

L'apport de la génomique pour un cheval sain et performant

Par :

- L. SCHIBLER
- INRA, Centre de recherche de Jouy en Josas
UMR1313 Génétique Animale et Biologie Intégrative
Equipe Biologie Intégrative et Génétique Equine
78350 JOUY EN JOSAS

Résumé

La génomique est une discipline de la biologie qui étudie la structure, le fonctionnement et l'évolution des génomes par des approches systématiques et à haut débit. Elle permet l'acquisition de connaissances à une échelle et à une vitesse qui révolutionne les approches scientifiques classiques. Ces connaissances et les avancées technologiques offrent de multiples perspectives en amélioration génétique des animaux ou en termes de gestion des populations. De même, les applications en biologie prédictive pourraient contribuer à l'évolution des pratiques zootechniques et vétérinaire dans le sens d'une meilleure prise en compte des particularités individuelles. Mais, loin des idées naïves « du tout génomique » des années 1990, en aucun cas elle ne constitue une solution miracle. La complexité du vivant fait que nous sommes encore très loin de comprendre comment la séquence d'ADN dirige la construction et le fonctionnement d'un être vivant.

Après avoir présenté les principales caractéristiques du génome et les principales méthodes de la génomique, cet article tentera d'en évaluer les potentialités à moyen terme en vue de la production d'un cheval sain et performant.

Mots clés : Génomique, transcriptomique, protéomique, anomalies génétiques, caractères complexes

Summary

Genomics studies the structure, the function and the genome evolution using high-throughput methods and holistic approaches. The advent of genomic technologies has brought a revolution as it dramatically transformed biologists' ability to acquire and analyze a huge amount of data. Genomic knowledge and technologies bring new opportunities in terms of livestock genetic improvement or genetic diversity management. Likewise, the emerging fields of predictive biology and personal genomics are promising and may lead to changes in animal management and veterinary medicine.

However, unraveling the genomic mysteries of life is more complicated than initially expected in the 90's, when genomics apologists claimed that miraculous applications were around the corner. Indeed, we are still far from understanding how the DNA sequence directs the development and function of living organisms.

We will first review some basic biological properties of the genome and DNA and describe the principles and methods underlying genomics. Finally, this paper will attempt to review genomic applications and opportunities for the equine industry in terms of horse health and performance.

Key-words : Genomics, transcriptomics, proteomics, genetic disease, QTL

Introduction

La génomique est une discipline de la biologie qui étudie la structure, le fonctionnement et l'évolution des génomes par des approches systématiques et à haut débit. Elle comporte 1) la génomique structurale qui étudie l'architecture et l'évolution des génomes et dont l'objectif est l'obtention de la séquence du génome; 2) la génomique fonctionnelle, qui vise à déterminer la fonction et l'expression des gènes en caractérisant l'ensemble des transcrits (transcriptome) et des protéines (protéome) ainsi que les éléments de leur régulation.

Si ces travaux ont un rôle essentiel dans le développement des connaissances et la compréhension du vivant, ils offrent également des applications directes en amélioration génétique des espèces d'élevage. Des applications en biologie prédictive pourraient de plus contribuer à l'évolution des pratiques zootechniques et vétérinaire dans le sens d'une plus grande prise en compte des particularités individuelles. Conduite différenciée ou adaptée des animaux et pharmacogénomique pourraient ainsi connaître un essor important au cours de la prochaine décennie.

Après avoir présenté les principales caractéristiques du génome et les principales méthodes de la génomique, cet article tentera d'en évaluer les potentialités à moyen terme en vue de la production d'un cheval sain et performant.

1. Structure et expression du génome

Les organismes vivants se composent de cellules renfermant toute l'information définissant leurs caractéristiques biologiques, leur permettant de se développer et de se reproduire. Les travaux de Mendel en 1865 ont montré que ces caractéristiques étaient transmises à la descendance (caractères héréditaires) et Johannsen en 1909 dénomma **gènes** les unités héréditaires codant ces caractères. L'ensemble des gènes d'un organisme constitue son **génome**. Un gène peut prendre plusieurs formes (allèles) et l'ensemble des allèles d'un organisme constitue son **génotype**. L'interaction génotype et environnement conditionne l'ensemble des caractères de l'organisme (**phénotype**).

