEGs and their neighbours were detected.

Network topology analyses. The networks' topology was analysed using the NetworkAnalyzer Cytoscape plugin³⁷. NetworkAnalyzer provides insights into the organization and structure of complex networks formed by the interacting molecules. It computes a comprehensive set of topological parameters, such as network diameter, density, centralization, heterogeneity, and clustering coefficient, neighbourhood connectivity, average clustering coefficients and the distribution of node degrees³⁷. The degree of a node corresponds to the number of its directly connected neighbours. Briefly, the distribution of node degree $P(k)$ is defined as the number of nodes with a degree k of 0, 1, 2, and so on. The dependency pattern can be visualized by fitting a line to the node degree data. NetworkAnalyzer calculates the positive coordinate value for fitting the line with the power law curve of the form $y = \beta x^a$. The R^2 value is a statistical measure of the linearity of the curve fit and used to quantify the fit to the power line. When the fit is good, R^2 is very close to 1.

Use of GGMs to validate DE miRNA-gene association networks in a validation set. GGMs can infer direct relationships between variables within a set of repeated observations and in the absence or presence of *a priori* knowledge⁵⁸. In the GGM method, networks are represented as undirected graphs. Each node represents a gene or a miRNA, and an edge connects two nodes if they are partially correlated. In contrast to correlation analyses (which measure both direct and indirect interactions between pairs of variables), partial correlation analyses measure the strength of direct interaction only⁵⁸. A direct relationship between two variables corresponds to a non-zero entry in the partial correlation matrix. In a large-dimension context in which the number of variables exceeds the sample size, regularisation methods are needed for the estimation procedure. Moreover, in order to increase the estimation's precision, some interactions between variables may be neglected by setting some edges to zero. We considered animals at T0 and animals at T1 separately. For each scenario, GGMs were built to model interactions between enriched miRNAs and DEGs that were inversely correlated and were targeted by at least one of the enriched miRNAs. In the correlation matrix, the following edges were set to zero: (i) miRNA-miRNA, (ii) miRNA-mRNA when the mRNA was not a validated target of the miRNA, and (iii) mRNA-mRNA when the interaction was not found in the BioGRID human PPI dataset (Supplementary Fig. S3). The analyses were carried out with the R package *glasso*, which uses a lasso penalty for GGM estimation and allows a specified list of edges to be set to zero. The value of the constant in the lasso penalty (which has to be specified in the *glasso* package) was set to 0.03 by minimising the cross-validation MSE described below.

Furthermore, GGMs can be viewed as regression models of each variable with regard to all the other variables; regression coefficients are then expressed as a function of the partial correlation matrix. The regression coefficient of a variable i with regard to a variable j is equal to zero if and only if there is an edge between nodes i and j ⁵⁹. The GGM method represents a refinement of simple linear regression models of genes with regard to each miRNA because it takes account of mRNA-mRNA correlations. We made use of this property to build a cross-validation procedure and thus estimate the predictive ability of each miRNA from the list of target mRNAs. We considered two criteria: (i) the cross-validation MSE, which increases with the precision of the prediction, and (ii) the R^2 for predicted vs observed values, which decreases as the linear correlation between predicted and observed values increases. For each miRNA, MSE and R^2 were computed on the network composed by the selected enriched miRNAs and their inversely correlated target genes. These MSE and R^2 values were then compared with those obtained using the same procedure and the same number of genes randomly selected from the transcriptome. More precisely, for 500 repetitions, the genes were randomly selected from among a total of 24,415. Lastly, we computed permutation p -values, i.e. the proportion of repetitions leading to a smaller MSE (or a larger R^2) with respect to the values observed with the originally selected genes.

RT-qPCR validation of the transcriptome and miRNome results. To validate the technical robustness of the transcriptome data generated in the microarray study, quantitative RT-qPCR was carried out on a subset of candidate genes (Supplementary Table S13). We selected (i) genes with significant differences in expression levels when comparing T1 and T0 (a dynamic range of at least \log_2 (ratio) > 0.485), and (ii) genes of biological interest (e.g. *IRF3*, *CASB5*). Specific PCR primers for these genes were designed using ABI Primer Express software (version 3.0.1, Applied Biosystems) with melting temperatures of 58 °C to 60 °C and product lengths of between 100 and 150 bp. Reverse transcription of 500 ng of isolated total RNA was performed using the SuperScript[®] VILLO[™] cDNA Synthesis Kit (Invitrogen, France), according to the manufacturer's protocol. Dilutions (1/100) were used for qPCR with ABsolute Blue qPCR SYBR Green ROX mix (ThermoFisher Scientific, USA) in an Eppendorf Mastercycler RealPlex 4 (Eppendorf, France). The diluted cDNA samples were mixed with 1 × SYBR Green Master Mix and the specific reverse and forward primers (final concentration: 300 nM) in a final volume of 25 µl. Cycling conditions were 95 °C for 15 min, then 40 cycles of 95 °C for 15 sec and 60 °C for 1 min. For each sample and each gene, qPCR runs were performed in triplicate (in accordance with the manufacturer's protocol). In order to quantify and normalise the expression data, we used the $\Delta\Delta C_t$ method and the geometric mean C_t value from the succinate dehydrogenase complex, subunit A (*SDHA*), and beta actin (*ACTB*) as the endogenous reference genes. The set of genes chosen for confirmation by RT-qPCR was analysed in 10 animals of the experimental set using a linear effect model, including group (T0 or T1) as a fixed effect. The threshold for statistical significance was set to $p < 0.05$.

For analysis of the miRNome, five candidate miRNAs were selected from the microarray dataset for validation in the miRCURY LNA[™] Universal RT microRNA PCR system (Exiqon, Denmark), using the same RNA samples used in the microarray profiling. cDNAs were obtained with the Universal cDNA synthesis kit II (product #203301, Exiqon, Denmark). The primer sets and their product numbers are given in Supplementary Table S14. The qPCR reaction was performed according to the kit manufacturer's instructions (ExiLENT SYBR[®] Green

master mix, product #203403, Exiqon). The hsa-miR191 (#204306, Exiqon, Denmark)) was used as endogenous control. As for mRNA, the set of miRNAs chosen for confirmation by RT-qPCR was analysed in 10 animals of the experimental set using a linear effect model, including group (T0 or T1) as a fixed effect. The threshold for statistical significance was set to $p < 0.05$.

References

- Russell, A. P. *et al.* Regulation of miRNAs in human skeletal muscle following acute endurance exercise and short-term endurance training. *J. Physiol.* **591**, 4637–4653 (2013).
- Munoz, A., Riber, C., Trigo, P., Castejon-Riber, C. & Castejon, F. M. Dehydration, electrolyte imbalances and renin-angiotensin-aldosterone-vasopressin axis in successful and unsuccessful endurance horses. *Equine Vet. J.* **42**, 83–90 (2010).
- Snow, D. H., Baxter, P. & Rose, R. J. Muscle fibre composition and glycogen depletion in horses competing in an endurance ride. *Vet. record* **108**, 374–378 (1981).
- Davies, K. J., Packer, L. & Brooks, G. A. Biochemical adaptation of mitochondria, muscle, and whole-animal respiration to endurance training. *Arch. biochem. biophys.* **209**, 539–554 (1981).
- Hoppeler, H., Klossner, S. & Fluck, M. Gene expression in working skeletal muscle. *Adv. Exp. Med. Biol.* **618**, 245–254 (2007).
- Cannell, I. G., Kong, Y. W. & Bushell, M. How do microRNAs regulate gene expression? *Biochem. Soc. Trans.* **36**, 1224–1231 (2008).
- Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. cell* **27**, 91–105 (2007).
- Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
- Kosaka, N. *et al.* Secretory mechanisms and intercellular transfer of microRNAs in living cells. *J. Biol. Chem.* **285**, 17442–17452 (2010).
- Collino, F. *et al.* Microvesicles derived from adult human bone marrow and tissue specific mesenchymal stem cells shuttle selected pattern of miRNAs. *PLoS one* **5**, e11803 (2010).
- Lira, F. S. *et al.* Acute high-intensity exercise with low energy expenditure reduced LDL-c and total cholesterol in men. *Eur. J. Appl. Physiol.* **107**, 203–210 (2009).
- Arroyo, J. D. *et al.* Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci. USA* **108**, 5003–5008 (2011).
- Baggish, A. L. *et al.* Dynamic regulation of circulating microRNA during acute exhaustive exercise and sustained aerobic exercise training. *J. Physiol.* **589**, 3983–3994 (2011).
- Bye, A. *et al.* Circulating microRNAs and aerobic fitness—the HUNT-Study. *PLoS one* **8**, e57496 (2013).
- Nielsen, S. *et al.* The miRNA plasma signature in response to acute aerobic exercise and endurance training. *PLoS one* **9**, e87308 (2014).
- Mooren, F. C., Viereck, J., Kruger, K. & Thum, T. Circulating microRNAs as potential biomarkers of aerobic exercise capacity. *Am. J. Physiol. Heart Circ. Physiol.* **306**, 557–563 (2014).
- Sawada, S. *et al.* Profiling of Circulating MicroRNAs after a Bout of Acute Resistance Exercise in Humans. *PLoS one* **8**, e70823 (2013).
- Wardle, S. L. *et al.* Plasma MicroRNA Levels Differ between Endurance and Strength Athletes. *PLoS one* **10**, e0122107 (2015).
- Uhlemann, M. *et al.* Circulating microRNA-126 increases after different forms of endurance exercise in healthy adults. *Eur. J. Prev. Cardiol.* **21**, 484–491 (2014).
- Radom-Aizik, S. *et al.* Effects of exercise on microRNA expression in young males peripheral blood mononuclear cells. *Clin. Transl. Sci.* **5**, 32–38 (2012).
- Radom-Aizik, S., Zaldivar, F. Jr., Oliver, S., Galassetti, P. & Cooper, D. M. Evidence for microRNA involvement in exercise-associated neutrophil gene expression changes. *J. Appl. Physiol.* **109**, 252–261 (2010).
- Tonevitsky, A. G. *et al.* Dynamically regulated miRNA-mRNA networks revealed by exercise. *BMC Physiol.* **13**, 9–20 (2013).
- Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* **11**, 597–610 (2010).
- Liu, H. *et al.* Identifying mRNA targets of microRNA dysregulated in cancer: with application to clear cell Renal Cell Carcinoma. *BMC Syst. Biol.* **4**, 51–68 (2010).
- Boveris, A. & Navarro, A. Systemic and mitochondrial adaptive responses to moderate exercise in rodents. *Free Radic. Biol. Med.* **44**, 224–229 (2008).
- Capomaccio, S. *et al.* RNA sequencing of the exercise transcriptome in equine athletes. *PLoS one* **8**, e83504 (2013).
- Scott, J. M. *et al.* Cardiovascular Consequences of Completing a 160-km Ultramarathon. *Med. Sci. Sport Exerc.* **41**, 25–33 (2009).
- Mohr, S. & Liew, C. C. The peripheral-blood transcriptome: new insights into disease and risk assessment. *Trends Mol. Med.* **13**, 422–432 (2007).
- Barrey, E., Mucher, E., Robert, C., Amiot, F. & Gidrol, X. Gene expression profiling in blood cells of endurance horses completing competition or disqualified due to metabolic disorder. *Equine Vet. J. Suppl.* **36**, 43–49 (2006).
- Gim, J. A. *et al.* Transcriptional expression changes of glucose metabolism genes after exercise in thoroughbred horses. *Gene* **547**, 152–158 (2014).
- Kavazis, A. N., Smuder, A. J. & Powers, S. K. Effects of short-term endurance exercise training on acute doxorubicin-induced FoxO transcription in cardiac and skeletal muscle. *J. Appl. Physiol.* **117**, 223–230 (2014).
- Bianchessi, V. *et al.* The mitochondrial lncRNA ASncmtRNA-2 is induced in aging and replicative senescence in Endothelial Cells. *J. Mol. Cell Cardiol.* **81**, 62–70 (2015).
- Holloszy, J. O. & Coyle, E. F. Adaptations of skeletal muscle to endurance exercise and their metabolic consequences. *J. Appl. Physiol. Respir. Environ. Exerc. Physiol.* **56**, 831–838 (1984).
- Makarova, J. A. *et al.* Exercise immunology meets miRNAs. *Exerc. Immunol. Rev.* **20**, 135–164 (2014).
- Safdar, A., Abadi, A., Akhtar, M., Hettinga, B. P. & Tarnopolsky, M. A. miRNA in the regulation of skeletal muscle adaptation to acute endurance exercise in C57Bl/6J male mice. *PLoS one* **4**, e5610 (2009).
- Gomes, C. P. *et al.* Circulating miR-1, miR-133a, and miR-206 levels are increased after a half-marathon run. *Biomarkers* **19**, 585–589 (2014).
- Tzimagiorgis, G., Michailidou, E. Z., Kritis, A., Markopoulos, A. K. & Kouidou, S. Recovering circulating extracellular or cell-free RNA from bodily fluids. *Cancer Epidemiol.* **35**, 580–589 (2011).
- Moldovan, L. *et al.* Methodological challenges in utilizing miRNAs as circulating biomarkers. *J. Cell Mol. Med.* **18**, 371–390 (2014).
- Koulmann, N. & Bigard, A. Interaction between signalling pathways involved in skeletal muscle responses to endurance exercise. *Pflugers Archiv-Europ. J. Physiol.* **452**, 125–139 (2006).
- Thum, T. *et al.* MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature* **456**, 980–984 (2008).
- Godwin, J. G. *et al.* Identification of a microRNA signature of renal ischemia reperfusion injury. *Proc. Natl. Acad. Sci. USA* **107**, 14339–14344 (2010).
- Pellegrini Masini, A., Tedeschi, D., Badagli, P., C. S. & G. L. Exercise-induced intravascular haemolysis in standardbred horses. *Comp. Clin. Pathol.* **12**, 45–48 (2003).
- Inoue, Y. *et al.* Effect of exercise on iron metabolism in horses. *Biol. Trace Elem. Res.* **107**, 33–42 (2005).

44. Le Moyec, L. *et al.* Protein catabolism and high lipid metabolism associated with long distance exercise are revealed by plasma NMR metabolomics in endurance horses. *PLoS one* **9**, e90730 (2014).
45. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
46. Buza, T., Arick, M. 2nd, Wang, H. & Peterson, D. G. Computational prediction of disease microRNAs in domestic animals. *BMC Res. Notes* **7**, 403–416 (2014).
47. Desjardin, C. *et al.* Next-generation sequencing identifies equine cartilage and subchondral bone miRNAs and suggests their involvement in osteochondrosis physiopathology. *BMC genomics* **15**, 798–809 (2014).
48. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, 25–35 (2009).
49. Kauffmann, A., Gentleman, R. & Huber, W. ArrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics* **25**, 415–416 (2009).
50. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
51. Bland, J. M. & Altman, D. G. Multiple significance tests: the Bonferroni method. *BMJ* **310**, 170 (1995).
52. Lopez-Romero, P. Pre-processing and differential expression analysis of Agilent microRNA arrays using the AgiMicroRNA Bioconductor library. *BMC Genomics* **12**, 64–72 (2011).
53. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
54. Janky, R. *et al.* iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput. Biol.* **10**, e1003731 (2014).
55. Drugs to treat overweight and obesity. *J. Psychosoc. Nurs. Ment. Health Serv.* **52**, 21–22 (2014).
56. Ge, Y., Sealton, S. C. & Speed, T. P. Some Step-down Procedures Controlling the False Discovery Rate under Dependence. *Stat. Sin.* **18**, 881–904 (2008).
57. Assenov, Y., Ramirez, F., Schelhorn, S. E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics* **24**, 282–284 (2008).
58. Villers, F., Schaeffer, B., Bertin, C. & Huet, S. Assessing the validity domains of graphical Gaussian models in order to infer relationships among components of complex biological systems. *Stat. Appl. Genet. Mol. Biol.* **7**, Article 14 (2008).
59. Meinshausen, N. & Bühlmann, P. High-dimensional graphs and variable selection with the Lasso. *Ann. Stat.* **34**, 1436–1462 (2006).

Acknowledgements

This work was funded by grants from the *Fonds Eperon*, the *Institut Français du Cheval et de l'Équitation* (IFCE), the *Association du Cheval Arabe* (ACA), the Swiss National Science Foundation (grant No. 310030–138295) and Swiss Institute of Equine Medicine Research. The authors acknowledge the assistance of David Garal from Open University of Catalonia, Barcelona, Spain) who helped with the mitochondrial small RNA analysis, and the assistance of Xavier Mata, Sylvain Marthey, Rachel Legendre, Sean Kennedy and Laurent Schibler who designed the custom miRNA array. The authors also thank Stéphane Robin (from the *Mathématiques et Informatique Appliquées* group at INRA/AgroParisTech, Paris, France) for his recommendation of the hypergeometric variation test. Lastly, we also thank all the horse owners, riders and endurance race organizers who participated in the study and David Fraser from Biotech communication for the English editing.

Author Contributions

N.M. carried out the functional biological analysis and the mitochondrial sRNAseq bioinformatics analysis, wrote the main manuscript text, and prepared all the figures. S.P. designed and carried out the differential expression analysis, the hypergeometric tests and the GGM analyses. A.V. isolated the RNA. J.L. and M.M. performed the RNA labelling, microarray hybridization/scans and the exploratory statistical analysis. A.P. and M.B. helped to interpret data. J.R. carried out the RT-qPCR analyses. C.M. helped to check the quality of microarray design. A.N. helped with the functional biological analysis. C.R. and E.B. designed the study. All authors reviewed the manuscript and approved the final version.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Mach, N. *et al.* Integrated mRNA and miRNA expression profiling in blood reveals candidate biomarkers associated with endurance exercise in the horse. *Sci. Rep.* **6**, 22932; doi: 10.1038/srep22932 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

3. Résultats, conclusion et perspectives

Notre étude nous a permis d'identifier un ensemble de **2453 gènes** et de **167 miARN** présentant une **différence significative d'expression** entre les échantillons prélevés **avant et après la course** d'endurance. Nous avons utilisé un test hypergéométrique et sa généralisation afin de mieux cerner les fonctions biologiques régulées par les miARN à profil d'expression modifié suite à la course : **44** d'entre eux **régulent** vraisemblablement un ensemble de **351 gènes cibles sous-exprimés** suite à la course, impliqués dans les voies biologiques du **métabolisme glucidique**, de **l'oxydation des acides gras**, de **la biogénèse mitochondriale**, et de **la réponse immunitaire**. L'application de méthodes statistiques sur un échantillon indépendant d'animaux a permis de valider les miARN **miR-21-5p**, **miR-181b-5p**, et **miR-505-5p** comme molécules **candidates régulatrices** de la **réponse** à l'exercice **d'endurance chez le cheval**. Il s'agit, à notre connaissance, de la première étude fournissant une vue globale, complète et intégrée des réseaux de co-régulation transcriptomique et microtranscriptomique à rôle potentiel dans le contrôle des régulations post-transcriptionnelles liées à l'effort d'endurance. Nos résultats suggèrent en outre qu'un grand nombre de mécanismes sous-tendent les modifications transcriptomiques de réponse à ce type d'effort, et que les changements d'expression de quelques effecteurs régulateurs clés peuvent contribuer à une amélioration de la réponse métabolique et inflammatoire à l'exercice chez les sportifs de haut niveau.

L'idéal serait maintenant de compléter cette étude transcriptomique et microtranscriptomique par une approche métabolomique : en effet, grâce aux prélèvements sanguins effectués sur ces chevaux d'endurance, nous disposons d'un accès au métabolome plasmatique avant et après la course d'un bon nombre d'entre eux. Il s'agit donc de déterminer si nous pouvons identifier des biomarqueurs métabolomiques de l'effort d'endurance chez le cheval, et les associer aux données transcriptomiques et microtranscriptomiques que nous avons dégagées dans notre étude.

Au delà de ces analyses, il serait intéressant d'étudier les relations entre le microbiote des chevaux d'endurance et leurs données métabolomiques. En effet, on peut penser que des modulations de la flore intestinale de ces chevaux sont associées à une meilleure aptitude à valoriser les aliments distribués, notamment les fourrages, et à une meilleure performance énergétique, tandis que d'autres changements associés au microbiote de ces chevaux peuvent perturber le fonctionnement digestif et ainsi favoriser la survenue de troubles métaboliques.