1.1. Les chromosomes et l'ADN, support de l'hérédité

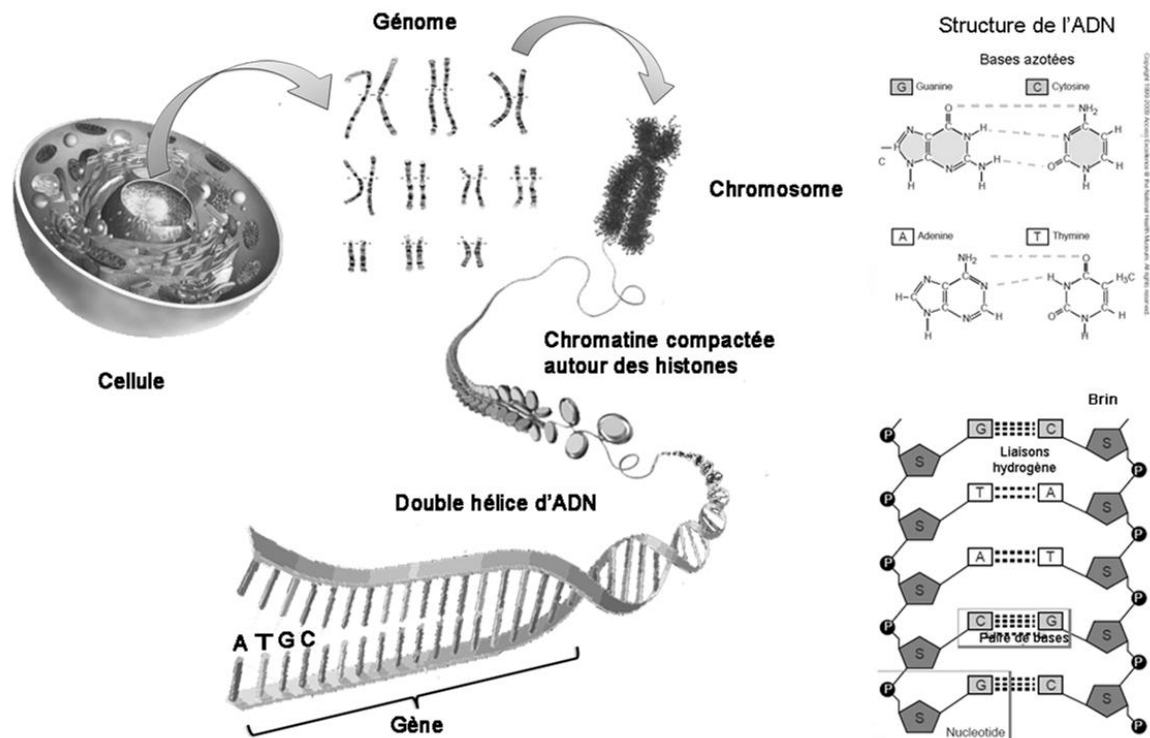
Selon la théorie chromosomique de l'hérédité établie par Morgan, les gènes sont portés par les **chromosomes**, des éléments microscopiques contenus dans le noyau des cellules (Figure I). Chez les mammifères, les cellules comportent deux lots de N chromosomes homologues, l'un d'origine paternelle, l'autre d'origine maternelle : elles sont **diploïdes** (2N). Une division cellulaire particulière permet un retour à l'état haploïde (N) au cours de la formation des gamètes : la **méiose**. Les gamètes héritent ainsi d'une combinaison aléatoire de chacun des deux chromosomes d'une même paire et sont donc tous différents.