CONCLUSION

Cette thèse m'a permis de présenter de manière détaillée la méthode, actuellement très en vogue, d'intégration de données « omiques » d'origines diverses, afin de comprendre les mécanismes sous-tendant l'acquisition d'un phénotype particulier, voire de prédire, dans certains cas, des caractères phénotypiques d'un individu. Elle a aussi été l'occasion d'effectuer un tour d'horizon des travaux dans lesquels cette approche novatrice est utilisée au jour d'aujourd'hui, et de présenter en détail notre travail expérimental portant sur la thématique de l'endurance équine. Ce dernier a mis en évidence des biomarqueurs « omiques » d'un exercice physique intense et prolongé chez les chevaux (3 microARN validés comme régulateurs de gènes montrant des différences significatives d'expression avant et après une course d'endurance), traduisant les modifications physiologiques à court et moyen termes engendrées par ce type d'effort sportif. Cette application très concrète d'analyse intégrée illustre bien la puissance de ce genre d'approche, qui va sûrement prendre une ampleur considérable dans les années à venir.

En plus d'ouvrir la voie à de nouvelles perspectives d'analyse combinée transcriptomique / microtranscriptomique dans l'espèce équine, notre travail, considéré non pas seul mais associé aux autres études de GenEndurance, permettra de fournir un aperçu global de l'athlète équin le plus à même de performer dans cette discipline d'endurance en plein essor, avec un génotype et un phénotype particuliers, des biomarqueurs « omiques » favorables, dont un profil métabolomique caractéristique... Les efforts se concentrent désormais sur l'analyse du microbiote intestinal des chevaux utilisés dans notre étude, en lien avec leurs données métabolomiques. Les résultats préliminaires de cette analyse font clairement apparaître deux groupes de chevaux caractérisés par des flores intestinales de composition bien distincte. Il s'agit donc d'identifier le facteur expliquant cette différence significative : la distinction est-elle réellement due à un profil métabolomique différent, ou n'est-elle pas plutôt imputable au régime alimentaire des chevaux, ou à un protocole d'entraînement sportif différent avant la course ? La collecte d'informations complémentaires auprès des propriétaires est en cours : ces précisions viendront étoffer la base de données, ce qui permettra de mener l'analyse en considérant plus finement les apports et dépenses énergétiques de ces animaux. Nous espérons

ainsi aller plus loin dans les investigations pour compléter, à terme, notre projet de volet métabolomique.

Les discussions avec quelques cavaliers et personnes familières du milieu équin permettent de soulever certaines questions. Nos travaux, ainsi que les objectifs de GenEndurance et des projets similaires, suscitent effectivement quelques récriminations légitimes, à savoir, notamment, une méfiance envers un « conditionnement par la génétique » : certains craignent que la sélection de chevaux pour telle ou telle discipline sur la base seule de leurs données « omiques » expose ce milieu à une fermeture d'esprit considérable et à une « tendance de ne voir l'animal qu'à travers ses données génétiques ». Mais un argument relativisant le poids de ces considérations est qu'il semble plus pertinent d'aider et d'encourager les cavaliers et propriétaires à « mieux » choisir leur monture, un cheval le plus susceptible de regrouper les caractéristiques favorables à réaliser de bonnes performances dans une discipline donnée, si cela peut permettre par exemple d'éviter ou de limiter le problème croissant de dopage dans le milieu sportif équin, sous prétexte de vouloir atteindre les performances souhaitées. Par ailleurs, le projet GenEndurance, et plus globalement l'endurance équine, doivent faire face à des contestations émanant de fervents partisans de la défense du bien-être animal, qui pointent du doigt l'effort, démesuré selon eux, demandé aux chevaux à l'occasion de ces courses de très longue durée. Il est certes vrai que ce type d'événement engendre un bouleversement physiologique et biochimique impressionnant chez ces athlètes, mais ces derniers sont entraînés pour, et aptes à tolérer de telles conditions et à récupérer relativement vite. Du moins, c'est à nous, vétérinaires, qu'il revient de s'assurer que les chevaux engagés dans ces courses le soient, afin d'éviter au maximum toute douleur ou souffrance de nos sportifs à quatre fers.

REFERENCES BIBLIOGRAPHIQUES

- ABE, S., MIYAKE, Y-I., KAGEYAMA, S-I., WATANABE, G., TAYA, K. and KAWAKURA, K. (1999). Deletion of the Sry region on the Y chromosome detected in a case of equine gonadal hypoplasia (XY female) with abnormal hormonal profiles. *Equine Veterinary Journal*, **31** (4), 236–238.
- ADAMU, L., NORANIZA, M. A., RASEDEE, A. and BASHIR, A. (2012). Metabolic responses in endurance horses during racing in relation to uric acid profile, leucocytes, heart rate and plasma biochemical parameters. *Veterinarni Medicina*, **57** (11), 591–596.
- AHLBORG, G., FELIG, P., HAGENFELDT, L., HENDLER, R. and WAHREN, J. (1974). Substrate turnover during prolonged exercise in man : splanchnic and leg metabolism of glucose, free fatty acids, and amino acids. *Journal of Clinical Investigation*, **53** (4), 1080.
- ALEMAN, M., NIETO, J.E. and MAGDESIAN, K.G. (2009). Malignant Hyperthermia Associated with Ryanodine Receptor 1 (C7360G) Mutation in Quarter Horses. *Journal of Veterinary Internal Medicine*, **23** (2), 329–334.
- ALEMAN, M., RIEHL, J., ALDRIDGE, B., LECOUTEUR, R., STOTT, J. and PESSAH, I. (2004). Association of a mutation in the ryanodine receptor 1 gene with equine malignant hyperthermia. *Muscle & Nerve*, **30** (3), 356–365.
- ALLAIS, S. (2011). *Détection et validation de marqueurs génétiques impliqués dans la qualité de la viande bovine*. Thèse de doctorat en génétique animale, AgroParisTech, 258 p.
- AMETAJ, B., ZEBELI, Q., SALEEM, F., PSYCHOGIOS, N., LEWIS, M., DUNN, S., XIA, J. and WISHART, D. (2010). Metabolomics reveals unhealthy alterations in rumen metabolism with increased proportion of cereal grain in the diet of dairy cows. *Metabolomics*, **6** (4), 583–594.
- ANDERSON, E. and SHIVAKUMAR, G. (2013). Effects of exercise and physical activity on anxiety. *Frontiers in Psychiatry*, **4** (27), 1-4.
- ANDERSSON, L. S., AXELSSON, J., DUBIELZIG, R., LINDGREN, G. and EKESTEN, B. (2011). Multiple congenital ocular anomalies in Icelandic horses. *BMC veterinary research*, **7** (1), 1-5.
- ANDERSSON, L. S., JURAS, R., RAMSEY, D., EASON-BUTLER, J., EWART, S., COTHRAN, G. and LINDGREN, G. (2008). Equine Multiple Congenital Ocular Anomalies maps to a 4.9 megabase interval on horse chromosome 6. *BMC Genetics*, **9** (1), 88.
- ANDERSSON, L. S., LARHAMMAR, M., MEMIC, F., WOOTZ, H., SCHWOCHOW, D., RUBIN, C.-J., PATRA, K., ARNASON, T., WELLBRING, L., HJÄLM, G., IMSLAND, F., PETERSEN, J., MCCUE, M., MICKELSON, J., COTHRAN, G., AHITUV, N., ROEPSTORFF, L., MIKKO, S., VALLSTEDT, A., LINDGREN, G., ANDERSSON, L. and KULLANDER, K. (2012). Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature*, **488** (7413), 642–646.
- ANDERSSON, L. S., WILBE, M., VILUMA, A., COTHRAN, G., EKESTEN, B., EWART, S. and LINDGREN, G. (2013). Equine Multiple Congenital Ocular Anomalies and Silver Coat Colour Result from the Pleiotropic Effects of Mutant PMEL. *PLoS ONE*, **8** (9), e75639.
- AUTON, A., ABECASIS, G., and others (2015). A global reference for human genetic variation. *Nature*, **526** (7571), p. 68–74.
- BAGGISH, A., HALE, A., WEINER, R., LEWIS, G., SYSTROM, D., WANG, F., WANG, T. and CHAN, S., (2011). Dynamic regulation of circulating microRNA during acute exhaustive exercise and sustained aerobic exercise training : Circulating microRNA in exercise. *The Journal of Physiology*, **589** (16), 3983–3994.
- BAILEY, E., REID, R., LEAR, T., SKOW, L., MATHIASON, K. and MCGUIRE, T. (1997). Linkage of the gene

- for equine combined immunodeficiency disease to microsatellite markers HTG8 and HTG4 ; synteny and FISH mapping to ECA9. *Animal genetics*, **28** (4), 268–273.
- BAILEY, E. (2011). Screening for foal immunodeficiency syndrome. *Veterinary Record*, **169** (25), 653–654.
- BAILEY, E. (2014). Five things equine veterinarians should know about genomics. *Equine Veterinary Journal*, **46** (4), 404–407.
- BAILLY-CHOURIBERRY, L., NOGUIER, F., MANCHON, L., PIQUEMAL, D., GARCIA, P., POPOT, M.-A. and BONNAIRE, Y. (2010). Blood cells RNA biomarkers as a first long-term detection strategy for EPO abuse in horseracing. *Drug Testing and Analysis*, **2** (7), 339–345.
- BAIRD, J., MILLON, L., DILEANIS, S., PENEDO, M., CHARLESWORTH, A., SPIRITO, F. and MENEGUZZI, G. (2003). Junctional epidermolysis bullosa in Belgian draft horses. *Proceedings of the American Association of Equine Practitioners*, **49**, 122–126.
- BAKER, E., GALLOWAY, L., JACKSON, B., SCHMOYER, D. and SNODDY, J. (2004). MuTrack : a genome analysis system for large-scale mutagenesis in the mouse. *Bmc Bioinformatics*, **5** (1), 11.
- BARREY, E., BONNAMY, B., BARREY, E. J., MATA, X., CHAFFAUX, S. and GUERIN, G. (2010). Muscular microRNA expressions in healthy and myopathic horses suffering from polysaccharide storage myopathy or recurrent exertional rhabdomyolysis : microRNA profile in equine myopathies. *Equine Veterinary Journal*, **42**, 303–310.
- BARREY, E., MUCHER, E., ROBERT, C., AMIOT, F. and GIDROL, X. (2006). Gene expression profiling in blood cells of endurance horses completing competition or disqualified due to metabolic disorder. *Equine Veterinary Journal*, **38** (S36), 43–49.
- BARREY, E., COLIN DE VERDIÈRE, J., MATA, X., SCHIBLER, L. and ROBERT, C. (2012). Pre-ride gait variables and risk for elimination in endurance rides. In : *7th International Conference on Canine and Equine Locomotion*, 25-28 June 2012, Strömsholm (Suède). p. 113.
- BARREY, E. (2010). Genetics and genomics in equine exercise physiology : an overview of the new applications of molecular biology as positive and negative markers of performance and health. *Equine Veterinary Journal*, **42**, 561–568.
- BARTON, C., BECK, P., KAY, R., TEALE, P. and ROBERTS, J. (2009). Multiplexed LC-MS/MS analysis of horse plasma proteins to study doping in sport. *Proteomics*, **9** (11), 3058–3065.
- BASSINI, A. and CAMERON, L.-C. (2014). Sportomics : Building a new concept in metabolic studies and exercise science. *Biochemical and Biophysical Research Communications*, **445** (4), 708–716.
- BASSINI-CAMERON, A., SWEET, E., BOTTINO, A., BITTAR, C., VEIGA, C., and CAMERON, L.-C. (2007). Effect of caffeine supplementation on haematological and biochemical variables in elite soccer players under physical stress conditions. *British Journal of Sports Medicine*, **41** (8), 523–530.
- BEGER, R. (2007). Cambridge Healthtech Institute's 7th Annual, Identifying and Validating Metabolic Markers for Drug Development and Clinical Studies. *Expert Review of Molecular Diagnostics*, **7** (2), 113–115.
- BELLONE, R. R., HOLL, H., SETALURI, V., DEVI, S., MADDODI, N., ARCHER, S., SANDMEYER, L., LUDWIG, A., FOERSTER, D., PRUVOST, M., REISSMANN, M., BORTFELDT, R., ADELSON, D., LIM, S. L., NELSON, J., HAASE, B., ENGENSTEINER, M., LEEB, T., FORSYTH, G., MIENALTOWSKI, M., MAHADEVAN, P., HOFREITER, M., PAIJMANS, J., GONZALEZ-FORTES, G., GRAHN, B. and BROOKS, S. (2013). Evidence for a Retroviral Insertion in TRPM1 as the Cause of Congenital Stationary Night Blindness and Leopard Complex Spotting in the Horse. *PLoS ONE*, **8** (10), e78280.
- BELLONE, R. R., FORSYTH, G., LEEB, T., ARCHER, S., SIGURDSSON, S., IMSLAND, F., MAUCELI, E., ENGENSTEINER, M., BAILEY, E., SANDMEYER, L., GRAHN, B., LINDBLAD-TOH, K. and WADE, C. M. (2010). Fine-mapping and mutation analysis of TRPM1 : a candidate gene for leopard complex (LP) spotting and congenital stationary night blindness in horses. *Briefings in Functional Genomics*, **9** (3), 193–207.

- BELLONE, R. R. (2010). Pleiotropic effects of pigmentation genes in horses : Effects of equine pigmentation genes. *Animal Genetics*, **41**, 100–110.
- BERNARD, C., CASSAR-MALEK, I., LE CUNFF, M., DUBROEUCQ, H., RENAND, G. and HOCQUETTE, J.-F. (2007). New Indicators of Beef Sensory Quality Revealed by Expression of Specific Genes. *Journal of Agricultural and Food Chemistry*, **55** (13), 5229–5237.
- BERNARD, L., LEROUX, C. and CHILLIARD, Y. (2008). Expression and Nutritional Regulation of Lipogenic Genes in the Ruminant Lactating Mammary Gland. In : *Bioactive Components of Milk*. New York : Springer, 2008, p. 67–108. (Advances in Experimental Medicine and Biology, 606).
- BERNOCO, D. and BAILEY, E. (1998). Frequency of the SCID gene among Arabian horses in the USA. *Animal Genetics*, **29** (1), 41–42.
- BESSA, A., NISSENBAUM, M., MONTEIRO, A., GANDRA, P. G., NUNES, L. S., BASSINI-CAMERON, A., WERNECK-DE-CASTRO, J. P. S., DE MACEDO, D. V. and CAMERON, L.-C. (2008). High-intensity ultraendurance promotes early release of muscle injury markers. *British Journal of Sports Medicine*, **42** (11), 589–593.
- BINNS, M. M., BOEHLER, D. A. and LAMBERT, D. H. (2010). Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA : Identification of MSTN. *Animal Genetics*, **41**, 154–158.
- BIOMARKERS DEFINITIONS WORKING GROUP (2001). Biomarkers and surrogate endpoints : Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, **69** (3), 89–95.
- BIRNEY, E. and SORANZO, N. (2015). Human genomics : The end of the start for population sequencing. *Nature*, **526** (7571), 52–53.
- BLOKHUIS, H. J., VEISSIER, I., MIELE, M. and JONES, B. (2010). The Welfare Quality® project and beyond : Safeguarding farm animal well-being. *Acta Agriculturae Scandinavica, Section A - Animal Science*, **60** (3), 129–140.
- BOGARD, M., AMEZIANE, N. and LAMORIL, J. (2008). Microarray d'ADN et profils d'expression des gènes. *Immuno-analyse & Biologie Spécialisée*, **23** (2), 71–88.
- BOGUE, M. (2003). Mouse Phenome Project : understanding human biology through mouse genetics and genomics. *Journal of Applied Physiology*, **95** (4), 1335–1337.
- BOUCHARD, C., SARZYNSKI, M. A., RICE, T. K., KRAUS, W. E., CHURCH, T. S., SUNG, Y. J., RAO, D. C. and RANKINEN, T. (2011). Genomic predictors of the maximal O₂ uptake response to standardized exercise training programs. *Journal of Applied Physiology*, **110** (5), 1160–1170.
- BOUDRA, H., GODEJOHANN, M. and MORGAVI, D. (2008). Uncovering metabolic changes using a combined LC-MS/NMR approach : Ochratoxin's effect in sheep. In : *3èmes Journées Scientifiques du Réseau Français de Métabolomique et Fluxomique*, 7 et 8 février 2008, Bordeaux, France [Online]. Available from : <http://agris.fao.org/agris-search/search.do?recordID=FR2014010593> [Accessed 16 December 2015].
- BOULEY, J., MEUNIER, B., CHAMBON, C., DE SMET, S., HOCQUETTE, J. F. and PICARD, B. (2005). Proteomic analysis of bovine skeletal muscle hypertrophy. *Proteomics*, **5** (2), 490–500.
- BOUWMAN, F. G., VAN GINNEKEN, M. M.E., NOBEN, J.-P., ROYACKERS, E., DE GRAAF-ROELFSEMA, E., WIJNBERG, I. D., VAN DER KOLK, J. H., MARIMAN, E. C.M. and VAN BREDA, E. (2010). Differential expression of equine muscle biopsy proteins during normal training and intensified training in young standardbred horses using proteomics technology. *Comparative Biochemistry and Physiology Part D : Genomics and Proteomics*, **5** (1), 55–64.
- BOUWMAN, F. G., VAN GINNEKEN, M. M.E., VAN DER KOLK, J. H., VAN BREDA, E. and MARIMAN, E. C.M. (2010). Novel markers for tying-up in horses by proteomics analysis of equine muscle biopsies.

BRANTH, S., HAMBRAEUS, L., PIEHL-AULIN, K., ESSÉN-GUSTAVSSON, B., ÅKERFELDT, T., OLSSON, R., STRIDSBERG, M. and RONQUIST, G. (2009). Metabolic stress-like condition can be induced by prolonged strenuous exercise in athletes. *Uppsala Journal of Medical Sciences*, **114** (1), 12–25.

BRARD, S. (2015). *Quel cadre théorique et pratique pour l'utilisation de la sélection génomique dans l'amélioration génétique des chevaux ?* Thèse de doctorat de l'Université de Toulouse en Pathologie, Toxicologie, Génétique et Nutrition, Institut National Polytechnique de Toulouse, 183 p.

BRAULT, L. S., COOPER, C. A., FAMULA, T. R., MURRAY, J. D. and PENEDO, M. C. T. (2011). Mapping of equine cerebellar abiotrophy to ECA2 and identification of a potential causative mutation affecting expression of MUTYH. *Genomics*, **97** (2), 121–129.

BRAULT, L. S., FAMULA, T. R. and PENEDO, M. C. T. (2011). Inheritance of cerebellar abiotrophy in Arabians. *American journal of veterinary research*, **72** (7), 940–944.

BREIMAN, L. (2001). Random forests. *Machine learning*, **45** (1), 5–32.

BRIGHT, L. A., MUJAHID, N., NANDURI, B., MCCARTHY, F. M., COSTA, L. R. R., BURGESS, S. C. and SWIDERSKI, C. E. (2011). Functional modelling of an equine bronchoalveolar lavage fluid proteome provides experimental confirmation and functional annotation of equine genome sequences : Lung fluid proteome provides annotation of equine genome sequences. *Animal Genetics*, **42** (4), 395–405.

BRONZE-DA-ROCHA, E. (2014). MicroRNAs Expression Profiles in Cardiovascular Diseases. *BioMed Research International*, **2014**, 1–23.

BROOKS, S. A., GABRESKI, N., MILLER, D., BRISBIN, A., BROWN, H. E., STREETER, C., MEZEY, J., COOK, D. and ANTCZAK, D. F. (2010). Whole-Genome SNP Association in the Horse : Identification of a Deletion in Myosin Va Responsible for Lavender Foal Syndrome. *PLoS Genetics*, **6** (4), e1000909.

BROSNAHAN, M. M., BROOKS, S. A. and ANTCZAK, D. F. (2010). Equine clinical genomics : A clinician's primer. *Equine Veterinary Journal*, **42** (7), 658–670.

BROSNAHAN, M. M., MILLER, D. C., ADAMS, M. and ANTCZAK, D. F. (2012). IL-22 Is Expressed by the Invasive Trophoblast of the Equine (*Equus caballus*) Chorionic Girdle. *The Journal of Immunology*, **188** (9), 4181–4187.

BROSNAN, C. A. and VOINNET, O. (2009). The long and the short of noncoding RNAs. *Current Opinion in Cell Biology*, **21** (3), 416–425.

BUECHLI, M. E., LAMARRE, J. and KOCH, T. G. (2012). MicroRNA-140 Expression During Chondrogenic Differentiation of Equine Cord Blood-Derived Mesenchymal Stromal Cells. *Stem Cells and Development*, **22** (8), 1288–1296.

BUTTERFIELD, T. A., BEST, T. M. and MERRICK, M. A. (2006). The dual roles of neutrophils and macrophages in inflammation : a critical balance between tissue damage and repair. *Journal of athletic training*, **41** (4), 457.

CAPOMACCIO, S., CAPPELLI, K., BARREY, E., FELICETTI, M., SILVESTRELLI, M. and VERINI-SUPPLIZI, A. (2010). Microarray analysis after strenuous exercise in peripheral blood mononuclear cells of endurance horses : Exercise expression profiling in horse. *Animal Genetics*, **41**, 166–175.

CAPOMACCIO, S., VITULO, N., VERINI-SUPPLIZI, A., BARCACCIA, G., ALBIERO, A., D'ANGELO, M., CAMPAGNA, D., VALLE, G., FELICETTI, M., SILVESTRELLI, M. and CAPPELLI, K. (2013). RNA Sequencing of the Exercise Transcriptome in Equine Athletes. *PLoS ONE*, **8** (12), e83504.

CAPPELLI, K., FELICETTI, M., CAPOMACCIO, S., PIERAMATI, C., SILVESTRELLI, M. and VERINI-SUPPLIZI, A. (2009). Exercise-induced up-regulation of MMP-1 and IL-8 genes in endurance horses. *BMC Physiology*, **9** (1), 12.

- CAPPELLI, K., VERINI-SUPPLIZI, A., CAPOMACCIO, S. and SILVESTRELLI, M. (2007). Analysis of peripheral blood mononuclear cells gene expression in endurance horses by cDNA-AFLP technique. *Research in Veterinary Science*, **82** (3), 335–343.
- CAROLI, A., CARDILLO, M. T., GALEA, R. and BIASUCCI, L. M. (2013). Potential therapeutic role of microRNAs in ischemic heart disease. *Journal of Cardiology*, **61** (5), 315–320.
- CASSAR-MALEK, I., BERNARD, C., JURIE, C., BARNOLA, I., GENTÈS, G., DOZIAS, D., MICOL, D. and HOCQUETTE, J. F. (2005). Pasture-based beef production systems may influence muscle characteristics and gene expression. *Indicators of milk and beef quality* (eds J. F. Hocquette and S. Gigli), *EAAP Publication*, **112**, 385–390.
- CAVALLERI, J. M.V., METZGER, J., HELDIGE, M., LAMPE, V., STUCKENSCHNEIDER, K., TIPOLD, A., BEINEKE, A., BECKER, K., DISTL, O. and FEIGE, K. (2013). Morphometric magnetic resonance imaging and genetic testing in cerebellar atrophy in Arabian horses. *BMC veterinary research*, **9** (1), 105.
- CAVILL, R., KAMBUROV, A., ELLIS, J. K., ATHERSUCH, T. J., BLAGROVE, M. S. C., HERWIG, R., EBBELS, T. M. D. and KEUN, H. C. (2011). Consensus-Phenotype Integration of Transcriptomic and Metabolomic Data Implies a Role for Metabolism in the Chemosensitivity of Tumour Cells. *PLoS Computational Biology*, **7** (3), e1001113.
- CHAUSSABEL, D., PASCUAL, V. and BANCHEREAU, J. (2010). Assessing the human immune system through blood transcriptomics. *BMC biology*, **8** (1), 84.
- CHAZE, T., MEUNIER, B., CHAMBON, C., JURIE, C. and PICARD, B. (2008). In vivo proteome dynamics during early bovine myogenesis. *Proteomics*, **8** (20), 4236–4248.
- CHEN, X., BA, Y., MA, L., CAI, X., YIN, Y., WANG, K., GUO, J., ZHANG, Y., CHEN, J. and GUO, X. (2008). Characterization of microRNAs in serum : a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell research*, **18** (10), 997–1006.
- CHESLER, E. J., LU, L., WANG, J., WILLIAMS, R. W. and MANLY, K. F. (2004). WebQTL : rapid exploratory analysis of gene expression and genetic networks for brain and behavior. *Nature neuroscience*, **7** (5), 485–486.
- CHINWALLA, A. T., COOK, L. L., DELEHAUNTY, K. D., FEWELL, G. A., FULTON, L. A., FULTON, R. S., GRAVES, T. A., HILLIER, L. W., MARDIS, E. R. and MCPHERSON, J. D. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420** (6915), 520–562.
- CHO, W. C.S. (2007). Proteomics Technologies and Challenges. *Genomics, Proteomics & Bioinformatics*, **5** (2), 77–85.
- CHOW, J. and HEARD, E. (2009). X inactivation and the complexities of silencing a sex chromosome. *Current Opinion in Cell Biology*, **21** (3), 359–366.
- CHRISTOPHERSON, P. W., SANTEN, V. L., LIVESEY, L. and BOUDREAUX, M. K. (2007). A 10-Base-Pair Deletion in the Gene Encoding Platelet Glycoprotein IIb Associated with Glanzmann Thrombasthenia in a Horse. *Journal of veterinary internal medicine*, **21** (1), 196–198.
- CHRISTOPHERSON, P. W., INSALACO, T. A., VAN SANTEN, V. L., LIVESEY, L., BOURNE, C. and BOUDREAUX, M. K. (2006). Characterization of the cDNA encoding α IIb and β 3 in normal horses and two horses with Glanzmann thrombasthenia. *Veterinary Pathology Online*, **43** (1), 78–82.
- COLEMAN, S. J., ZENG, Z., WANG, K., LUO, S., KHREBTUKOVA, I., MIENALTOWSKI, M. J., SCHROTH, G. P., LIU, J. and MACLEOD, J. N. (2010). Structural annotation of equine protein-coding genes determined by mRNA sequencing : Structural annotation of equine protein-coding genes. *Animal Genetics*, **41**, 121–130.
- COLEMAN, S. J., ZENG, Z., LIU, J. and MACLEOD, J. N. (2010). Analysis of equine protein-coding gene structure and expression by RNA-sequencing. *BMC Bioinformatics*, **11** (S4), 8.

COLIN DE VERDIÈRE, J. (2014). *Caractérisation de la locomotion du cheval d'endurance : critères associés à la performance en course à vitesse libre*. Thèse de doctorat vétérinaire, Faculté de Médecine de Créteil, 110 p.

COMBOURIEU, B. (2007). Du métabolisme au métabolome : complémentarité des approches par RMN et spectrométrie de masse. *Spectra Analyse*, **36** (257), 20.

CONRAD, D. H., GOYETTE, J. and THOMAS, P. S. (2008). Proteomics as a Method for Early Detection of Cancer : A Review of Proteomics, Exhaled Breath Condensate, and Lung Cancer Screening. *Journal of General Internal Medicine*, **23** (S1), 78–84.

CORBIN, L. J., BLOTT, S. C., SWINBURNE, J. E., SIBBONS, C., FOX-CLIPSHAM, L. Y., HELWEGEN, M., PARKIN, T. D. H., NEWTON, J. R., BRAMLAGE, L. R., MCILWRAITH, C. W., BISHOP, S. C., WOOLLIAMS, J. A. and VAUDIN, M. (2012). A genome-wide association study of osteochondritis dissecans in the Thoroughbred. *Mammalian Genome*, **23** (3-4), 294–303.

COYLE, E. F. (2000). Physical activity as a metabolic stressor. *The American Journal of Clinical Nutrition*, **72** (2), 512S–520S.

CRISTIANINI, N. and SHAW-TAYLOR, J. (2000). *An introduction to support vector machines and other kernel-based learning methods* [Online]. Cambridge University Press, 189 p. Available from : https://books.google.fr/books?hl=fr&lr=&id=_PXJn_cxv0AC&oi=fnd&pg=PR9&dq=An+Introduction+to+Support+Vector+Machines+and+other+kernel+based+learning+methods&ots=xRRd7A_u_b&sig=QQOito_d1HDUGA9PBQpV8-v7r-w [Accessed 19 December 2015].

CUISINIER, C., WARD, R. J., FRANCAUX, M., STURBOIS, X. and DE WITTE, P. (2001). Changes in plasma and urinary taurine and amino acids in runners immediately and 24 h after a marathon. *Amino acids*, **20** (1), 13–23.

DAETWYLER, H. D., CAPITAN, A., PAUSCH, H., STOTHARD, P., VAN BINSBERGEN, R., BRØNDUM, R. F., LIAO, X., DJARI, A., RODRIGUEZ, S. C., GROHS, C., ESQUERRÉ, D., BOUCHEZ, O., ROSSIGNOL, M.-N., KLOPP, C., ROCHA, D., FRITZ, S., EGGEN, A., BOWMAN, P. J., COOTE, D., CHAMBERLAIN, A. J., ANDERSON, C., VANTASSELL, C. P., HULSEGG, I., GODDARD, M. E., GULDBRANDTSEN, B., LUND, M. S., VEERKAMP, R. F., BOICHARD, D. A., FRIES, R. and HAYES, B. J. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics*, **46** (8), 858–865.

DALLOUL, R. A., LONG, J. A., ZIMIN, A. V., ASLAM, L., BEAL, K., ANN BLOMBERG, L., BOUFFARD, P., BURT, D. W., CRASTA, O., CROOIJMANS, R. P. M. A., COOPER, K., COULOMBE, R. A., DE, S., DELANY, M. E., DODGSON, J. B., DONG, J. J., EVANS, C., FREDERICKSON, K. M., FLICEK, P., FLOREA, L., FOLKERTS, O., GROENEN, M. A. M., HARKINS, T. T., HERRERO, J., HOFFMANN, S., MEGENS, H.-J., JIANG, A., DE JONG, P., KAISER, P., KIM, H., KIM, K.-W., KIM, S., LANGENBERGER, D., LEE, M.-K., LEE, T., MANE, S., MARCAIS, G., MARZ, M., MCELROY, A. P., MODISE, T., NEFEDOV, M., NOTREDAME, C., PATON, I. R., PAYNE, W. S., PERTEA, G., PRICKETT, D., PUIU, D., QIOA, D., RAINERI, E., RUFFIER, M., SALZBERG, S. L., SCHATZ, M. C., SCHEURING, C., SCHMIDT, C. J., SCHROEDER, S., SEARLE, S. M. J., SMITH, E. J., SMITH, J., SONSTEGARD, T. S., STADLER, P. F., TAFER, H., TU, Z., VAN TASSELL, C. P., VILELLA, A. J., WILLIAMS, K. P., YORKE, J. A., ZHANG, L., ZHANG, H.-B., ZHANG, X., ZHANG, Y. and REED, K. M. (2010). Multi-Platform Next-Generation Sequencing of the Domestic Turkey (*Meleagris gallopavo*) : Genome Assembly and Analysis. *PLoS Biology*, **8** (9), e1000475.

DALRYMPLE, B. P., KIRKNESS, E. F., NEFEDOV, M., MCWILLIAM, S., RATNAKUMAR, A., BARRIS, W., ZHAO, S., SHETTY, J., MADDOX, J. F. and O'GRADY, M. (2007). Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome biology*, **8** (7), 1–20.

DE LA TORRE, A., GRUFFAT, D., DURAND, D., MICOL, D., PEYRON, A., SCISLOWSKI, V. and BAUCHART, D. (2006). Factors influencing proportion and composition of CLA in beef. *Meat Science*, **73** (2), 258–268.

DESJARDIN, C., VAIMAN, A., MATA, X., LEGENDRE, R., LAUBIER, J., KENNEDY, S. P., LALOE, D., BARREY, E., JACQUES, C. and CRIBIU, E. P. (2014). Next-generation sequencing identifies equine cartilage and subchondral bone miRNAs and suggests their involvement in osteochondrosis physiopathology. *BMC*

genomics, **15** (1), 798.

DE VANSSAY, A., BOUGÉ, A.-L., BOIVIN, A., HERMANT, C., TEYSSET, L., DELMARRE, V., ANTONIEWSKI, C. and RONSSERAY, S. (2012). Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature*, **490** (7418), 112–115.

DEVAUX, Y., AZUAJE, F., VAUSORT, M., YVORRA, C. and WAGNER, D. R. (2010). Integrated protein network and microarray analysis to identify potential biomarkers after myocardial infarction. *Functional & Integrative Genomics*, **10** (3), 329–337.

DOAN, R., COHEN, N. D., SAWYER, J., GHAFARI, N., JOHNSON, C. D. and DINDOT, S. V. (2012). Whole-genome sequencing and genetic variant analysis of a Quarter Horse mare. *BMC genomics*, **13** (1), 78.

DRÖGEMÜLLER, M., JAGANNATHAN, V., WELLE, M. M., GRAUBNER, C., STRAUB, R., GERBER, V., BURGER, D., SIGNER-HASLER, H., PONCET, P.-A., KLOPFENSTEIN, S., VON NIEDERHÄUSERN, R., TETENS, J., THALLER, G., RIEDER, S., DRÖGEMÜLLER, C. and LEEB, T. (2014). Congenital Hepatic Fibrosis in the Franches-Montagnes Horse Is Associated with the Polycystic Kidney and Hepatic Disease 1 (PKHD1) Gene. *PLoS ONE*, **9** (10), e110125.

DUCRO, B. J., SCHURINK, A., BASTIAANSEN, J. W. M., BOEGHEIM, I. J. M., VAN STEENBEEK, F. G., VOS-LOOHUIS, M., NIJMAN, I. J., MONROE, G. R., HELLINGA, I., DIBBITS, B. W., BACK, W. and LEEGWATER, P. A. J. (2015). A nonsense mutation in B3GALNT2 is concordant with hydrocephalus in Friesian horses. *BMC Genomics*, **16** (1), 761.

DUPUIS, M.-C., ZHANG, Z., DRUET, T., DENOIX, J.-M., CHARLIER, C., LEKEUX, P. and GEORGES, M. (2011). Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse. *Mammalian Genome*, **22** (9-10), 613–620.

EBERTH, J. E. (2013). *Chondrodysplasia-Like Dwarfism in the Miniature Horse*. Theses and Dissertations -- Veterinary Science. Paper 11. University of Kentucky, 83 p. http://uknowledge.uky.edu/gluck_etds/11.

EISSA, N. T. (2013). The exosome in lung diseases : Message in a bottle. *Journal of Allergy and Clinical Immunology*, **131** (3), 904–905.

ESCALONA, E. E., LENG, J., DONA, A. C., MERRIFIELD, C. A., HOLMES, E., PROUDMAN, C. J. and SWANN, J. R. (2015). Dominant components of the Thoroughbred metabolome characterised by 1H-nuclear magnetic resonance spectroscopy : A metabolite atlas of common biofluids : 1H-NMR metabolomic characterisation of equine blood, urine and faeces. *Equine Veterinary Journal*, **47** (6), 721–730.

ESVAN, S., DRAGAN, C., VARENNE, A., ASTRUC, J.M., BARILLET, F., BOICHARD, D., BRUNSCHWIG, P., DUBRULLE, A., FAUCON-LAHALLE, F. and FERLAY, A. (2010). PhénoFinlait, 1ers résultats : influence de l'alimentation, de l'état physiologique et de la génétique sur la composition en acides gras des laits de vache, brebis et chèvre. In : *1^{re} Edition Journées Rencontres Recherches Ruminants*, 6 et 7 décembre 2010, Paris, France [Online]. Available from : http://www.journees3r.fr/IMG/pdf/2010_12_06_Esvan.pdf [Accessed 14 December 2015].

FASANARO, P., GRECO, S., IVAN, M., CAPOGROSSI, M. C. and MARTELLI, F. (2010). microRNA: Emerging therapeutic targets in acute ischemic diseases. *Pharmacology & Therapeutics*, **125** (1), 92–104.

FAUCON-LAHALLE, F. (coord.) and MONTAGU, D. (coord.) (2011). *La révolution génomique animale*. Paris, France : France Agricole, 2011, 214 p. ISBN 978-2-85557-181-2.

FEBBRAIO, M. A. and PEDERSEN, B. K. (2002). Muscle-derived interleukin-6 : mechanisms for activation and possible biological roles. *The FASEB Journal*, **16** (11), 1335–1347.

FEHRENBACH, E. (2006). Multifarious microarray-based gene expression patterns in response to exercise. *Journal of Applied Physiology*, **102** (1), 7–8.

FIGEYS, D. and PINTO, D. (2001). Proteomics on a chip : Promising developments. *Electrophoresis*, **22** (2), 208–216.

- FINNO, C. J., STEVENS, C., YOUNG, A., AFFOLTER, V., JOSHI, N. A., RAMSAY, S. and BANNASCH, D. L. (2015). SERPINB11 Frameshift Variant Associated with Novel Hoof Specific Phenotype in Connemara Ponies. *PLoS Genetics*, **4**, e1005122.
- FINNO, C. J. and BANNASCH, D. L. (2014). Applied equine genetics. *Equine Veterinary Journal*, **46** (5), 538–544.
- FOX-CLIPSHAM, L. Y., CARTER, S. D., GOODHEAD, I., HALL, N., KNOTTENBELT, D. C., MAY, P. D. F., OLLIER, W. E. and SWINBURNE, J. E. (2011). Identification of a Mutation Associated with Fatal Foal Immunodeficiency Syndrome in the Fell and Dales Pony. *PLoS Genetics*, **7** (7), e1002133.
- FREIMER, N. and SABATTI, C. (2003). The Human Phenome Project. *Nature Genetics*, **34** (1), 15–21.
- FRISCHKNECHT, M., JAGANNATHAN, V., PLATTET, P., NEUDITSCHKO, M., SIGNER-HASLER, H., BACHMANN, I., PACHOLEWSKA, A., DRÖGEMÜLLER, C., DIETSCHI, E., FLURY, C., RIEDER, S. and LEEB, T. (2015). A Non-Synonymous HMGA2 Variant Decreases Height in Shetland Ponies and Other Small Horses. *PLoS ONE*, **10** (10), e0140749.
- GABRESKI, N. A., HAASE, B., ARMSTRONG, C. D., DISTL, O. and BROOKS, S. A. (2012). Investigation of allele frequencies for Lavender foal syndrome in the horse. *Animal Genetics*, **43** (5), 650.
- GALASSO, M., SANA, M. E. and VOLINIA, S. (2010). Non-coding RNAs : a key to future personalized molecular therapy. *Genome Medicine*, **2** (2), 12.
- GANESALINGAM, J. and BOWSER, R. (2010). The application of biomarkers in clinical trials for motor neuron disease. *Biomarkers in Medicine*, **4** (2), 281–297.
- GEISLER, S. and COLLER, J. (2013). RNA in unexpected places : long non-coding RNA functions in diverse cellular contexts. *Nature Reviews Molecular Cell Biology*, **14** (11), 699–712.
- GIBBS, R. A., WEINSTOCK, G. M., METZKER, M. L., MUZNY, D. M., SODERGREN, E. J., SCHERER, S., SCOTT, G., STEFFEN, D., WORLEY, K. C. and BURCH, P. E. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428** (6982), 493–521.
- GIM, J.-A., AYARPADIKANNAN, S., EO, J., KWON, Y.-J., CHOI, Y., LEE, H.-K., PARK, K.-D., YANG, Y. M., CHO, B.-W. and KIM, H.-S. (2014). Transcriptional expression changes of glucose metabolism genes after exercise in thoroughbred horses. *Gene*, **547** (1), 152–158.
- GOBERT, M., BOURGUET, C., TERLOUW, C., DEISS, V., BERDEAUX, O., COMTE, B., GRUFFAT, D., BAUCHART, D. and DURAND, D. (2009). Pre-slaughter stress and lipoperoxidation : protective effect of vitamin E and plant extracts rich in polyphenols given to finishing cattle. In : *Ruminant physiology. Digestion, metabolism and effects of nutrition on reproduction and welfare*. Proceedings of the 11th International Symposium on Ruminant Physiology, 6-9 September 2009, Clermont-Ferrand, France. Wageningen, Netherlands : Wageningen Academic Publishers, 814–815.
- GONÇALVES, L. C., BESSA, A., FREITAS-DIAS, R., LUZES, R., WERNECK-DE-CASTRO, J. P. S., BASSINI, A. and CAMERON, L.-C. (2012). A sportomics strategy to analyze the ability of arginine to modulate both ammonia and lymphocyte levels in blood after high-intensity exercise. *Journal of the International Society of Sports Nutrition*, **9**, 30.
- GOODWIN, S., MCPHERSON, J. D. and MCCOMBIE, W. R. (2016). Coming of age : ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, **17** (6), 333–351.
- GOTO, H., RYDER, O. A., FISHER, A. R., SCHULTZ, B., KOSAKOVSKY POND, S. L., NEKRUTENKO, A. and MAKOVA, K. D. (2011). A Massively Parallel Sequencing Approach Uncovers Ancient Origins and High Genetic Variability of Endangered Przewalski's Horses. *Genome Biology and Evolution*, **3** (0), 1096–1106.
- GRAVES, K. T., HENNEY, P. J. and ENNIS, R. B. (2009). Partial deletion of the LAMA3 gene is responsible for hereditary junctional epidermolysis bullosa in the American Saddlebred Horse. *Animal Genetics*, **40** (1), 35–