De plus, lors de la méiose, des échanges réciproques de fragments peuvent se produire entre deux chromosomes homologues : c'est la **recombinaison génétique**. Comme la probabilité de voir apparaître une recombinaison entre deux régions (loci) est fonction de la distance les séparant, Morgan proposa d'utiliser la fréquence de recombinaison comme mesure de distance entre gènes : 1 centimorgan (cM) correspond ainsi à 1 % de descendants au génotype recombiné.

Au plan chimique, les chromosomes sont composés d'une molécule d'acide désoxyribonucléique (**ADN**), plus ou moins compactée autour de protéines (histones), formant la chromatine. L'ADN est une macromolécule constituée de deux brins résultant de l'enchaînement de quatre nucléotides différents par leur base azotée (Figure I) : l'adénine (A), la guanine (G), la cytosine (C) et la thymine (T). Les deux brins sont appariés par leur bases complémentaires (un A est toujours en face d'un T, un C en face d'un G) et s'organisent selon une structure en double hélice, découverte en 1953 par Watson, Crick et Wilkins. Cette structure explique le maintien de l'information génétique au cours des divisions cellulaires. En effet, l'ADN est répliqué avant la division: la double hélice d'ADN s'ouvre et un brin complémentaire est synthétisé le long de chaque brin initial par ajout de nucléotides respectant la complémentarité des bases. Deux nouvelles molécules d'ADN double brin en hélice, identiques au modèle original sont ainsi produites *in fine*. Cet alphabet de 4 lettres permet de coder toutes les informations susceptibles de créer et de faire vivre un organisme.

Un **gène** correspond ainsi à un segment de la chaîne d'ADN dont la séquence de nucléotides à une signification fonctionnelle et code le plus souvent pour une protéine qui assure la fonction biologique. Certains gènes ne codent cependant pas pour une protéine, mais pour divers acides nucléiques ayant des fonctions dans la synthèse protéique ou la régulation de l'expression.

Figure I : Génome, Chromosome, ADN et Gènes
 Figure I: Genome, Chromosome, DNA and Genes



L'ADN peut subir des changements (**mutations**) sous l'effet d'agents extérieurs (UV, radioactivité, produits chimiques...) ou, bien que ce mécanisme soit d'une grande fidélité, lors de la réplication. Ces mutations peuvent être des substitutions d'une base par une autre, des insertions ou des délétions. Les **allèles** correspondent aux multiples versions d'un même gène qui diffèrent les uns des autres par une ou plusieurs mutations. Si les allèles apportés par chaque parent sont identiques, l'individu est **homozygote** pour ce gène et **hétérozygote** sinon. Dans ce dernier cas, si l'un des allèles s'exprime alors que l'autre est muet, le premier est dit **dominant** et le second **récessif**. Le phénotype est déterminé par l'allèle dominant. Si les deux allèles contribuent au phénotype, ils sont dits **codominants**.

1.2. De l'ADN aux protéines.

Les protéines sont les molécules les plus complexes et les plus variées des êtres vivants et près de 100 000 seraient codées par le génome des mammifères. Elles remplissent de nombreux rôles, dont des fonctions structurales, de transport, de communication, de défense ou d'enzyme. Il s'agit de macromolécules constituées d'un enchaînement de quelques à plusieurs centaines d'acides aminés. Vingt acides aminés entrent dans la composition des protéines et leur ordre, codé par le génome, constitue la séquence primaire. Les chaînes se replient pour adopter une structure tridimensionnelle et peuvent s'assembler entre elles, ce qui leur confère leurs propriétés fonctionnelles.