- GROENEN, M. A. M., ARCHIBALD, A. L., UENISHI, H., TUGGLE, C. K., TAKEUCHI, Y., ROTHSCHILD, M. F., ROGEL-GAILLARD, C., PARK, C., MILAN, D., MEGENS, H.-J., LI, S., LARKIN, D. M., KIM, H., FRANTZ, L. A. F., CACCAMO, M., AHN, H., AKEN, B. L., ANSELMO, A., ANTHON, C., AUVIL, L., BADAOU, B., BEATTIE, C. W., BENDIXEN, C., BERMAN, D., BLECHA, F., BLOMBERG, J., BOLUND, L., BOSSE, M., BOTTI, S., BUJIE, Z., BYSTROM, M., CAPITANU, B., CARVALHO-SILVA, D., CHARDON, P., CHEN, C., CHENG, R., CHOI, S.-H., CHOW, W., CLARK, R. C., CLEE, C., CROOIJMANS, R. P. M. A., DAWSON, H. D., DEHAIS, P., DE SAPIO, F., DIBBITS, B., DROU, N., DU, Z.-Q., EVERSOLE, K., FADISTA, J., FAIRLEY, S., FARAUT, T., FAULKNER, G. J., FOWLER, K. E., FREDHOLM, M., FRITZ, E., GILBERT, J. G. R., GIUFFRA, E., GORODKIN, J., GRIFFIN, D. K., HARROW, J. L., HAYWARD, A., HOWE, K., HU, Z.-L., HUMPHRAY, S. J., HUNT, T., HORNSHØJ, H., JEON, J.-T., JERN, P., JONES, M., JURKA, J., KANAMORI, H., KAPETANOVIC, R., KIM, J., KIM, J.-H., KIM, K.-W., KIM, T.-H., LARSON, G., LEE, K., LEE, K.-T., LEGGETT, R., LEWIN, H. A., LI, Y., LIU, W., LOVELAND, J. E., LU, Y., LUNNEY, J. K., MA, J., MADSEN, O., MANN, K., MATTHEWS, L., MCLAREN, S., MOROZUMI, T., MURTAUGH, M. P., NARAYAN, J., TRUONG NGUYEN, D., NI, P., OH, S.-J., ONTERU, S., PANITZ, F., PARK, E.-W., PARK, H.-S., PASCAL, G., PAUDEL, Y., PEREZ-ENCISO, M., RAMIREZ-GONZALEZ, R., REECY, J. M., RODRIGUEZ-ZAS, S., ROHRER, G. A., RUND, L., SANG, Y., SCHACHTSCHNEIDER, K., SCHRAIBER, J. G., SCHWARTZ, J., SCOBIE, L., SCOTT, C., SEARLE, S., SERVIN, B., SOUTHEY, B. R., SPERBER, G., STADLER, P., SWEEDLER, J. V., TAFER, H., THOMSEN, B., WALI, R., WANG, Ji., WANG, Ju., WHITE, S., XU, X., YERLE, M., ZHANG, G., ZHANG, J., ZHANG, J., ZHAO, S., ROGERS, J., CHURCHER, C. and SCHOOK, L. B. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, **491** (7424), 393–398.
- GRUBB, S. C., CHURCHILL, G. A. and BOGUE, M. A. (2004). A collaborative database of inbred mouse strain characteristics. *Bioinformatics*, **20** (16), 2857–2859.
- GRUBB, S. C., BULT, C. J. and BOGUE, M. A. (2014). Mouse Phenome Database. *Nucleic Acids Research*, **42** (D1), D825–D834.
- GRUFFAT, D., GOBERT, M., DURAND, D. and BAUCHART, D. (2011). Distinct metabolism of linoleic and linolenic acids in liver and adipose tissues of finishing Normande cull cows. *Animal*, **5** (7), 1090–1098.
- GRUFFAT, D., RÉMOND, C., DURAND, D., LOREAU, O. and BAUCHART, D. (2008). 9cis,11trans conjugated linoleic acid (CLA) is synthesised and desaturated into conjugated 18:3 in bovine adipose tissues. *Animal*, **2** (4), 645–652.
- GUÉGUEN, P., REDON, S. and LE MARÉCHAL, C. (2015). Puces à ADN (microArrays) et séquençage de nouvelle génération. *Revue Francophone des Laboratoires*, **2015** (473), 63–70.
- GUILLEMIN, N., CASSAR-MALEK, I., HOCQUETTE, J. F., JURIE, C., MICOL, D., LISTRAT, A., LEVEZIEL, H., RENAND, G. and PICARD, B. (2009). La maîtrise de la tendreté de la viande bovine : identification de marqueurs biologiques. *Productions animales*, **22** (4), 331.
- GUILLEMIN, N., BONNET, M., JURIE, C. and PICARD, B. (2011). Functional analysis of beef tenderness. *Journal of Proteomics*, **75** (2), 352–365.
- GU, J., ORR, N., PARK, S. D., KATZ, L. M., SULIMOVA, G., MACHUGH, D. E. and HILL, E. W. (2009). A Genome Scan for Positive Selection in Thoroughbred Horses. *PLoS ONE*, **4** (6), e5767.
- GU, J., MACHUGH, D. E., MCGIVNEY, B. A., PARK, S. D. E., KATZ, L. M. and HILL, E. W. (2010). Association of sequence variants in CKM (creatine kinase, muscle) and COX4I2 (cytochrome c oxidase, subunit 4, isoform 2) genes with racing performance in Thoroughbred horses : SNP association with elite racing performance. *Equine Veterinary Journal*, **42**, 569–575.
- GULMANN, C., SHEEHAN, K., KAY, E., LIOTTA, L. and PETRICOIN, E. (2006). Array-based proteomics : mapping of protein circuitries for diagnostics, prognostics, and therapy guidance in cancer. *The Journal of Pathology*, **208** (5), 595–606.
- HARTLEY, L. H., MASON, J. W., HOGAN, R. P., JONES, L. G., KOTCHEN, T. A., MOUGEY, E. H.,

- WHERRY, F. E., PENNINGTON, L. L. and RICKETTS, P. T. (1972). Multiple hormonal responses to graded exercise in relation to physical training. *Journal of Applied Physiology*, **33** (5), 602–606.
- HAYES, B. and GODDARD, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome*, **53** (11), 876–883.
- HENZE, A., AUMER, F., GRABNER, A., RAILA, J. and SCHWEIGERT, F. J. (2011). Genetic differences in the serum proteome of horses, donkeys and mules are detectable by protein profiling. *British Journal of Nutrition*, **106** (S1), S170–S173.
- HERSZBERG, B., MCCUE, M. E., LARCHER, T., MATA, X., VAIMAN, A., CHAFFAUX, S., CHÉREL, Y., VALBERG, S. J., MICKELSON, J. R. and GUÉRIN, G. (2009). A GYS1 gene mutation is highly associated with polysaccharide storage myopathy in Cob Normand draught horses. *Animal Genetics*, **40** (1), 94–96.
- HILL, E. W., GU, J., EIVERS, S. S., FONSECA, R. G., MCGIVNEY, B. A., GOVINDARAJAN, P., ORR, N., KATZ, L. M. and MACHUGH, D. (2010). A Sequence Polymorphism in MSTN Predicts Sprinting Ability and Racing Stamina in Thoroughbred Horses. *PLoS ONE*, **5** (1), e8645.
- HILL, E. W., MCGIVNEY, B. A., GU, J., WHISTON, R. and MACHUGH, D. E. (2010). A genome-wide SNP-association study confirms a sequence variant (g. 66493737C> T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. *BMC genomics*, **11** (1), 552.
- HILL, E. W., FONSECA, R. G., MCGIVNEY, B. A., GU, J., MACHUGH, D. E. and KATZ, L. M. (2012). MSTN genotype (g.66493737C/T) association with speed indices in Thoroughbred racehorses. *Journal of Applied Physiology*, **112** (1), 86–90.
- HILL, E. W., GU, J., MCGIVNEY, B. A. and MACHUGH, D. E. (2010). Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance : SNP association with elite racing performance. *Animal Genetics*, **41**, 56–63.
- HINTZ, R. L. (1980). Genetics of Performance in the Horse. *Journal of Animal Science*, **51** (3), 582-594.
- HO, T. K. (1995). Random decision forests. In : *Proceedings of the Third International Conference on Document Analysis and Recognition*, 14-16 August 1995, Montreal, Quebec. [Online] : IEEE, **1**, 278–282. Available from : http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=598994 [Accessed 19 December 2015].
- HOCQUETTE, J.-F., BOUDRA, H., CASSAR-MALEK, I., LEROUX, C., PICARD, B., SAVARY-AUZÉLOUX, I., BERNARD, L., CORNU, A., DURAND, D. and FERLAY, A. (2009). Perspectives offertes par les approches en “omique” pour l’amélioration de la durabilité de l’élevage des herbivores. *Productions animales*, **22** (5), 385.
- HOCQUETTE, J. F., BOUDRA, H., CASSAR-MALEK, I., LEROUX, C., PICARD, B., SAVARY-AUZÉLOUX, I., BERNARD, L., CORNU, A., DURAND, D. and FERLAY, A. (2008). Contribution des approches en «omique» à l’élevage durable des herbivores. *Rencontres autour des recherches sur les ruminants*. 439–442.
- HOCQUETTE, J.-F., LEHNERT, S., BARENDSE, W., CASSAR-MALEK, I. and PICARD, B. (2007). Recent advances in cattle functional genomics and their application to beef quality. *Animal*, **1** (1), 159–173.
- HOLMES, E., FOXALL, P. J.D., SPRAUL, M., DUNCAN FARRANT, R., NICHOLSON, J. K. and LINDON, J. C. (1997). 750 MHz 1H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism : 2-hydroxyglutaric aciduria and maple syrup urine disease. *Journal of Pharmaceutical and Biomedical Analysis*, **15** (11), 1647–1659.
- HONG, H., GOODSID, F., SHI, L. and TONG, W. (2010). Molecular biomarkers : a US FDA effort. *Biomarkers in Medicine*, **4** (2), 215–225.
- HONG, H. and SLIKKER, W. (2015). Advancing translation of biomarkers into regulatory science. *Biomarkers in Medicine*, **9** (11), 1043–1046.
- HONG, H. and TONG, W. (2014). Emerging efforts for discovering new biomarkers of liver disease and

hepatotoxicity. *Biomarkers in Medicine*, **8** (2), 143–146.

HOSSAIN, M. S., JOSHI, T. and STACEY, G. (2015). System approaches to study root hairs as a single cell plant model : current status and future perspectives. *Frontiers in Plant Science*, **6**, 363.

JAIN, A. K., MURTY, M. N. and FLYNN, P. J. (1999). Data clustering : a review. *ACM computing surveys (CSUR)*, **31** (3), 264–323.

JAIN, K. K. (2005). Personalised medicine for cancer : from drug development into clinical practice. *Expert Opinion on Pharmacotherapy*, **6** (9), 1463–1476.

JANSEN, T., FORSTER, P., LEVINE, M. A., OELKE, H., HURLES, M., RENFREW, C., WEBER, J. and OLEK, K. (2002). Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences*, **99** (16), 10905–10910.

JANSSON, E. and KAIJSER, L. (1987). Substrate utilization and enzymes in skeletal muscle of extremely endurance-trained men. *Journal of Applied Physiology*, **62** (3), 999–1005.

JEVSINEK SKOK, D., GODNIC, I., ZORC, M., HORVAT, S., DOVC, P., KOVAC, M. and KUNEJ, T. (2013). Genome-wide in silico screening for microRNA genetic variability in livestock species. *Animal Genetics*, **44** (6), 669–677.

JOSEFSSON, T., LINDWALL, M. and ARCHER, T. (2014). Physical exercise intervention in depressive disorders : Meta-analysis and systematic review. *Scandinavian Journal of Medicine & Science in Sports*, **24** (2), 259–272.

JOYNER, M. J. and COYLE, E. F. (2008). Endurance exercise performance : the physiology of champions : Factors that make champions. *The Journal of Physiology*, **586** (1), 35–44.

KAWAHARA-MIKI, R., SANO, S., NUNOME, M., SHIMMURA, T., KUWAYAMA, T., TAKAHASHI, S., KAWASHIMA, T., MATSUDA, Y., YOSHIMURA, T. and KONO, T. (2013). Next-generation sequencing reveals genomic features in the Japanese quail. *Genomics*, **101** (6), 345–353.

KELLER, M. D., POLLITT, C. C. and MARX, U. C. (2011). Nuclear magnetic resonance-based metabonomic study of early time point laminitis in an oligofructose-overload model : Metabolic changes in an early time point laminitis model. *Equine Veterinary Journal*, **43** (6), 737–743.

KELLER, P., VOLLAARD, N. B. J., GUSTAFSSON, T., GALLAGHER, I. J., SUNDBERG, C. J., RANKINEN, T., BRITTON, S. L., BOUCHARD, C., KOCH, L. G. and TIMMONS, J. A. (2011). A transcriptional map of the impact of endurance exercise training on skeletal muscle phenotype. *Journal of Applied Physiology*, **110** (1), 46–59.

KIM, H. J., LEE, Y. H. and KIM, C. K. (2009). Changes in serum cartilage oligomeric matrix protein (COMP), plasma CPK and plasma hs-CRP in relation to running distance in a marathon (42.195 km) and an ultra-marathon (200 km) race. *European Journal of Applied Physiology*, **105** (5), 765–770.

KIM, M.-C., LEE, S.-W., RYU, D.-Y., CUI, F.-J., BHAK, J. and KIM, Y. (2014). Identification and Characterization of MicroRNAs in Normal Equine Tissues by Next Generation Sequencing. *PLoS ONE*, **9** (4), e93662.

KŁAPCIŃSKA, B., WAŚKIEWICZ, Z., CHRAPUSTA, S. J., SADOWSKA-KRĘPA, E., CZUBA, M. and LANGFORT, J. (2013). Metabolic responses to a 48-h ultra-marathon run in middle-aged male amateur runners. *European Journal of Applied Physiology*, **113** (11), 2781–2793.

KLOSE, J. and KOBALZ, U. (1995). Two-dimensional electrophoresis of proteins : an updated protocol and implications for a functional analysis of the genome. *Electrophoresis*, **16** (6), 1034–1059.

KOBOLDT, D. C., STEINBERG, K. M., LARSON, D. E., WILSON, R. K. and MARDIS, E. R. (2013). The Next-Generation Sequencing Revolution and Its Impact on Genomics. *Cell*, **155** (1), 27–38.

- KOUZARIDES, T. (2007). Chromatin Modifications and Their Function. *Cell*, **128** (4), 693–705.
- KRAUS, R. H.S., KERSTENS, H. H.D., VAN HOOFT, P., CROOIJMANS, R. P.M.A., VAN DER POEL, J. J., ELMBERG, J., VIGNAL, A., HUANG, Y., LI, N. and PRINS, H. H.T. (2011). Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). *BMC genomics*, **12** (1), 1.
- KULBROCK, M., LEHNER, S., METZGER, J., OHNESORGE, B. and DISTL, O. (2013). A Genome-Wide Association Study Identifies Risk Loci to Equine Recurrent Uveitis in German Warmblood Horses. *PLoS ONE*, **8** (8), e71619.
- KUTTER, C. and SVOBODA, P. (2008). miRNA, siRNA, piRNA : Knowns of the unknown. *RNA Biology*, **5** (4), 181–188.
- LAFENETRE, P., LESKE, O., WAHLE, P. and HEUMANN, R. (2011). The Beneficial Effects of Physical Activity on Impaired Adult Neurogenesis and Cognitive Performance. *Frontiers in Neuroscience*, **5** (51).
- LAMORIL, J., AMEZIANE, N., DEYBACH, J.-C., BOUIZEGARÈNE, P. and BOGARD, M. (2008). Les techniques de séquençage de l'ADN : une révolution en marche. Première partie. *Immuno-analyse & Biologie Spécialisée*, **23** (5), 260–279.
- LAMORIL, J., AMEZIANE, N., DEYBACH, J.-C., BOUIZEGARÈNE, P. and BOGARD, M. (2010a). Le monde complexe et mouvant des ARN. Première partie. *Immuno-analyse & Biologie Spécialisée*, **25** (1), 4–25.
- LAMORIL, J., BOUIZEGARÈNE, P. and BOGARD, M. (2010b). Le monde complexe et mouvant des ARN. Seconde partie : les microARNs. *Immuno-analyse & Biologie Spécialisée*, **25** (5-6), 219–240.
- LAMORIL, J., BOUIZEGARÈNE, P. and BOGARD, M. (2011). Le monde complexe et mouvant des ARN. Troisième partie. *Immuno-analyse & Biologie Spécialisée*, **26**, 225–238.
- LANGLOIS, B. (1980). Heritability of racing ability in thoroughbreds — A review. *Livestock Production Science*, **7** (6), 591–605.
- LE BAIL, P.-Y., BUGEON, J., DAMERON, O., FATET, A., GOLIK, W., HOCQUETTE, J.-F., HURTAUD, C., HUE, I., JONDREVILLE, C., JORET, L., MEUNIER-SALAÜN, M.-C., VERNET, J., NEDELLEC, C., REICHSTADT, M., CHEMINEAU, P. (2014). Un langage de référence pour le phénotypage des animaux d'élevage : l'ontologie ATOL. *INRA Productions Animales*, **27** (3), 195–208.
- LEE, J. T. (2012). Epigenetic regulation by long noncoding RNAs. *Science*, **338** (6113), 1435–1439.
- LE GALL, G., NOOR, S. O., RIDGWAY, K., SCOVELL, L., JAMIESON, C., JOHNSON, I. T., COLQUHOUN, I. J., KEMSLEY, E. K. and NARBAD, A. (2011). Metabolomics of Fecal Extracts Detects Altered Metabolic Activity of Gut Microbiota in Ulcerative Colitis and Irritable Bowel Syndrome. *Journal of Proteome Research*, **10** (9), 4208–4218.
- LE MOYEC, L., MILLE-HAMARD, L., TRIBA, M. N., BREUNEVAL, C., PETOT, H. and BILLAT, V. L. (2012). NMR metabolomics for assessment of exercise effects with mouse biofluids. *Analytical and Bioanalytical Chemistry*, **404** (2), 593–602.
- LE MOYEC, L., ROBERT, C., TRIBA, M. N., BILLAT, V. L., MATA, X., SCHIBLER, L. and BARREY, E. (2014). Protein Catabolism and High Lipid Metabolism Associated with Long-Distance Exercise Are Revealed by Plasma NMR Metabolomics in Endurance Horses. *PLoS ONE*, **9** (3), e90730.
- LEVIN, G. (2014). *Role of physical exercise in reducing depression and improving mental health in cancer survivors*. Thesis for the degree of Doctor of Philosophy, Faculty of Health, Engineering and Science, Edith Cowan University, Joondalup, Western Australia, 221 p.
- LIETO, L.D. and COTHRAN, E.G. (2003). The epitheliogenesis imperfecta locus maps to equine chromosome 8 in American Saddlebred horses. *Cytogenetic and Genome Research*, **102** (1-4), 207–210.
- LI, J., WIJFFELS, G., YU, Y., NIELSEN, L. K., NIEMEYER, D. O., FISHER, A. D., FERGUSON, D. M. and

SCHIRRA, H. J. (2011). Altered Fatty Acid Metabolism in Long Duration Road Transport : An NMR-based Metabonomics Study in Sheep. *Journal of Proteome Research*, **10** (3), 1073–1087.

LINDBLAD-TOH, K., GARBER, M., ZUK, O., LIN, M. F., PARKER, B. J., WASHIETL, S., KHERADPOUR, P., ERNST, J., JORDAN, G., MAUCELI, E., WARD, L. D., LOWE, C. B., HOLLOWAY, A. K., CLAMP, M., GNERRE, S., ALFÖLDI, J., BEAL, K., CHANG, J., CLAWSON, H., CUFF, J., DI PALMA, F., FITZGERALD, S., FLICEK, P., GUTTMAN, M., HUBISZ, M. J., JAFFE, D. B., JUNGREIS, I., KENT, W. J., KOSTKA, D., LARA, M., MARTINS, A. L., MASSINGHAM, T., MOLTKE, I., RANEY, B. J., RASMUSSEN, M. D., ROBINSON, J., STARK, A., VILELLA, A. J., WEN, J., XIE, X., ZODY, M. C., BALDWIN, J., BLOOM, T., WHYE CHIN, C., HEIMAN, D., NICOL, R., NUSBAUM, C., YOUNG, S., WILKINSON, J., WORLEY, K. C., KOVAR, C. L., MUZNY, D. M., GIBBS, R. A., CREE, A., DIHN, H. H., FOWLER, G., JHANGIANI, S., JOSHI, V., LEE, S., LEWIS, L. R., NAZARETH, L. V., OKWUONU, G., SANTIBANEZ, J., WARREN, W. C., MARDIS, E. R., WEINSTOCK, G. M., WILSON, R. K., DELEHAUNTY, K., DOOLING, D., FRONIK, C., FULTON, L., FULTON, B., GRAVES, T., MINX, P., SODERGREN, E., BIRNEY, E., MARGULIES, E. H., HERRERO, J., GREEN, E. D., HAUSSLER, D., SIEPEL, A., GOLDMAN, N., POLLARD, K. S., PEDERSEN, J. S., LANDER, E. S. and KELLIS, M. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, **478** (7370), 476–482.

LINDBLAD-TOH, K., WADE, C. M., MIKKELSEN, T. S., KARLSSON, E. K., JAFFE, D. B., KAMAL, M., CLAMP, M., CHANG, J. L., KULBOKAS, E. J., ZODY, M. C., MAUCELI, E., XIE, X., BREEN, M., WAYNE, R. K., OSTRANDER, E. A., PONTING, C. P., GALIBERT, F., SMITH, D. R., DEJONG, P. J., KIRKNESS, E., ALVAREZ, P., BIAGI, T., BROCKMAN, W., BUTLER, J., CHIN, C.-W., COOK, A., CUFF, J., DALY, M. J., DECAPRIO, D., GNERRE, S., GRABHERR, M., KELLIS, M., KLEBER, M., BARDELEBEN, C., GOODSTADT, L., HEGER, A., HITTE, C., KIM, L., KOEPFLI, K.-P., PARKER, H. G., POLLINGER, J. P., SEARLE, S. M. J., SUTTER, N. B., THOMAS, R., WEBBER, C., BROAD INSTITUTE GENOME SEQUENCING PLATFORM and LANDER, E. S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438** (7069), 803–819.

LIU, J., JENNINGS, S. F., TONG, W. and HONG, H. (2011). Next generation sequencing for profiling expression of miRNAs : technical progress and applications in drug development. *Journal of Biomedical Science and Engineering*, **4** (10), 666–676.

LIU, Y., UBOH, C. E., SOMA, L. R., LI, X., GUAN, F., YOU, Y. and CHEN, J.-W. (2011). Efficient Use of Retention Time for the Analysis of 302 Drugs in Equine Plasma by Liquid Chromatography-MS/MS with Scheduled Multiple Reaction Monitoring and Instant Library Searching for Doping Control. *Analytical Chemistry*, **83** (17), 6834–6841.

LIU, Y., DEVESCOVI, V., CHEN, S. and NARDINI, C. (2013). Multilevel omic data integration in cancer cell lines : advanced annotation and emergent properties. *BMC systems biology*, **7** (1), 14.

LOOR, J. J., DANN, H. M., GURETZKY, N. A. J., EVERTS, R. E., OLIVEIRA, R., GREEN, C. A., LITHERLAND, N. B., RODRIGUEZ-ZAS, S. L., LEWIN, H. A. and DRACKLEY, J. K. (2006). Plane of nutrition prepartum alters hepatic gene expression and function in dairy cows as assessed by longitudinal transcript and metabolic profiling. *Physiological Genomics*, **27** (1), 29–41.

LUCK, M. M., LE MOYEC, L., BARREY, E., TRIBA, M. N., BOUCHEMAL, N., SAVARIN, P. and ROBERT, C. (2015). Energetics of endurance exercise in young horses determined by nuclear magnetic resonance metabolomics. *Frontiers in Physiology*, **6**, 198.

MACH, N., JACOBS, A. A. A., KRUIJT, L., VAN BAAL, J. and SMITS, M. A. (2011). Alteration of gene expression in mammary gland tissue of dairy cows in response to dietary unsaturated fatty acids. *Animal*, **5** (8), 1217–1230.

MACH, N., KEUNING, E., KRUIJT, L., HORTOS, M., ARNAU, J. and TE PAS, M. F. W. (2010). Comparative proteomic profiling of 2 muscles from 5 different pure pig breeds using surface-enhanced laser desorption/ionization time-of-flight proteomics technology. *Journal of Animal Science*, **88** (4), 1522–1534.

MACH, N., BERRI, M., ESTELLÉ, J., LEVENEZ, F., LEMONNIER, G., DENIS, C., LEPLAT, J.-J., CHEVALEYRE, C., BILLON, Y., DORÉ, J., ROGEL-GAILLARD, C. and LEPAGE, P. (2015). Early-life establishment of the swine gut microbiome and impact on host phenotypes : Role of early-life gut microbiome on

pigs' health. *Environmental Microbiology Reports*, **7** (3), 554–569.

MACH, N., VAN BAAL, J., KRUIJT, L., JACOBS, A. and SMITS, M. (2011). Dietary unsaturated fatty acids affect the mammary gland integrity and health in lactating dairy cows. In : *BMC Proceedings, International Symposium on Animal Genomics for Animal Health*, 31 May - 2 June 2010, Paris, France. BioMed Central Ltd, 2011, **5** (S4), S35.

MACH, N., ZOM, R. L. G., WIDJAJA, H. C. A., VAN WIKSELAAR, P. G., WEURDING, R. E., GOSELINK, R. M. A., VAN BAAL, J., SMITS, M. A. and VAN VUUREN, A. M. (2013). Dietary effects of linseed on fatty acid composition of milk and on liver, adipose and mammary gland metabolism of periparturient dairy cows. *Journal of Animal Physiology and Animal Nutrition*, **97**, 89–104.

MADDATU, T. P., GRUBB, S. C., BULT, C. J. and BOGUE, M. A. (2012). Mouse Phenome Database (MPD). *Nucleic Acids Research*, **40** (D1), D887–D894.

MAKVANDI-NEJAD, S., HOFFMAN, G. E., ALLEN, J. J., CHU, E., GU, E., CHANDLER, A. M., LOREDO, A. I., BELLONE, R. R., MEZEY, J. G., BROOKS, S. A. and SUTTER, N. B. (2012). Four Loci Explain 83% of Size Variation in the Horse. *PLoS ONE*, **7** (7), e39929.

MALLON, A.-M., BLAKE, A. and HANCOCK, J. M. (2007). EuroPhenome and EMPReSS : online mouse phenotyping resource. *Nucleic Acids Research*, **36**, Database Issue, D715–D718.

MARCOS, B., GOU, P., GUÀRDIA, M.D., HORTÓS, M., COLLEO, M., MACH, N., TE PAS, M.F.W., KEUNING, E., KRUIJT, L., TIBAU, J., GISPERT, M. and ARNAU, J. (2013). Surface-enhanced laser desorption/ionisation time-of-flight mass spectrometry : A tool to predict pork quality. *Meat Science*, **95** (3), 688–693.

MARTIN, A.-M., ELLIOTT, J. A., DUFFY, P., BLAKE, C. M., ATTIA, S. B., KATZ, L. M., BROWNE, J. A., GATH, V., MCGIVNEY, B. A., HILL, E. W. and MURPHY, B. A. (2010). Circadian regulation of locomotor activity and skeletal muscle gene expression in the horse. *Journal of Applied Physiology*, **109** (5), 1328–1336.

MASCIA, G., PERROTTA, L., GALANTI, G. and PADELETTI, L. (2012). Atrial Fibrillation in Athletes. *International Journal of Sports Medicine*, **34** (5), 379–384.

MATTES, J., COLLISON, A., PLANK, M., PHIPPS, S. and FOSTER, P. S. (2009). Antagonism of microRNA-126 suppresses the effector function of TH2 cells and the development of allergic airways disease. *Proceedings of the National Academy of Sciences*, **106** (44), 18704–18709.

MATTICK, J. S. (2009). The Genetic Signatures of Noncoding RNAs. *PLoS Genetics*, **5** (4), e1000459.

MATTICK, J. S. and MAKUNIN, I. V. (2006). Non-coding RNA. *Human Molecular Genetics*, **15** (90001), R17–R29.

MCCOY, A. M., SCHAEFER, R., PETERSEN, J. L., MORRELL, P. L., SLAMKA, M. A., MICKELSON, J. R., VALBERG, S. J. and MCCUE, M. E. (2014). Evidence of Positive Selection for a Glycogen Synthase (GYS1) Mutation in Domestic Horse Populations. *Journal of Heredity*, **105** (2), 163–172.

MCCUE, M.E., VALBERG, S.J., JACKSON, M., BORGIA, L., LUCIO, M. and MICKELSON, J.R. (2009). Polysaccharide storage myopathy phenotype in quarter horse-related breeds is modified by the presence of an RYR1 mutation. *Neuromuscular Disorders*, **19** (1), 37–43.

MCCUE, M.E., VALBERG, S.J., LUCIO, M. and MICKELSON, J.R. (2008). Glycogen Synthase 1 (GYS1) Mutation in Diverse Breeds with Polysaccharide Storage Myopathy. *Journal of Veterinary Internal Medicine*, **22** (5), 1228–1233.

MCCUE, M. E., BANNASCH, D. L., PETERSEN, J. L., GURR, J., BAILEY, E., BINNS, M. M., DISTL, O., GUÉRIN, G., HASEGAWA, T., HILL, E. W., LEEB, T., LINDGREN, G., PENEDO, M. C. T., RØED, K. H., RYDER, O. A., SWINBURNE, J. E., TOZAKI, T., VALBERG, S. J., VAUDIN, M., LINDBLAD-TOH, K., WADE, C. M. and MICKELSON, J. R. (2012). A High Density SNP Array for the Domestic Horse and Extant Perissodactyla : Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies. *PLoS Genetics*, **8** (1),

e1002451.

MCCUE, M. E., VALBERG, S. J., MILLER, M. B., WADE, C., DIMAURO, S., AKMAN, H. O. and MICKELSON, J. R. (2008). Glycogen synthase (GYS1) mutation causes a novel skeletal muscle glycogenosis. *Genomics*, **91** (5), 458–466.

MCGIVNEY, B. A., BROWNE, J. A., FONSECA, R. G., KATZ, L. M., MACHUGH, D. E., WHISTON, R. and HILL, E. W. (2012). MSTN genotypes in Thoroughbred horses influence skeletal muscle gene expression and racetrack performance. *Animal Genetics*, **43** (6), 810–812.

MCGIVNEY, B. A., MCGETTIGAN, P. A., BROWNE, J. A., EVANS, A. C.O., FONSECA, R. G., LOFTUS, B. J., LOHAN, A., MACHUGH, D. E., MURPHY, B. A., KATZ, L. M. and HILL, E. W. (2010). Characterization of the equine skeletal muscle transcriptome identifies novel functional responses to exercise training. *BMC genomics*, **11** (1), 398.

MEISSNER, A., MIKKELSEN, T. S., GU, H., WERNIG, M., HANNA, J., SIVACHENKO, A., ZHANG, X., BERNSTEIN, B. E., NUSBAUM, C., JAFFE, D. B., GNIRKE, A., JAENISCH, R. and LANDER, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454** (7205), 766–770.

MERCER, T. R., DINGER, M. E. and MATTICK, J. S. (2009). Long non-coding RNAs : insights into functions. *Nature Reviews Genetics*, **10** (3), 155–159.

MERCER, T. R. and MATTICK, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural & Molecular Biology*, **20** (3), 300–307.

MERRIFIELD, C. A., LEWIS, M., CLAUS, S. P., BECKONERT, O. P., DUMAS, M.-E., DUNCKER, S., KOCHHAR, S., REZZI, S., LINDON, J. C., BAILEY, M., HOLMES, E. and NICHOLSON, J. K. (2011). A metabolic system-wide characterisation of the pig : a model for human physiology. *Molecular BioSystems*, **7** (9), 2577.

METALLINOS, D.L., BOWLING, A.T. and RINE, J.A. (1998). A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome : an equine version of Hirschsprung Disease. *Mammalian Genome*, **9** (6), 426–431.

METZGER, J., OHNESORGE, B. and DISTL, O. (2012). Genome-Wide Linkage and Association Analysis Identifies Major Gene Loci for Guttural Pouch Tympany in Arabian and German Warmblood Horses. *PLoS ONE*, **7** (7), e41640.

METZGER, J., SCHRIMPF, R., PHILIPP, U. and DISTL, O. (2013). Expression levels of LCORL are associated with body size in horses. *PLoS ONE*, **8** (2), e56497.

MIENALTOWSKI, M. J., HUANG, L., FRISBIE, D. D., MCILWRAITH, C. W., STROMBERG, A. J, BATHKE, A. C. and MACLEOD, J. N. (2009). Transcriptional profiling differences for articular cartilage and repair tissue in equine joint surface lesions. *BMC Medical Genomics*, **2** (1), 60.

MIKKELSEN, T. S., KU, M., JAFFE, D. B., ISSAC, B., LIEBERMAN, E., GIANNOUKOS, G., ALVAREZ, P., BROCKMAN, W., KIM, T.-K., KOICHE, R. P., LEE, W., MENDENHALL, E., O'DONOVAN, A., PRESSER, A., RUSS, C., XIE, X., MEISSNER, A., WERNIG, M., JAENISCH, R., NUSBAUM, C., LANDER, E. S. and BERNSTEIN, B. E. (2007b). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448** (7153), 553–560.

MIKKELSEN, T. S., WAKEFIELD, M. J., AKEN, B., AMEMIYA, C. T., CHANG, J. L., DUKE, S., GARBER, M., GENTLES, A. J., GOODSTADT, L., HEGER, A., JURKA, J., KAMAL, M., MAUCELI, E., SEARLE, S. M. J., SHARPE, T., BAKER, M. L., BATZER, M. A., BENOS, P. V., BELOV, K., CLAMP, M., COOK, A., CUFF, J., DAS, R., DAVIDOW, L., DEAKIN, J. E., FAZZARI, M. J., GLASS, J. L., GRABHERR, M., GREALLY, J. M., GU, W., HORE, T. A., HUTTLEY, G. A., KLEBER, M., JIRTLE, R. L., KOINA, E., LEE, J. T., MAHONY, S., MARRA, M. A., MILLER, R. D., NICHOLLS, R. D., ODA, M., PAPPENFUSS, A. T., PARRA, Z. E., POLLOCK, D. D., RAY, D. A., SCHEIN, J. E., SPEED, T. P., THOMPSON, K., VANDEBERG, J. L., WADE, C. M., WALKER, J. A., WATERS, P. D., WEBBER, C., WEIDMAN, J. R., XIE, X., ZODY, M. C., BROAD INSTITUTE GENOME SEQUENCING PLATFORM, BROAD INSTITUTE WHOLE GENOME ASSEMBLY

- TEAM, GRAVES, J. A. M., PONTING, C. P., BREEN, M., SAMOLLOU, P. B., LANDER, E. S. and LINDBLAD-TOH, K. (2007a). Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature*, **447** (7141), 167–177.
- MILENKOVIC, D., CHAFFAUX, S., TAOURIT, S. and GUERIN, G. (2003). A mutation in the *LAMC2* gene causes the Herlitz junctional epidermolysis bullosa (H-JEB) in two French draft horse breeds. *Genetics Selection Evolution*, **35** (2), 249–256.
- MILES, L. and CASWELL, H. (2008). Advancing beef safety and quality : ProSafeBeef. *Nutrition Bulletin*, **33** (2), 140–144.
- MILLER, I., FRIEDLEIN, A., TSANGARIS, G., MARIS, A., FOUNTOULAKIS, M. and GEMEINER, M. (2004). The serum proteome of *Equus caballus*. *Proteomics*, **4** (10), 3227–3234.
- MOBUCHON, L., MARTHEY, S., LE GUILLOU, S., LALOË, D., LE PROVOST, F. and LEROUX, C. (2015). Food Deprivation Affects the miRNome in the Lactating Goat Mammary Gland. *PLoS ONE*, **10** (10), e0140111.
- MOBUCHON, L., MARTHEY, S., LE GUILLOU, S., LALOË, D., LEROUX, C. and LE PROVOST, F. (2014). Nutritional regulation of microRNAs in the lactating goat mammary gland. In : *Gordon Research Conference on Mammary Gland Biology*, 8-13 June 2014, Lucca, Italy [Online]. Available from : <https://hal.archives-ouvertes.fr/hal-01193895/> [Accessed 16 December 2015].
- MONTHOUX, C., DE BROU, S., JACKSON, M., BLEUL, U. and WALTER, J. (2015). Skin malformations in a neonatal foal tested homozygous positive for Warmblood Fragile Foal Syndrome. *BMC Veterinary Research*, **11** (1), 12.
- MOOREN, F. C., VIREECK, J., KRUGER, K. and THUM, T. (2014). Circulating micrnas as potential biomarkers of aerobic exercise capacity. *AJP : Heart and Circulatory Physiology*, **306** (4), H557–H563.
- MOORE, R. E, KNOTTENBELT, D., MATTHEWS, J. B, BEYNON, R. J and WHITFIELD, P. D (2008). Biomarkers for ragwort poisoning in horses : identification of protein targets. *BMC Veterinary Research*, **4** (1), 30.
- MORZEL, M., TERLOUW, C., CHAMBON, C., MICOL, D. and PICARD, B. (2008). Muscle proteome and meat eating qualities of *Longissimus thoracis* of “Blonde d’Aquitaine” young bulls : A central role of HSP27 isoforms. *Meat Science*, **78** (3), 297–304.
- MURPHY, D. (2002). GENE EXPRESSION STUDIES USING MICROARRAYS : PRINCIPLES, PROBLEMS, AND PROSPECTS. *Advances in Physiology Education*, **26** (4), 256–270.
- NAGANO, T. and FRASER, P. (2011). No-Nonsense Functions for Long Noncoding RNAs. *Cell*, **145** (2), 178–181.
- NAGARAJ, S. H., GASSER, R. B. and RANGANATHAN, S. (2006). A hitchhiker’s guide to expressed sequence tag (EST) analysis. *Briefings in Bioinformatics*, **8** (1), 6–21.
- NETTO, G. J. and EPSTEIN, J. I. (2010). Theranostic and prognostic biomarkers : genomic applications in urological malignancies. *Pathology*, **42** (4), 384–394.
- NICHOLSON, J. K., FOXALL, P. J.D., SPRAUL, M., FARRANT, R. D. and LINDON, J. C. (1995). 750 MHz ¹H and ¹H-¹³C NMR spectroscopy of human blood plasma. *Analytical chemistry*, **67** (5), 793–811.
- NIETO, J.E. and ALEMAN, M. (2009). A Rapid Detection Method for the Ryanodine Receptor 1 (C7360G) Mutation in Quarter Horses. *Journal of Veterinary Internal Medicine*, **23** (3), 619–622.
- NOSCHKA, E., VANDENPLAS, M. L., HURLEY, D. J. and MOORE, J. N. (2009). Temporal aspects of laminar gene expression during the developmental stages of equine laminitis. *Veterinary Immunology and Immunopathology*, **129** (3-4), 242–253.
- NYBERG, N. T., NIELSEN, M. O. and JAROSZEWSKI, J. W. (2010). Metabolic trajectories based on ¹H NMR