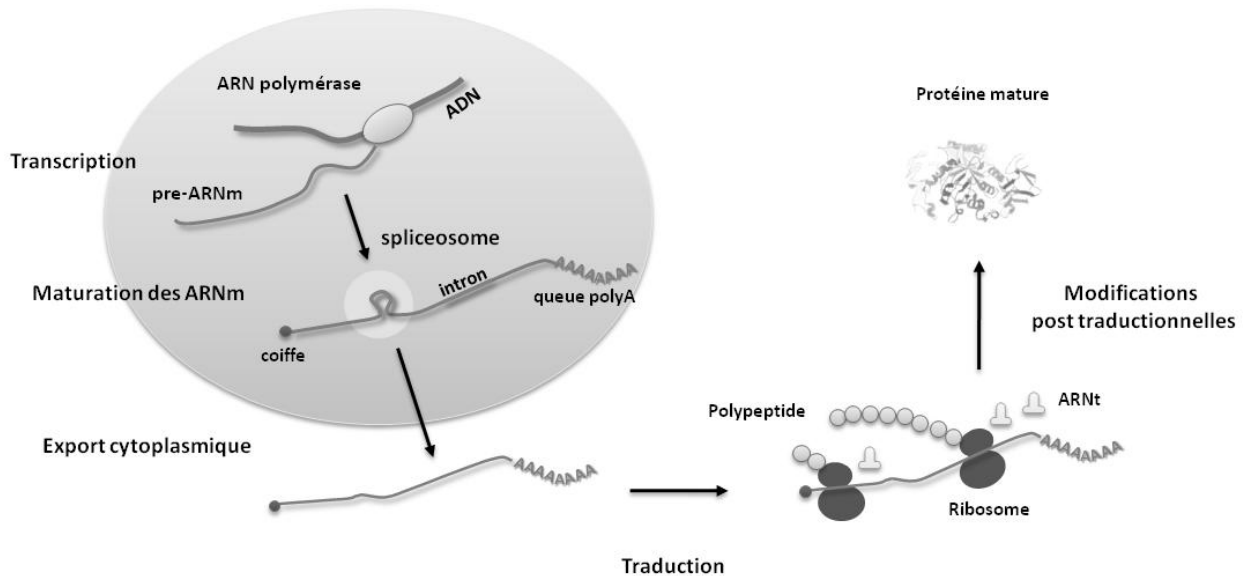
La synthèse des protéines débute par la création de plusieurs «copies de travail» des gènes : c'est la **transcription** en acide ribonucléique (**ARN**). Tout comme l'ADN, les ARN sont constitués de l'enchaînement de nucléotides mais ils sont simple brin, comportent un ribose et une uracile remplace la thymine. La transcription est assurée par une ARN polymérase, à partir d'un site promoteur et jusqu'à un site de terminaison. Seul un des brins d'ADN est transcrit et donne naissance à un ARN pré-messager comportant des **introns** et des **exons**. Celui-ci est modifié dans le noyau par l'ajout d'une coiffe, l'adjonction d'une séquence de 20 à 250 nucléotides d'adénine (queue poly A) et l'excision des introns (**épissage**), avant d'être exporté hors du noyau sous forme d'ARN messager (**ARNm**) (Figure II).

L'information génétique des ARNm, codées par un alphabet à 4 bases, est alors traduite en alphabet protéique à 20 acides aminés pour synthétiser les protéines : c'est la **traduction**. Le système de transcodage constitue le code génétique et repose sur une combinaison de 3 nucléotides (codon). Plusieurs codons codent pour le même acide aminé, 3 codons commandent l'arrêt de la synthèse (codons stop) et le codon AUG est le signal d'initiation de la traduction (codon-initiateur). Ce processus complexe repose sur l'action des **ribosomes** et des ARN de transfert (**ARNt**).

Les protéines subissent d'abondantes modifications pendant et après leur synthèse (**modifications post-traductionnelles**). Il peut s'agir de la modification de certains acides aminés par ajout ou retrait de glucides, d'acides gras, de groupements divers (phosphates, sulfate, carboxyle...) ou par protéolyse contrôlée. Ces modifications contribuent à la régulation de leur activité, permettent leur adressage vers certains compartiments cellulaires ou leur reconnaissance par les systèmes de dégradation...

Ainsi, à partir d'un gène, plusieurs centaines de milliers d'exemplaires d'une même protéine peuvent être produits rapidement. De plus, un même gène peut donner naissance à plusieurs types de protéines, via l'utilisation de promoteurs alternatifs ou par épissage alternatif.

Figure II : De l'ADN aux protéines.
Figure II: From DNA to proteins.