- spectra of urines from sheep exposed to nutritional challenges during prenatal and early postnatal life. *Metabolomics*, **6** (4), 489–496.
- O'CONNOR, A. and WYNESS, L. (2012). A summary of findings from the 5-year ProSafeBeef project : The ProSafeBeef project. *Nutrition Bulletin*, **37** (1), 72–77.
- O'KEEFE, J. H., PATIL, H. R., LAVIE, C. J., MAGALSKI, A., VOGEL, R. A. and MCCULLOUGH, P. A. (2012). Potential Adverse Cardiovascular Effects From Excessive Endurance Exercise. *Mayo Clinic Proceedings*, **87** (6), 587–595.
- OLIVEIRA-CARVALHO, V., DA SILVA, M. M. F., GUIMARÃES, G. V., BACAL, F. and BOCCHI, E. A. (2013). MicroRNAs : new players in heart failure. *Molecular Biology Reports*, **40** (3), 2663–2670.
- OLIVER, S., WINSON, M., KELL, D. and BAGANZ, F. (1998). Systematic functional analysis of the yeast genome. *Trends in biotechnology*, **16** (9), 373–378.
- ORLANDO, L., GINOLHAC, A., ZHANG, G., FROESE, D., ALBRECHTSEN, A., STILLER, M., SCHUBERT, M., CAPPELLINI, E., PETERSEN, B., MOLTKE, I., JOHNSON, P. L. F., FUMAGALLI, M., VILSTRUP, J. T., RAGHAVAN, M., KORNELIUSSEN, T., MALASPINAS, A.-S., VOGT, J., SZKLARCZYK, D., KELSTRUP, C. D., VINTHER, J., DOLOCAN, A., STENDERUP, J., VELAZQUEZ, A. M. V., CAHILL, J., RASMUSSEN, M., WANG, X., MIN, J., ZAZULA, G. D., SEGUIN-ORLANDO, A., MORTENSEN, C., MAGNUSSEN, K., THOMPSON, J. F., WEINSTOCK, J., GREGERSEN, K., ROED, K. H., EISENMANN, V., RUBIN, C. J., MILLER, D. C., ANTCZAK, D. F., BERTELSEN, M. F., BRUNAK, S., AL-RASHEID, K. A. S., RYDER, O., ANDERSSON, L., MUNDY, J., KROGH, A., GILBERT, M. T. P., KJAER, K., SICHERITZ-PONTEN, T., JENSEN, L. J., OLSEN, J. V., HOFREITER, M., NIELSEN, R., SHAPIRO, B., WANG, J. and WILLERSLEV, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, **499** (7456), 74–78.
- OZSOLAK, F. and MILOS, P. M. (2011). RNA sequencing : advances, challenges and opportunities. *Nature Reviews Genetics*, **12** (2), 87–98.
- PAILHOUX, E., CRIBIU, E. P., PARMA, P. and COTINOT, C. (1995). Molecular Analysis of an XY Mare with Gonadal Dysgenesis. *Hereditas*, **122** (2), 109–112.
- PANGANIBAN, R. P., PINKERTON, M. H., MARU, S. Y., JEFFERSON, S. J., ROFF, A. N. and ISHMAEL, F. T. (2012). Differential microRNA expression in asthma and the role of miR-1248 in regulation of IL-5. *American Journal of Clinical and Experimental Immunology*, **1** (2), 154–165.
- PAN, Z. and RAFTERY, D. (2007). Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Analytical and Bioanalytical Chemistry*, **387** (2), 525–527.
- PAPPALARDO, L., PELCZER, I. and RALSTON, S. L. (2013). Metabolic Differences Between Draft-cross and Mustang Horses Detected by Metabonomic Analyses. *Journal of Equine Veterinary Science*, **33** (12), 1044–1049.
- PECHLIVANIS, A., KOSTIDIS, S., SARASLANIDIS, P., PETRIDOU, A., TSALIS, G., VESELKOV, K., MIKROS, E., MOUGIOS, V. and THEODORIDIS, G. A. (2013). ¹H NMR Study on the Short- and Long-Term Impact of Two Training Programs of Sprint Running on the Metabolic Fingerprint of Human Serum. *Journal of Proteome Research*, **12** (1), 470–480.
- PEDERSEN, B. K. and FISCHER, C. P. (2007). Beneficial health effects of exercise – the role of IL-6 as a myokine. *Trends in Pharmacological Sciences*, **28** (4), 152–156.
- PEDERSEN, B. K. and HOFFMAN-GOETZ, L. (2000). Exercise and the immune system : regulation, integration, and adaptation. *Physiological reviews*, **80** (3), 1055–1081.
- PEDERSEN, B. K. and FEBBRAIO, M. A. (2008). Muscle as an Endocrine Organ : Focus on Muscle-Derived Interleukin-6. *Physiological Reviews*, **88** (4), 1379–1406.
- PEDERSEN, B. K., STEENSBERG, A., FISCHER, C., KELLER, C., OSTROWSKI, K. and SCHJERLING, P. (2001). Exercise and cytokines with particular focus on muscle derived IL-6. *Exercise immunology review*, **7**, 18–

PEDERSEN, B. K., STEENSBERG, A., FISCHER, C., KELLER, C., KELLER, P., PLOMGAARD, P., FEBBRAIO, M. and SALTIN, B. (2003). Searching for the exercise factor : is IL-6 a candidate ? *Journal of Muscle Research & Cell Motility*, **24** (2-3), 113–119.

PEDERSEN, B. K., STEENSBERG, A., FISCHER, C., KELLER, C., KELLER, P., PLOMGAARD, P., WOLSK-PETERSEN, E. and FEBBRAIO, M. (2004). The metabolic role of IL-6 produced during exercise : is IL-6 an exercise factor ? *Proceedings of the Nutrition Society*, **63** (2), 263–267.

PEDERSEN, B. K. (2009). Edward F. Adolph Distinguished Lecture : Muscle as an endocrine organ : IL-6 and other myokines. *Journal of Applied Physiology*, **107** (4), 1006–1014.

PELECHANO, V. and STEINMETZ, L. M. (2013). Gene regulation by antisense transcription. *Nature Reviews Genetics*, **14** (12), 880–893.

PETERSEN, A. and PEDERSEN, B. (2006). The role of IL-6 in mediating the anti inflammatory. *Journal of Physiology and Pharmacology*, **57** (S10), 43–51.

PETERSEN, J. L., MICKELSON, J. R., RENDAHL, A. K., VALBERG, S. J., ANDERSSON, L. S., AXELSSON, J., BAILEY, E., BANNASCH, D., BINNS, M. M., BORGES, A. S., BRAMA, P., DA CÂMARA MACHADO, A., CAPOMACCIO, S., CAPPELLI, K., COTHRAN, E. G., DISTL, O., FOX-CLIPSHAM, L., GRAVES, K. T., GUÉRIN, G., HAASE, B., HASEGAWA, T., HEMMANN, K., HILL, E. W., LEEB, T., LINDGREN, G., LOHI, H., LOPES, M. S., MCGIVNEY, B. A., MIKKO, S., ORR, N., PENEDO, M. C. T., PIERCY, R. J., RAEKALLIO, M., RIEDER, S., RØED, K. H., SWINBURNE, J., TOZAKI, T., VAUDIN, M., WADE, C. M. and MCCUE, M. E. (2013). Genome-Wide Analysis Reveals Selection for Important Traits in Domestic Horse Breeds. *PLoS Genetics*, **9** (1), e1003211.

PINAUDEAU, D. (2015). Séquenceurs haut débit. *IRBM News*, **36** (6), 191–195.

PONTING, C. P., OLIVER, P. L. and REIK, W. (2009). Evolution and Functions of Long Noncoding RNAs. *Cell*, **136** (4), 629–641.

PONTIUS, J. U., MULLIKIN, J. C., SMITH, D. R., AGENCOURT SEQUENCING TEAM, LINDBLAD-TOH, K., GNERRE, S., CLAMP, M., CHANG, J., STEPHENS, R., NEELAM, B., VOLFOVSKY, N., SCHAFFER, A. A., AGARWALA, R., NARFSTROM, K., MURPHY, W. J., GIGER, U., ROCA, A. L., ANTUNES, A., MENOTTI-RAYMOND, M., YUHKI, N., PECON-SLATTEY, J., JOHNSON, W. E., BOURQUE, G., TESLER, G., NISC COMPARATIVE SEQUENCING PROGRAM and O'BRIEN, S. J. (2007). Initial sequence and comparative analysis of the cat genome. *Genome Research*, **17** (11), 1675–1689.

PRACHE, S., MARTIN, B., NOZIÈRE, P., ENGEL, E., BESLE, J. M., FERLAY, A., MICOL, D., CORNU, A., CASSAR-MALEK, I. and ANDUEZA, D. (2007). Authentification de l'alimentation des ruminants à partir de la composition de leurs produits et tissus. *INRA Productions Animales*, **20** (4), 295-308.

PUJAR, S. and MEYERS-WALLEN, V. (2012). Sequence Variations in Equine Candidate Genes For XX and XY Inherited Disorders of Sexual Development : Equine DSD Candidate Gene Variants. *Reproduction in Domestic Animals*, **47** (5), 827–834.

PURVIS, D., GONSALVES, S. and DEUSTER, P. A. (2010). Physiological and Psychological Fatigue in Extreme Conditions : Overtraining and Elite Athletes. *PM&R*, **2** (5), 442–450.

RALSTON, S. L., PAPPALARDO, L. and PELCZER, I. (2011). Breed and age effects on metabolic profiles of young horses using NMR-based Metabonomic analyses of serum. *Journal of Equine Veterinary Science*, **31** (5), 304–305.

RALSTON, S. L., PELCZER, I., ZIEGLER, K., BAEZ, E. and SPEARS, P. F. (2009). NMR-based metabonomic analysis of serum from Standardbred yearlings with or without hock osteochondrosis dissecans lesions : metabolic markers for OCD. In : *Applied Equine Nutrition and Training : Equine NUTrition and TRAINing CONference (ENUTRACO) 2009*, 26-29 September 2009, Madrid, Spain. Wageningen, Netherlands : Arno Lindner (Wageningen Academic Publishers), p. 215.

- RANKINEN, T., BRAY, M. S., HAGBERG, J. M., PÉRUSSE, L., ROTH, S. M., WOLFARTH, B. and BOUCHARD, C. (2006). The Human Gene Map for Performance and Health-Related Fitness Phenotypes : The 2005 Update. *Medicine & Science in Sports & Exercise*, **38** (11), 1863–1888.
- RASSOULZADEGAN, M., GRANDJEAN, V., GOUNON, P., VINCENT, S., GILLOT, I. and CUZIN, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature*, **441** (7092), 469–474.
- RAUDSEPP, T., DURKIN, K., LEAR, T. L., DAS, P. J., AVILA, F., KACHROO, P. and CHOWDHARY, B. P. (2010). Molecular heterogeneity of XY sex reversal in horses : ECAY deletions and SRY-negative XY sex reversal in horses. *Animal Genetics*, **41** (S2), 41–52.
- RÉVAY, T., VILLAGÓMEZ, D.A.F., BREWER, D., CHENIER, T. and KING, W.A. (2012). GTG Mutation in the Start Codon of the Androgen Receptor Gene in a Family of Horses with 64,XY Disorder of Sex Development. *Sexual Development*, **6** (1-3), 108–116.
- RICARD, A., BRUNS, E. and CUNNINGHAM, E. P. (2000). Genetics of performance traits. In : BOWLING, A. T. and RUVINSKY, A., *The genetics of the horse*. New York, U.S.A. : CABI Publishing, 2000, 411–538. ISBN 978-0851994291.
- RICARD, A. and TOUVAIS, M. (2007). Genetic parameters of performance traits in horse endurance races. *Livestock Science*, **110** (1-2), 118–125.
- RIEDER, S. (2009). Molecular tests for coat colours in horses. *Journal of Animal Breeding and Genetics*, **126** (6), 415–424.
- RINN, J. L. and CHANG, H. Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*, **81** (1), 145–166.
- ROBERT, C., BARREY, E. and PATRICE, T. (2014). Effect of A 90-km Endurance Race on Resistance to Oxidation and Inflammation Levels in 6-Year Old Horses. *Equine Veterinary Journal*, **46** (S46), 19.
- ROBERT, C., SCHIBLER, L., MATA, X., BLOUIN, C., RICARD, A. and BARREY, E. (2012). Mesurations et niveau de qualification des chevaux d'endurance de race Arabe et croisé Arabe inclus dans le projet GenEndurance 2011. In : *38e Journée de la Recherche Equine*, 1^{er} mars 2012, Paris, France. Paris : Librairie Institut Français du Cheval et de l'Equitation, 223–226.
- ROBINSON, N. E. (2001). International workshop on equine chronic airway disease, Michigan State University, 16–18 June 2000. *Equine veterinary journal*, **33** (1), 5–19.
- ROSENGREN PIELBERG, G., GOLOVKO, A., SUNDSTROM, E., CURIK, I., LENNARTSSON, J., SELTENHAMMER, M. H, DRUML, T., BINNS, M., FITZSIMMONS, C., LINDGREN, G., SANDBERG, K., BAUMUNG, R., VETTERLEIN, M., STROMBERG, S., GRABHERR, M., WADE, C., LINDBLAD-TOH, K., PONTEN, F., HELDIN, C.-H., SOLKNER, J. and ANDERSSON, L. (2008). A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nature Genetics*, **40** (8), 1004–1009.
- ROSENSON, R. S. (2010). New Technologies Personalize Diagnostics and Therapeutics. *Current Atherosclerosis Reports*, **12** (3), 184–186.
- RUDOLPH, J. A., SPIER, S. J., BYRNS, G. and HOFFMAN, E. P. (1992). Linkage of hyperkalaemic periodic paralysis in Quarter horses to the horse adult skeletal muscle sodium channel gene. *Animal Genetics*, **23** (3), 241–250.
- RUDOLPH, J. A., SPIER, S. J., BYRNS, G., ROJAS, C. V., BERNOCO, D. and HOFFMAN, E. P. (1992). Periodic paralysis in Quarter Horses : a sodium channel mutation disseminated by selective breeding. *Nature Genetics*, **2** (2), 144–147.
- RUSSELL, A. P., LAMON, S., BOON, H., WADA, S., GÜLLER, I., BROWN, E. L., CHIBALIN, A. V., ZIERATH, J. R., SNOW, R. J., STEPTO, N., WADLEY, G. D. and AKIMOTO, T. (2013). Regulation of miRNAs in human skeletal muscle following acute endurance exercise and short-term endurance training : miRNA in

muscle after exercise. *The Journal of Physiology*, **591** (18), 4637–4653.

SAN CRISTOBAL, M., SANCHEZ, M.-P., MERCAT, M.-J., ROHART, F., LIAUBET, L., TRIBOUT, T., CANLET, C., MULLER, N., MOLINA, J., IANUCCELLI, N., LAURENT, B., VILLA-VIALANEIX, N., PARIS, A. and MILAN, D. (2014). Le métabolome, un moyen pour trouver de nouveaux biomarqueurs ? *Viandes et Produits Carnés - Février 2014*, 1–5.

SANDMEYER, L. S., BELLONE, R. R., ARCHER, S., BAUER, B. S., NELSON, J., FORSYTH, G. and GRAHN, B. H. (2012). Congenital stationary night blindness is associated with the leopard complex in the miniature horse : Miniature horse congenital stationary night blindness is associated with the leopard complex. *Veterinary Ophthalmology*, **15** (1), 18–22.

SANTSCHI, E. M., PURDY, A. K., VALBERG, S. J., VROTSOS, P. D., KAESE, H. and MICKELSON, J. R. (1998). Endothelin receptor B polymorphism associated with lethal white foal syndrome in horses. *Mammalian genome*, **9** (4), 306–309.

SANZ, M. G., WILLS, T. B., CHRISTOPHERSON, P. and HINES, M. T. (2011). Glanzmann thrombasthenia in a 17-year-old Peruvian Paso mare : Glanzmann thrombasthenia in a mare. *Veterinary Clinical Pathology*, **40** (1), 48–51.

SATO, Y., NAGASAKI, M., NAKAI, N. and FUSHIMI, T. (2003). Physical Exercise Improves Glucose Metabolism in Lifestyle-Related Diseases. *Experimental Biology and Medicine*, **228** (10), 1208–1212.

SAUVAGEAU, M., GOFF, L. A., LODATO, S., BONEV, B., GROFF, A. F., GERHARDINGER, C., SANCHEZ-GOMEZ, D. B., HACISULEYMAN, E., LI, E., SPENCE, M., LIAPIS, S. C., MALLARD, W., MORSE, M., SWERDEL, M. R., D'ECCLESSIS, M. F., MOORE, J. C., LAI, V., GONG, G., YANCOPOULOS, G. D., FRENDEWEY, D., KELLIS, M., HART, R. P., VALENZUELA, D. M., ARLOTTA, P. and RINN, J. L. (2013). Multiple knockout mouse models reveal lncRNAs are required for life and brain development. *eLife*, **2**, e01968.

SAVAGNER, F., LE PENNEC, S., RIVALIN, R. and EYER, J. (2015). Les microARNs. *Revue Francophone des Laboratoires*, **2015** (473), 49–54.

SCARTH, J. P., SPENCER, H. A., TIMBERS, S. E., HUDSON, S. C. and HILLYER, L. L. (2010). The use of in vitro technologies coupled with high resolution accurate mass LC-MS for studying drug metabolism in equine drug surveillance. *Drug Testing and Analysis*, **2**, 1-10.

SCHAUER, S. N., SONTAKKE, S. D., WATSON, E. D., ESTEVES, C. L. and DONADEU, F. X. (2013). Involvement of miRNAs in equine follicle development. *Reproduction*, **146** (3), 273–282.

SCHENA, M., SHALON, D., DAVIS, R. W. and BROWN, P. O. (1995). Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*, **270** (5235), 467–470.

SCHIBLER, L. (2012). LA GÉNOMIQUE : UN OUTIL PERFORMANT POUR UNE GESTION INTÉGRÉE DE L'ÉLEVAGE, DE L'ENTRAÎNEMENT ET DE LA SANTÉ DES CHEVAUX. *Bulletin de l'Académie vétérinaire de France*, **165** (3), 197-204.

SCHMITZ, J. and BROSIUS, J. (2011). Exonization of transposed elements : A challenge and opportunity for evolution. *Biochimie*, **93** (11), 1928–1934.

SCHRIMPF, R., DIERKS, C., MARTINSSON, G., SIEME, H. and DISTL, O. (2014). Genome-Wide Association Study Identifies Phospholipase C zeta 1 (PLCz1) as a Stallion Fertility Locus in Hanoverian Warmblood Horses. *PLoS ONE*, **9** (10), e109675.

SCHRÖDER, W., KLOSTERMANN, A., STOCK, K. F. and DISTL, O. (2012). A genome-wide association study for quantitative trait loci of show-jumping in Hanoverian warmblood horses : Genome-wide association for show-jumping. *Animal Genetics*, **43** (4), 392–400.

SCHURINK, A., WOLC, A., DUCRO, B. J., FRANKENA, K., GARRICK, D. J., DEKKERS, J. C.M. and VAN ARENDONK, J. A.M. (2012). Genome-wide association study of insect bite hypersensitivity in two horse

populations in the Netherlands. *Genetics Selection Evolution*, **44** (1), 31.

SCOPPETTA, F., TARTAGLIA, M., RENZONE, G., AVELLINI, L., GAITI, A., SCALONI, A. and CHIARADIA, E. (2012). Plasma protein changes in horse after prolonged physical exercise : A proteomic study. *Journal of Proteomics*, **75** (14), 4494–4504.

SEYEDNASROLLAH, F., LAIHO, A. and ELO, L. L. (2015). Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in Bioinformatics*, **16** (1), 59–70.

SHIN, E. K., PERRYMAN, L. E. and MEEK, K. (1997a). A kinase-negative mutation of DNA-PK(CS) in equine SCID results in defective coding and signal joint formation. *Journal of Immunology (Baltimore, Md.: 1950)*, **158** (8), 3565–3569.

SHIN, E. K., PERRYMAN, L. E. and MEEK, K. (1997b). Evaluation of a test for identification of Arabian horses heterozygous for the severe combined immunodeficiency trait. *Journal of the American Veterinary Medical Association*, **211** (10), 1268–1270.

SHIROMA, E. J. and LEE, I.-M. (2010). Physical Activity and Cardiovascular Health : Lessons Learned From Epidemiological Studies Across Age, Gender, and Race/Ethnicity. *Circulation*, **122** (7), 743–752.

SIGAUX, F. (2005). Comment entrer dans l'ère de la biologie intégrative (Systems Biology). *Hématologie*, **11** (2), 95.

SILVA, A. L. and ROMÃO, L. (2009). The mammalian nonsense-mediated mRNA decay pathway : To decay or not to decay ! Which players make the decision ? *FEBS Letters*, **583** (3), 499–505.

SIMPSON, L. J., PATEL, S., BHAKTA, N. R., CHOY, D. F., BRIGHTBILL, H. D., REN, X., WANG, Y., PUA, H. H., BAUMJOHANN, D., MONTOYA, M. M., PANDURO, M., REMEDIOS, K. A., HUANG, X., FAHY, J. V., ARRON, J. R., WOODRUFF, P. G. and ANSEL, K. M. (2014). A microRNA upregulated in asthma airway T cells promotes TH2 cytokine production. *Nature Immunology*, **15** (12), 1162–1170.

SPECK, R. M., COURNEYA, K. S., MÂSSE, L. C., DUVAL, S. and SCHMITZ, K. H. (2010). An update of controlled physical activity trials in cancer survivors : a systematic review and meta-analysis. *Journal of Cancer Survivorship*, **4** (2), 87–100.

SPIRITO, F., CHARLESWORTH, A., ORTONNE, J.-P., MENEGUZZI, G., LINDER, K. and BAIRD, J. (2002). Animal Models for Skin Blistering Conditions : Absence of Laminin 5 Causes Hereditary Junctional Mechanobullous Disease in the Belgian Horse. *Journal of Investigative Dermatology*, **119** (3), 684–691.

STANTON, R. and REABURN, P. (2014). Exercise and the treatment of depression : A review of the exercise program variables. *Journal of Science and Medicine in Sport*, **17** (2), 177–182.

STOCK, K. F. and DISTL, O. (2007). Genetic correlations between performance traits and radiographic findings in the limbs of German Warmblood riding horses. *Journal of Animal Science*, **85** (1), 31–41.

STOCK, K. F. and DISTL, O. (2008). Multiple-trait selection for radiographic health of the limbs, conformation and performance in Warmblood riding horses. *Animal*, **2** (12), 1724.

STRICKLAND, J. (2014). *Predictive Modeling and Analytics*. First Edition. Colorado Springs, Colorado, U.S.A. : Lulu.com, 488 p. ISBN 978-1-312-37544-4.

SUDMANT, P. H., RAUSCH, T., GARDNER, E. J., HANDSAKER, R. E., ABYZOV, A., HUDDLESTON, J., ZHANG, Y., YE, K., JUN, G., HSI-YANG FRITZ, M., KONKEL, M. K., MALHOTRA, A., STÜTZ, A. M., SHI, X., PAOLO CASALE, F., CHEN, J., HORMOZDIARI, F., DAYAMA, G., CHEN, K., MALIG, M., CHAISSON, M. J. P., WALTER, K., MEIERS, S., KASHIN, S., GARRISON, E., AUTON, A., LAM, H. Y. K., JASMINE MU, X., ALKAN, C., ANTAKI, D., BAE, T., CERVEIRA, E., CHINES, P., CHONG, Z., CLARKE, L., DAL, E., DING, L., EMERY, S., FAN, X., GUJRAL, M., KAHVECI, F., KIDD, J. M., KONG, Y., LAMEIJER, E.-W., MCCARTHY, S., FLICEK, P., GIBBS, R. A., MARTH, G., MASON, C. E., MENELAOU, A., MUZNY, D. M., NELSON, B. J., NOOR, A., PARRISH, N. F., PENDLETON, M., QUITADAMO, A., RAEDER, B., SCHADT, E. E., ROMANOVITCH, M., SCHLATT, A., SEBRA, R., SHABALIN, A. A., UNTERGASSER, A.,

- WALKER, J. A., WANG, M., YU, F., ZHANG, C., ZHANG, J., ZHENG-BRADLEY, X., ZHOU, W., ZICHNER, T., SEBAT, J., BATZER, M. A., MCCARROLL, S. A., MILLS, R. E., GERSTEIN, M. B., BASHIR, A., STEGLE, O., DEVINE, S. E., LEE, C., EICHLER, E. E. and KORBEL, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, **526** (7571), 75–81.
- SUDRE, K., CASSAR-MALEK, I., LEROUX, C., LISTRAT, A., UEDA, Y., JURIE, C., RENAND, G., MARTIN, P. and HOCQUETTE, J. F. (2003). Transcriptome analysis of muscle in order to identify genes which determine muscle characteristics and sensory quality traits of beef. *Sciences des aliments*, **23** (1), 65–69.
- SUZUKI, K., PEAKE, J., NOSAKA, K., OKUTSU, M., ABBISS, C. R., SURRIANO, R., BISHOP, D., QUOD, M. J., LEE, H., MARTIN, D. T. and LAURSEN, P. B. (2006). Changes in markers of muscle damage, inflammation and HSP70 after an Ironman triathlon race. *European Journal of Applied Physiology*, **98** (6), 525–534.
- TEALE, P., BARTON, C., DRIVER, P. M. and KAY, R. G. (2009). Biomarkers : unrealized potential in sports doping analysis. *Bioanalysis*, **1** (6), 1103–1118.
- TETENS, J., WIDMANN, P., KÜHN, C. and THALLER, G. (2013). A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. *Animal Genetics*, **44** (4), 467–471.
- THE 1000 GENOMES PROJECT CONSORTIUM (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491** (7422), 56–65.
- THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM, ELSIK, C. G., TELLAM, R. L. and WORLEY, K. C. (2009). The Genome Sequence of Taurine Cattle : A Window to Ruminant Biology and Evolution. *Science*, **324** (5926), 522–528.
- THE CHIMPANZEE SEQUENCING AND ANALYSIS CONSORTIUM (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, **437** (7055), 69–87.
- THE INTERNATIONAL CHICKEN GENOME SEQUENCING CONSORTIUM (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432** (7018), 695–716.
- THE INTERNATIONAL SHEEP GENOMICS CONSORTIUM, ARCHIBALD, A. L., COCKETT, N. E., DALRYMPLE, B. P., FARAUT, T., KIJAS, J. W., MADDOX, J. F., MCEWAN, J. C., HUTTON ODDY, V., RAADSMA, H. W., WADE, C., WANG, J., WANG, W. and XUN, X. (2010). The sheep genome reference sequence : a work in progress. *Animal Genetics*, **41** (5), 449–453.
- THE MOUSE PHENOTYPE DATABASE INTEGRATION CONSORTIUM (2007). Integration of mouse phenome data resources. *Mammalian Genome*, **18** (3), 157–163.
- THOMAES, T., THOMIS, M., ONKELINX, S., FAGARD, R., MATTHIJS, G., BUYS, R., SCHEPERS, D., CORNELISSEN, V. and VANHEES, L. (2011). A genetic predisposition score for muscular endophenotypes predicts the increase in aerobic power after training : the CAREGENE study. *BMC genetics*, **12** (1), 1.
- TIDBALL, J. G. (2004). Inflammatory processes in muscle injury and repair. *American Journal of Physiology : Regulatory, Integrative and Comparative Physiology*, **288** (2), R345–R353.
- TOUMI, H., F'GUYER, S. and BEST, T. M. (2006). The role of neutrophils in injury and repair following muscle stretch. *Journal of anatomy*, **208** (4), 459–470.
- TOWERS, R. E., MURGIANO, L., MILLAR, D. S., GLEN, E., TOPF, A., JAGANNATHAN, V., DRÖGEMÜLLER, C., GOODSHIP, J. A., CLARKE, A. J. and LEEB, T. (2013). A Nonsense Mutation in the IKBKG Gene in Mares with Incontinentia Pigmenti. *PLoS ONE*, **8** (12), e81625.
- TOZAKI, T., MIYAKE, T., KAKOI, H., GAWAHARA, H., SUGITA, S., HASEGAWA, T., ISHIDA, N., HIROTA, K. and NAKANO, Y. (2010). A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene : A genome-wide scan for racing performances.

Animal Genetics, **41** (S2), 28–35.

TRACHSEL, D. S., GIRAUDET, A., HERVÉ, G., MASO, D. and ROBERT, C. (2014). Evolution of Echocardiographic Dimensions with Age and Training in Young Endurance Horses. *Equine Veterinary Journal*, **46**, 26.

TRANG, P., WEIDHAAS, J. B. and SLACK, F. J. (2008). MicroRNAs as potential cancer therapeutics. *Oncogene*, **27**, S52–S57.

TRIBOUT, P. (2013). *Etude morphométrique du cheval d'endurance de race arabe et croisé arabe en relation avec la performance*. Thèse de doctorat vétérinaire, Faculté de Médecine de Créteil, 94 p.

TRYON, R. C., WHITE, S. D. and BANNASCH, D. L. (2007). Homozygosity mapping approach identifies a missense mutation in equine cyclophilin B (PPIB) associated with HERDA in the American Quarter Horse. *Genomics*, **90** (1), 93–102.

VALBERG, S. J., WARD, T. L., RUSH, B., KINDE, H., HIRARAGI, H., NAHEY, D., FYFE, J. and MICKELSON, J. R. (2001). Glycogen branching enzyme deficiency in quarter horse foals. *Journal of Veterinary Internal Medicine*, **15** (6), 572–580.

VANCE, D. D., CHEN, G. L., STOUTENBERG, M., MYERBURG, R. J., JACOBS, K., NATHANSON, L., PERRY, A., SEO, D., GOLDSCHMIDT-CLERMONT, P. J. and RAMPERSAUD, E. (2014). Cardiac performance, biomarkers and gene expression studies in previously sedentary men participating in half-marathon training. *BMC sports science, medicine and rehabilitation*, **6** (1), 6.

VAN DER KOLK, J.H., PACHOLEWSKA, A. and GERBER, V. (2015). The role of microRNAs in equine medicine : a review. *Veterinary Quarterly*, **35** (2), 88–96.

VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. and KINZLER, K. W. (1995). Serial Analysis of Gene Expression. *Science*, **270** (5235), 484–487.

VENTER, J. C., ADAMS, M. D., MYERS, E. W., LI, P. W., MURAL, R. J., SUTTON, G. G., SMITH, H. O., YANDELL, M., EVANS, C. A., HOLT, R. A., GOCAYNE, J. D., AMANATIDES, P., BALLEW, R. M., HUSON, D. H., WORTMAN, J. R., ZHANG, Q., KODIRA, C. D., ZHENG, X. H., CHEN, L., SKUPSKI, M., SUBRAMANIAN, G., THOMAS, P. D., ZHANG, J., MIKLOS, G. L. G., NELSON, C., BRODER, S., CLARK, A. G., NADEAU, V. A., MCKUSICK, V. A., ZINDER, N., LEVINE, A. J., ROBERTS, R. J., SIMON, M., SLAYMAN, D., HUNKAPILLER, M., BOLANOS, R., DELCHER, A., DEW, I., FASULO, D., FLANIGAN, M., FLOREA, L., HALPERN, A., HANNENHALLI, S., KRAVITZ, S., LEVY, S., MOBARRY, C., REINERT, K., REMINGTON, K., ABU-THREIDEH, J., BEASLEY, E., BIDDICK, K., BONAZZI, V., BRANDON, R., CARGILL, M., CHANDRAMOULISWARAN, I., CHARLAB, R., CHATURVEDI, K., DENG, Z., FRANCESCO, V. D., DUNN, P., EILBECK, K., EVANGELISTA, C., GABRIELIAN, A. E., GAN, W., GE, W., GONG, F., GU, Z., GUAN, P., HEIMAN, T. J., HIGGINS, M. E., JI, R.-R., KE, Z., KETCHUM, K. A., LAI, Z., LEI, Y., LI, Z., LI, J., LIANG, Y., LIN, X., LU, F., MERKULOV, G. V., MILSHINA, N., MOORE, H. M., NAIK, A. K., NARAYAN, V. A., NEELAM, B., NUSSKERN, D., RUSCH, D. B., SALZBERG, S., SHAO, W., SHUE, B., SUN, J., WANG, Z. Y., WANG, A., WANG, X., WANG, J., WEI, M.-H., WIDES, R., XIAO, C., YAN, C., YAO, A., YE, J., ZHAN, M., ZHANG, W., ZHANG, H., ZHAO, Q., ZHENG, L., ZHONG, F., ZHONG, W., ZHU, S. C., ZHAO, S., GILBERT, D., BAUMHUETER, S., SPIER, G., CARTER, C., CRAVCHIK, A., WOODAGE, T., ALI, F., AN, H., AWE, A., BALDWIN, D., BADEN, H., BARNSTEAD, M., BARROW, I., BEESON, K., BUSAM, D., CARVER, A., CENTER, A., CHENG, M. L., CURRY, L., DANAHER, S., DAVENPORT, L., DESILETS, R., DIETZ, S., DODSON, K., DOUP, L., FERRIERA, S., GARG, N., GLUECKSMANN, A., HART, B., HAYNES, J., HAYNES, C., HEINER, C., HLDUN, S., HOSTIN, D., HOUCK, J., HOWLAND, T., IBEGWAM, C., JOHNSON, J., KALUSH, F., KLINE, L., KODURU, S., LOVE, A., MANN, F., MAY, D., MCCAWLEY, S., MCINTOSH, T., MCMULLEN, I., MOY, M., MOY, L., MURPHY, B., NELSON, K., PFANNKOCH, C., PRATTS, E., PURI, V., QURESHI, H., REARDON, M., RODRIGUEZ, R., ROGERS, Y.-H., ROMBLAD, D., RUHFEL, B., SCOTT, R., SITTER, C., SMALLWOOD, M., STEWART, E., STRONG, R., SUH, E., THOMAS, R., TINT, N. N., TSE, S., VECH, C., WANG, G., WETTER, J., WILLIAMS, S., WILLIAMS, M., WINDSOR, S., WINN-DEEN, E., WOLFE, K., ZAVERI, J., ZAVERI, K., ABRIL, J. F., GUIGÓ, R., CAMPBELL, M. J., SJOLANDER, K. V., KARLAK, B., KEJARIWAL, A., MI, H., LAZAREVA, B., HATTON, T., NARECHANIA, A., DIEMER, K., MURUGANUJAN, A., GUO, N., SATO, S., BAFNA, V., ISTRAIL, S., LIPPERT, R., SCHWARTZ, R., WALENZ, B., YOOSEPH, S., ALLEN, D., BASU,

A., BAXENDALE, J., BLICK, L., CAMINHA, M., CARNES-STINE, J., CAULK, P., CHIANG, Y.-H., COYNE, M., DAHLKE, C., MAYS, A. D., DOMBROSKI, M., DONNELLY, M., ELY, D., ESPARHAM, S., FOSLER, C., GIRE, H., GLANOWSKI, S., GLASSER, K., GLODEK, A., GOROKHOV, M., GRAHAM, K., GROPMAN, B., HARRIS, M., HEIL, J., HENDERSON, S., HOOVER, J., JENNINGS, D., JORDAN, C., JORDAN, J., KASHA, J., KAGAN, L., KRAFT, C., LEVITSKY, A., LEWIS, M., LIU, X., LOPEZ, J., MA, D., MAJOROS, W., MCDANIEL, J., MURPHY, S., NEWMAN, M., NGUYEN, T., NGUYEN, N., NODELL, M., PAN, S., PECK, J., PETERSON, M., ROWE, W., SANDERS, R., SCOTT, J., SIMPSON, M., SMITH, T., SPRAGUE, A., STOCKWELL, T., TURNER, R., VENTER, E., WANG, M., WEN, M., WU, D., WU, M., XIA, A., ZANDIEH, A. and ZHU, X. (2001). The Sequence of the Human Genome. *Science*, **291** (5507), 1304–1351.

VIGNAL, A. (2011). Etat actuel du séquençage et de la connaissance du génome des espèces animales. *INRA Productions Animales*, **24** (4), 387-404.

VIKLUND, A., BRAAM, A., NÄSHOLM, A., STRANDBERG, E. and PHILIPSSON, J. (2010). Genetic variation in competition traits at different ages and time periods and correlations with traits at field tests of 4-year-old Swedish Warmblood horses. *Animal*, **4** (05), 682-691.

VIKLUND, A., NÄSHOLM, A., STRANDBERG, E. and PHILIPSSON, J. (2010). Effects of long-time series of data on genetic evaluations for performance of Swedish Warmblood riding horses. *Animal*, **4** (11), 1823–1831.

VITZTHUM, F., BEHRENS, F., ANDERSON, N. L. and SHAW, J. H. (2005). Proteomics : From Basic Research to Diagnostic Application. A Review of Requirements & Needs. *Journal of Proteome Research*, **4** (4), 1086–1097.

VUOLTEENAHO, K., LEPPÄNEN, T., KEKKONEN, R., KORPELA, R. and MOILANEN, E. (2014). Running a Marathon Induces Changes in Adipokine Levels and in Markers of Cartilage Degradation – Novel Role for Resistin. *PLoS ONE*, **9** (10), e110481.

WADE, C. M., GIULOTTO, E., SIGURDSSON, S., ZOLI, M., GNERRE, S., IMSLAND, F., LEAR, T. L., ADELSON, D. L., BAILEY, E., BELLONE, R. R., BLÖCKER, H., DISTL, O., EDGAR, R. C., GARBER, M., LEEB, T., MAUCELLI, E., MACLEOD, J. N., PENEDO, M. C. T., RAISON, J. M., SHARPE, T., VOGEL, J., ANDERSSON, L., ANTCZAK, D. F., BIAGI, T., BINNS, M. M., CHOWDHARY, B. P., COLEMAN, S.J., DELLA VALLE, G., FRYC, S., GUÉRIN, G., HASEGAWA, T., HILL, E. W., JURKA, J., KIIALAINEN, A., LINDGREN, G., LIU, J., MAGNANI, E., MICKELSON, J. R., MURRAY, J., NERGADZE, S. G., ONOFRIO, R., PEDRONI, S., PIRAS, M. F., RAUDSEPP, T., ROCCHI, M., RØED, K. H., RYDER, O. A., SEARLE, S., SKOW, L., SWINBURNE, J. E., SYVÄNEN, A. C., TOZAKI, T., VALBERG, S. J., VAUDIN, M., WHITE, J. R., ZODY, M. C., LANDER, E. S. and LINDBLAD-TOH, K. (2009). Genome sequence, comparative analysis and population genetics of the domestic horse (*Equus caballus*). *Science*, **326** (5954), 865–867.

WALLBERG, L., MIKAEL MATTSSON, C., ENQVIST, J. K. and EKBLÖM, B. (2011). Plasma IL-6 concentration during ultra-endurance exercise. *European Journal of Applied Physiology*, **111** (6), 1081–1088.

WALSH, N. P., GLEESON, M., SHEPHARD, R. J., GLEESON, M., WOODS, J. A., BISHOP, N. C., FLESHNER, M., GREEN, C., PEDERSEN, B. K., HOFFMAN-GOETZ, L., ROGERS, C. J., NORTHOFF, H., ABBASI, A. and SIMON, P. (2011). Position statement part one : immune function and exercise. *Exercise Immunology Review*, **17**, 6-63.

WANG, X., MILLER, D. C., CLARK, A. G. and ANTCZAK, D. F. (2012). Random X inactivation in the mule and horse placenta. *Genome Research*, **22** (10), 1855–1863.

WANG, X., MILLER, D. C., HARMAN, R., ANTCZAK, D. F. and CLARK, A. G. (2013). Paternally expressed genes predominate in the placenta. *Proceedings of the National Academy of Sciences*, **110** (26), 10705–10710.

WANSCHER, J. H. (1975). The History of Wilhelm Johannsen's Genetical Terms and Concepts from the Period 1903 to 1926. *Centaurus*, **19** (2), 125–147.

WARD, R. J., FRANCAUX, M., CUISINIER, C., STURBOIS, X. and DE WITTE, P. (1999). Changes in plasma taurine levels after different endurance events. *Amino acids*, **16** (1), 71–77.

WARD, T. L., VALBERG, S. J., ADELSON, D. L., ABBEY, C. A., BINNS, M. M. and MICKELSON, J. R. (2004). Glycogen branching enzyme (GBE1) mutation causing equine glycogen storage disease IV. *Mammalian*

Genome, **15** (7), 570-577.

WATSON, R., GEE, A., POLKINGHORNE, R. and PORTER, M. (2008). Consumer assessment of eating quality – development of protocols for Meat Standards Australia (MSA) testing. *Australian Journal of Experimental Agriculture*, **48** (11), 1360.

WEBBON, P. (2012). Harnessing the genetic toolbox for the benefit of the racing Thoroughbred : Thoroughbred genetics. *Equine Veterinary Journal*, **44** (1), 8–12.

WEILAND, M., GAO, X.-H., ZHOU, L. and MI, Q.-S. (2012). Small RNAs have a large impact : Circulating microRNAs as biomarkers for human diseases. *RNA Biology*, **9** (6), 850–859.

WEILER, J., HUNZIKER, J. and HALL, J. (2006). Anti-miRNA oligonucleotides (AMOs) : ammunition to target miRNAs implicated in human disease? *Gene therapy*, **13** (6), 496–502.

WEST-WRIGHT, C. N., HENDERSON, K. D., SULLIVAN-HALLEY, J., URSIN, G., DEAPEN, D., NEUHAUSEN, S., REYNOLDS, P., CHANG, E., MA, H. and BERNSTEIN, L. (2009). Long-Term and Recent Recreational Physical Activity and Survival After Breast Cancer : The California Teachers Study. *Cancer Epidemiology Biomarkers & Prevention*, **18** (11), 2851–2859.