1.3. La régulation de l'expression du génome.

Toutes les cellules d'un organisme contiennent les mêmes gènes mais seuls certains gènes s'expriment dans un type cellulaire donné. De même, l'expression du génome est finement contrôlée en réponse aux conditions environnementales. Différents mécanismes de régulation ont été découverts, contribuant aux interactions génome-environnement et au lien entre génotype et phénotype.

La régulation chromatinienne constitue un premier niveau de régulation, jouant sur la condensation de la chromatine et son accessibilité. La chromatine « ouverte » ou euchromatine est ainsi particulièrement active en termes de transcription, au contraire de l'hétérochromatine condensée. Diverses modifications des histones (acétylation ou méthylation) ou de l'ADN (méthylation des cytosines) modulent la condensation de la chromatine et l'expression des gènes.

La transcription est régulée par des éléments régulateurs situés dans le promoteur. Ces sites permettent la fixation de divers facteurs facilitant ou bloquant la transcription. Des séquences situées à distance de la séquence transcrite augmentent l'action du promoteur (activateurs ou enhancers) ou la répriment (répresseurs ou silencers).

A la fin des années 90, deux nouveaux types de petits ARN non codants ont été découverts : les micro-ARN (**miRNA**) et les petits ARN interférents (**siRNA**). Les miRNA sont codés par des gènes spécifiques, alors que les siRNA sont produits par utilisation des parties exoniques d'un gène. Après maturation par des complexes enzymatiques, ces petits ARN peuvent s'apparier sur leurs ARNm cibles, provoquant l'inhibition de leur traduction ou leur dégradation.

2. Les outils de la génomique

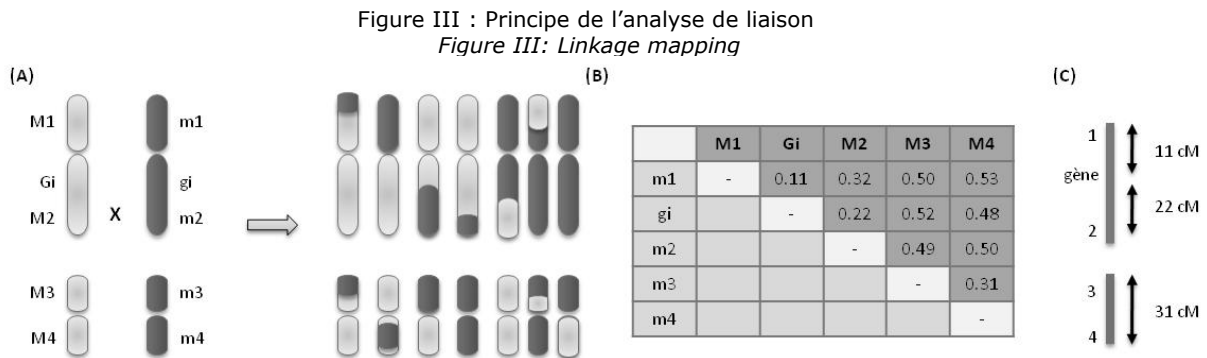
La génomique a stimulé les développements technologiques et la construction de nombreux outils, dont beaucoup sont devenus obsolètes avec l'assemblage des séquences complètes des génomes. Seuls seront donc évoqués ici les concepts et outils actuels.

2.1. La cartographie

2.1.1 L'analyse de liaison

L'analyse de liaison exploite la recombinaison génétique au sein de familles pour positionner de façon relative des loci les uns par rapport aux autres. Il est ainsi possible non seulement d'établir des cartes génétiques du génome, mais également de localiser des gènes d'intérêt sur la seule base du phénotype, que le caractère soit simple (1 gène gouverne le caractère) ou complexe (plusieurs gènes interviennent), binaire (sain versus malade) ou quantitatif (taille des animaux). Dans le cas de caractères complexes et quantitatifs, la méthode permet de détecter des loci expliquant chacun une part de la variance phénotypique observée : ce sont les QTL (quantitative trait loci).