WIJNBERG, I. D., OWCZAREK-LIPSKA, M., SACCHETTO, R., MASCARELLO, F., PASCOLI, F., GRÜNBERG, W., VAN DER KOLK, J. H. and DRÖGEMÜLLER, C. (2012). A missense mutation in the skeletal muscle chloride channel 1 (CLCN1) as candidate causal mutation for congenital myotonia in a New Forest pony. *Neuromuscular Disorders*, **22** (4), 361–367.

WILKINS, M. (2009). Proteomics data mining. *Expert Review of Proteomics*, **6** (6), 599–603.

WILUSZ, J. E., SUNWOO, H. and SPECTOR, D. L. (2009). Long noncoding RNAs : functional surprises from the RNA world. *Genes & Development*, **23** (13), 1494–1504.

WINAND, N. J. (2012). *Identification of the causative mutation for inherited connective tissue disorders in equines and methods for testing for same*. Cornell University, N.Y., U.S.A., Patent WO2008121727 (A1), Google Patents, 9 October 2008.

WORLD CANCER RESEARCH FUND / AMERICAN INSTITUTE FOR CANCER RESEARCH (2007). *Food, Nutrition, Physical Activity, and the Prevention of Cancer : a Global Perspective*. First Edition. Washington DC, U.S.A. : American Institute for Cancer Research, 517 p. ISBN 978-0-9722522-2-5.

XIANG, L., REHM, K. E. and MARSHALL, G. D. (2014). Effects of strenuous exercise on Th1/Th2 gene expression from human peripheral blood mononuclear cells of marathon participants. *Molecular Immunology*, **60** (2), 129–134.

XU, X., LAN, J. and KORFMACHER, W. A. (2005). Rapid LC/MS/MS Method Development for Drug Discovery. *Analytical Chemistry*, **77** (19), 389A–394A.

YANG, G. C., CROAKER, D., ZHANG, A. L., MANGLUICK, P., CARTMILL, T. and CASS, D. (1998). A dinucleotide mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome (LWFS) ; a horse variant of Hirschsprung disease (HSCR). *Human Molecular Genetics*, **7** (6), 1047–1052.

YEOMAN, C. J., THOMAS, S. M., MILLER, M. E. B., ULANOV, A. V., TORRALBA, M., LUCAS, S., GILLIS, M., CREGGER, M., GOMEZ, A., HO, M., LEIGH, S. R., STUMPF, R., CREEDON, D. J., SMITH, M. A., WEISBAUM, J. S., NELSON, K. E., WILSON, B. A. and WHITE, B. A. (2013). A Multi-Omic Systems-Based Approach Reveals Metabolic Markers of Bacterial Vaginosis and Insight into the Disease. *PLoS ONE*, **8** (2), e56111.

YEUNG, R. R. (1996). The acute effects of exercise on mood state. *Journal of Psychosomatic Research*, **40** (2), 123–141.

YOUNES, M., ROBERT, C., COTTIN, F. and BARREY, E. (2015). Speed and Cardiac Recovery Variables Predict the Probability of Elimination in Equine Endurance Events. *PLoS ONE*, **10** (8), e0137013.

ZABEL, C. and KLOSE, J. (2009). High-resolution large-gel 2DE. *Methods in Molecular Biology (Clifton, N.J.)*, **519**, 311–338.

ZHANG, C., HONG, H., MENDRICK, D. L, TANG, Y. and CHENG, F. (2015). Biomarker-based drug safety assessment in the age of systems pharmacology : from foundational to regulatory science. *Biomarkers in Medicine*, **9** (11), 1241–1252.

ZHANG, C. (2008). MicroRNomics : a newly emerging approach for disease biology. *Physiological Genomics*, **33** (2), 139–147.

ZHANG, X., FANG, A., RILEY, C. P., WANG, M., REGNIER, F. E. and BUCK, C. (2010). Multi-dimensional liquid chromatography in proteomics—A review. *Analytica Chimica Acta*, **664** (2), 101–113.

ZHOU, M., WANG, Q., SUN, J., LI, X., XU, L., YANG, H., SHI, H., NING, S., CHEN, L., LI, Y., HE, T. and ZHENG, Y. (2009). In silico detection and characteristics of novel microRNA genes in the Equus caballus genome using an integrated ab initio and comparative genomic approach. *Genomics*, **94** (2), 125–131.

GLOSSAIRE :

Acide aminé : Unité élémentaire des protéines. C'est une molécule organique possédant un squelette carboné et deux fonctions : une fonction amine (-NH₂) et une fonction acide carboxylique (-COOH).

ADN : Acide désoxyribonucléique, molécule contenant l'information héréditaire et constituant essentiel des chromosomes. L'ADN est formé par l'enchaînement de quatre unités moléculaires (bases : adénine, cytosine, guanine, thymidine), dont l'ordre (la séquence) est spécifique d'un gène (par exemple) et détermine sa fonction. Le génome haploïde d'un animal est composé approximativement de 3 milliards de paires de bases.

Allèles : Différentes versions d'un même gène dérivant les unes des autres par mutation et dont l'expression peut éventuellement conduire à des phénotypes différents. Chaque allèle se distingue par une ou plusieurs différences de la séquence de bases. Tous les allèles d'un gène occupent le même locus (emplacement) sur un même chromosome.

Apoptose : Mécanisme de mort cellulaire.

ARN : Acide ribonucléique, produit de la transcription de l'ADN.

ARNm : ARN messager, ARN codant vecteur d'information entre le noyau et le cytoplasme.

ARNnc : ARN non codant, non traduit en protéine.

Biopuce : Aussi appelée puce à acide nucléique (puce à ADN, à ARN...), il s'agit d'une lame (de verre le plus souvent, de silicium ou de plastique) sur laquelle sont fixés de manière ordonnée des fragments d'ADN ou d'ARN appelés « sondes ».

Biomarqueur : Ce terme désigne tout paramètre biologique (protéine, ADN, ARN...) utilisé comme indicateur d'un processus physiologique, pathologique ou de réponse pharmacologique à un traitement, mesuré objectivement (dosé dans les liquides biologiques ou tissus par exemple), et pouvant remplacer un critère d'évaluation clinique sur la base de preuves épidémiologiques, thérapeutiques, physiopathologiques, etc.

Carte génétique / génomique : Connaissance de l'ordre de l'ensemble des gènes ou marqueurs de l'ADN distribués le long d'un chromosome ; la densité des cartes dépend du nombre de marqueurs localisés (quelques centaines à quelques dizaines de milliers).

Chromosome : Structures présentes dans les noyaux des cellules, associant ADN et protéines, constituant le support physique de l'hérédité. Parmi les chromosomes, on distingue les autosomes des gonosomes (chromosomes sexuels). Les cellules équine possèdent 32 paires de chromosomes, les cellules humaines 23.

Copy Number Variants (CNV) : Variations du nombre de copies de segments d'ADN pouvant ou non contenir des gènes.

Codant (ARN ou gène) : ARN messager, ou partie d'un gène qui définit directement la séquence en acides aminés de la protéine correspondante.

Codon stop : Ensemble de 3 bases successives sur un brin d'ADN (codon) dont la traduction est à l'origine d'un arrêt de l'élongation de la chaîne protéique (terminaison de la traduction).

Délétion : Mécanisme de mutation caractérisée par la perte d'une partie du matériel génétique pouvant aller d'un seul nucléotide à plusieurs gènes.

Déséquilibre de liaison (*linkage disequilibrium, LD*) : Association non aléatoire d'allèles à des loci différents.

Dominant :

- Allèle dominant : allèle dont le phénotype associé s'exprime à l'état homozygote et hétérozygote.
- Transmission dominante d'un caractère / caractère ou affection à déterminisme dominant : la présence d'une seule copie mutée suffit à entraîner le développement de la maladie ou l'expression du caractère.

Épigénétique : Domaine étudiant comment l'environnement et l'histoire individuelle influent sur l'expression des gènes, et plus précisément l'ensemble des modifications transmissibles (via la mitose, et, éventuellement, la méiose) et réversibles de l'expression génique sans altération des séquences nucléotidiques.

Épigénome : Ensemble des marques épigénétiques, c'est-à-dire ensemble des modifications réversibles et hérissables de l'ADN (sans altération de séquence) d'une cellule, ou d'un ensemble de cellules, qui régulent l'expression génomique.

Épigénomique : Science étudiant les épigénomes.

Exome : Ensemble des exons.

Exon : Portion exprimée d'un gène, c'est-à-dire transcrite en ARNm, représentée dans l'ARNm mature puis traduite au sein d'une protéine.

Fluxome : Ensemble des flux métaboliques (de nutriments par exemple) au niveau cellulaire, tissulaire, d'un organe spécifique ou du corps entier.

Fluxomique : Science étudiant les fluxomes.

Gène : Déterminant héréditaire pour un caractère, ou partie de l'ADN codant une protéine.

Génétique : Science qui cherche à comprendre l'origine et la transmission des caractères héréditaires.

Génome : Ensemble des gènes et des séquences intergéniques.

Génomique : Science étudiant la structure, le fonctionnement, la dynamique et l'évolution des génomes.

Génomique fonctionnelle : Science d'étude de la fonction et de l'expression du génome par caractérisation de l'ensemble des transcrits, protéines et métabolites, ainsi que des éléments interagissant avec ces derniers et des éléments régulant leur expression.

Génomique structurale : Science d'étude de l'architecture (structure et localisation des gènes et séquences intergéniques, variations génomiques...) et de l'évolution des génomes.

Génotype : Ensemble des combinaisons alléliques d'un individu.

Haplotype : Séquences adjacentes d'ADN sur un chromosome.

Héritabilité : Part de variation observée au niveau d'un caractère phénotypique due à des différences de valeur génétique additive (partie transmissible de la valeur d'un génome) entre les individus d'une population. Autrement dit, l'héritabilité correspond au rapport entre la variance des valeurs génétiques additives et la variance phénotypique. Elle est un indicateur de la facilité de sélection d'un caractère.

Hétérozygote : Individu possédant deux copies différentes d'un marqueur / gène en un locus donné.

Homozygote : Individu possédant deux copies identiques d'un marqueur / gène en un locus donné.

InDels : Insertion / Délétion.

Insertion : Mécanisme de mutation correspondant à l'addition d'une séquence d'ADN (d'un à plusieurs nucléotides) dans une molécule d'ADN donnée.

Interactome : Ensemble des interactions entre macromolécules d'une cellule (réseaux biologiques)

Intron : Partie d'un gène, située entre deux exons, et éliminée au cours de la maturation de l'ARNm.

Marqueur moléculaire : Fragment d'ADN repère (facilement génotypable) pour suivre la transmission d'un segment de chromosome d'une génération à l'autre (microsatellites, *SNP*, etc.). Un marqueur polymorphe présente une variabilité de séquence.

Métabolome : Ensemble des métabolites (acides aminés, acides gras...) présents dans un fluide biologique.

Métabolomique : Science étudiant les métabolomes.

miARN, microARN : courts ARN de 20 à 24 paires de bases, régulateurs post-transcriptionnels de l'expression génique (répresseurs) par dégradation des ARNm cibles ou inhibition de la traduction de ces derniers. Ils assurent ce rôle répressif associés aux protéines de la famille Argonaute (AGO). Leur expression est tissu-spécifique.

Microtranscriptome, miRNome : Ensemble des microARN.

Microtranscriptomique, miRNomique : Sous-discipline de la génomique fonctionnelle et de la transcriptomique, identifiant les miARN, analysant leur expression, leur biogenèse, leur structure, leurs cibles et leurs fonctions biologiques à une échelle génomique, ainsi que leur rôle dans la régulation de l'expression génomique.

(Affection) monogénique : Affection liée à la mutation d'un seul gène.

Mutation : Variation dans la séquence d'ADN, pouvant provoquer une modification apparente d'un caractère. On parle aussi de variant génétique ou de polymorphisme. Certaines mutations peuvent avoir une localisation ponctuelle unique (c'est-à-dire ne toucher qu'une base), on parle de mutation ponctuelle. Le mécanisme peut être une substitution (à l'origine de *SNP*), une insertion, une délétion, une duplication...

Mutation non-sens : Mutation ponctuelle induisant le remplacement d'un codon correspondant à un acide aminé (codon codant) par un codon stop, à l'origine d'un arrêt prématuré de la traduction. La protéine résultante, tronquée et donc incomplète, est généralement non-fonctionnelle.

myomiRs : MicroARN spécifiques des muscles squelettiques et cardiaque, dont les niveaux d'expression sont modifiés en réponse à l'exercice.

Next Generation Sequencing (NGS) : Séquençage de nouvelle génération ou séquençage massif parallèle, par des séquenceurs de haut débit.

Nutrigénomique : Science d'étude des interactions entre génome et nutriments.

Phénoème : Ensemble des caractères phénotypiques.

Phénomique : Science d'étude du phénoème.

Phénotype : Au sens large, il s'agit de l'ensemble des caractères observables d'un individu ; c'est la manifestation observée de l'expression de l'information génomique propre à un individu dans le milieu dans lequel ce dernier évolue. Autrement dit, c'est l'ensemble apparent des performances résultant de l'expression du génome et des effets de milieu sur différents caractères.

Protéome : Ensemble des protéines exprimées dans un compartiment cellulaire, une cellule, un tissu ou un organisme entier.

Protéomique : Science d'étude du protéome, c'est-à-dire étude à grande échelle des protéines, en particulier de leurs expression, structures, fonctions et interactions, dans un système biologique donné.

Quantitative Trait Locus (QTL) : Région du génome contribuant, par son polymorphisme, à une partie de la variation de caractère(s) quantitatif(s).

Récessif :

- Allèle récessif : allèle dont le phénotype associé s'exprime uniquement à l'état homozygote.

- Transmission récessive d'un caractère / caractère ou affection à déterminisme récessif : le développement de la maladie ou l'expression du caractère phénotypique par l'individu nécessite la présence de deux allèles identiques.

(Taux de) Recombinaison : Phénomène conduisant à l'apparition dans une cellule ou dans un individu de fragments d'ADN (gènes ou séquences intergéniques) ou de caractères héréditaires dans une association différente de celle observée chez les cellules ou individus parentaux.

Séquençage : Technique d'identification de la succession linéaire des bases azotés (A, C, T, G) dans un acide nucléique, ou plus généralement, détermination de l'ordre linéaire des composants d'une macromolécule.

Single Nucleotide Polymorphism (SNP) : mutation ponctuelle d'une base nucléotidique dans la séquence d'ADN.

Traduction : Synthèse d'une chaîne d'acides aminés à partir d'une matrice d'ARNm.

Transcription : Synthèse d'ARN à partir d'une matrice d'ADN.

Transcriptome : Ensemble des transcrits (ARN codés par le génome), reflet direct du niveau d'expression des gènes.

Transcriptomique : Science d'étude des transcrits ; c'est une approche permettant de quantifier et décrire le transcriptome.

Transposon : Aussi appelé élément transposable / instable / mobile, il s'agit d'un fragment d'ADN susceptible de se déplacer, de s'exciser, de s'insérer d'un endroit du génome à un autre.

AGREMENT SCIENTIFIQUE

En vue de l'obtention du permis d'imprimer de la thèse de doctorat vétérinaire

Je soussigné, Alain DUCOS, Enseignant-chercheur, de l'Ecole Nationale Vétérinaire de Toulouse, directeur de thèse, certifie avoir examiné la thèse de **NEVOT Alizée** intitulée « **Prédiction des statuts métabolique et de santé de chevaux d'endurance via l'intégration de données « omiques » : vers l'identification de biomarqueurs** » et que cette dernière peut être imprimée en vue de sa soutenance.

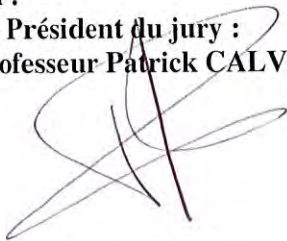
Fait à Toulouse, le 1 septembre 2016
Professeur Alain DUCOS
Enseignant chercheur
de l'Ecole Nationale Vétérinaire de Toulouse



Vu :
La Directrice de l'Ecole Nationale
Vétérinaire de Toulouse
Isabelle CHMITTEIN



Vu :
Le Président du jury :
Professeur Patrick CALVAS



Vu et autorisation de l'impression :
Président de l'Université
Paul Sabatier
Monsieur Jean-Pierre VINEL

Le Président de l'Université Paul Sabatier
par délégation,
La Vice-Présidente de la CFVU


Régine ANDRE-OBRECHT

Mlle NEVOT Alizée
a été admis(e) sur concours en : 2011
a obtenu son diplôme d'études fondamentales vétérinaires le : 25/06/2015
a validé son année d'approfondissement le : 07/07/2016 ALFORT
n'a plus aucun stage, ni enseignement optionnel à valider.



Prédiction des statuts métabolique et de santé de chevaux d'endurance via l'intégration de données « omiques » : vers l'identification de biomarqueurs.

RESUME :

Cette thèse est composée d'une partie bibliographique et d'une partie expérimentale.

La première commence par présenter l'intérêt de l'intégration de données de différentes natures (phénomiques, génomiques, épigénomiques, transcriptomiques et microtranscriptomiques, protéomiques, et métabolomiques) dans une perspective de prédiction phénotypique (de performance, de santé, zootechnique...) puis donne un panorama des données « omiques » produites jusqu'à présent chez le cheval, et de leur valorisation dans les domaines de la recherche, de la médecine vétérinaire et de la gestion des populations. Les méthodes statistiques actuellement utilisées pour l'analyse et l'intégration de ces jeux de données de très grandes dimensions sont brièvement décrites. La partie bibliographique recentre ensuite la thèse sur l'utilisation et l'intégration de données « omiques » pour la prédiction de performances sportives d'endurance, avec en particulier la présentation du projet IFCE-INRA-Fonds Eperon GENENDURANCE, programme de recherche de grande envergure spécifiquement dédié à la performance d'endurance chez le cheval.

La deuxième partie de cette thèse présente des travaux expérimentaux inscrits dans le cadre du programme GENENDURANCE, dont les résultats ont fait l'objet d'une publication scientifique (Mach et *al.*, Scientific Reports | 6:22932 | DOI: 10.1038/srep22932, article paru le 16 mars 2016). Ce travail a permis de montrer que des relations entre miARN et ARNm sanguins régulés de manière spécifique par l'exercice d'endurance chez les chevaux révèlent des biomarqueurs uniques de la réponse de stress à l'endurance, et fournissent un aperçu du contrôle moléculaire de cette réponse (régulations post-transcriptomiques liées à un effort d'endurance) : 44 miARN réguleraient un ensemble de 351 gènes cibles sous-exprimés suite à une course d'endurance, gènes impliqués dans les voies biologiques du métabolisme glucidique, de l'oxydation des acides gras, de la biogenèse mitochondriale et de la réponse immunitaire, et 3 miARN (miR-21-5p, miR-181b-5p et miR-505-5p) ont été validés comme molécules candidates régulatrices de la réponse à l'exercice d'endurance chez le cheval.

MOTS CLES : Cheval, endurance, intégration de données, prédiction de performances, biomarqueurs, miARN.

Prediction of metabolic and health condition of endurance horses through –omic data integration: towards the identification of biomarkers.

SUMMARY:

This thesis consists in a bibliographic section and an experimental part.

The former sets forth the interest of integrating different kinds of data (phenomic, genomic, epigenomic, transcriptomic and microtranscriptomic, proteomic, and metabolomic) in the context of a phenotype prediction (performance, health, zootechnical...). It then provides a general overview of –omic data generated so far on horses, and of their promotion in the fields of research, veterinary medicine, and management of populations. A brief outline of statistical methods currently used for analysis and integration of these very large-dimension data sets is given. The bibliographic part of the thesis then focuses on –omics' use and integration for athletic performance prediction in endurance, especially with the presentation of the IFCE-INRA-Fonds Eperon GENENDURANCE project, a large-scale research program specifically dedicated to endurance performance in horses.

The latter presents the experimental work within the framework of the GENENDURANCE program, whose results were the subject of a scientific article (Mach & *al.*, Scientific Reports | 6:22932 | DOI: 10.1038/srep22932, article published on March 16th, 2016). This work demonstrated that whole blood miRNA-mRNA relationships specifically regulated by endurance exercise in horses reveal unique biomarkers of the stress response to endurance, and provide significant insights into the molecular control of this response (post-transcriptomic regulation associated with endurance exercise): 44 miRNA would indeed regulate a total of 351 target genes, underexpressed following an endurance race, involved in glucose metabolism, fatty acid oxidation, mitochondrion biogenesis, and immune response pathways, and 3 miRNA (miR-21-5p, miR-181b-5p and miR-505-5p) have been confirmed as candidate regulatory molecules for the response to endurance exercise in the horse.

KEYWORDS: Horse, endurance, data integration, performance prediction, biomarkers, miRNA.