La précision de localisation dépend du nombre de recombinaisons observées entre 2 loci. Améliorer la localisation nécessite donc de multiplier les familles et le nombre de descendants par familles.



(A) Le croisement de 2 génomes parentaux (haploïdes pour simplifier le schéma) génère des chromosomes recombinants.

(B) Le génotype de la descendance est déterminé pour de nombreux marqueurs moléculaires (2 allèles M et m) et le marqueur phénotypique (2 allèles Gi et gi). La fréquence de recombinaison est calculée pour chaque paire de marqueurs.

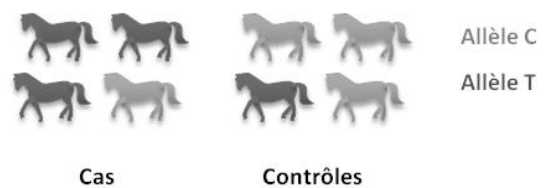
(C) La carte génétique la plus probable est calculée à partir de toutes les données à l'aide d'un logiciel de cartographie

2.1.2 L'analyse d'association

Cette approche cherche à déceler une association entre un variant génétique et le caractère, à un niveau populationnel et non plus familial uniquement. Il ne s'agit plus de corrélérer la transmission d'un locus et celle du caractère, mais de comparer la fréquence des allèles au sein des groupes d'individus classés en fonction de leur statut. Elle exploite les nombreuses recombinaisons générées au cours du temps au sein de populations, permettant une cartographie très précise.

L'étude d'association est utilisée classiquement pour tester l'implication possible de gènes candidats dans un phénotype donné. Depuis quelques années, la communauté scientifique porte beaucoup d'espoir sur les approches d'association pangénomique (GWAS pour Genome Wide Association Studies) pour l'analyse des caractères complexes. Ces approches GWAS nécessitent cependant une forte densité de marqueurs et des effectifs importants.

Figure IV : Principe de l'analyse d'association

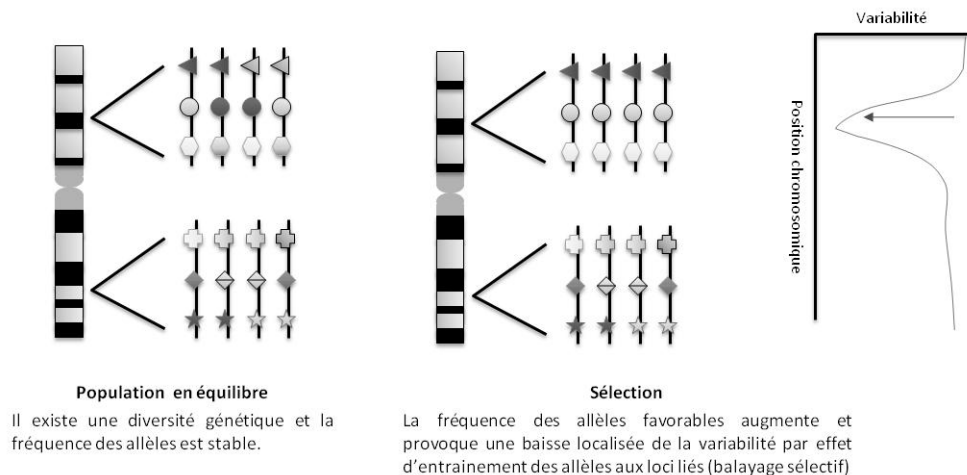


Si un marqueur est suffisamment proche du locus d'intérêt, un de ces allèles (ici l'allèle T) est présent à une fréquence plus élevée dans la population des cas que dans la population contrôle

2.1.3 L'étude des signatures de sélection

L'apparition d'une mutation favorable apportant un avantage sélectif est sélectionnée, augmentant progressivement la fréquence des allèles favorables dans la population. Les allèles des loci présents au voisinage sont également sélectionnés par effet d'entraînement (autostop génétique). Il en résulte une baisse locale de variabilité génétique.

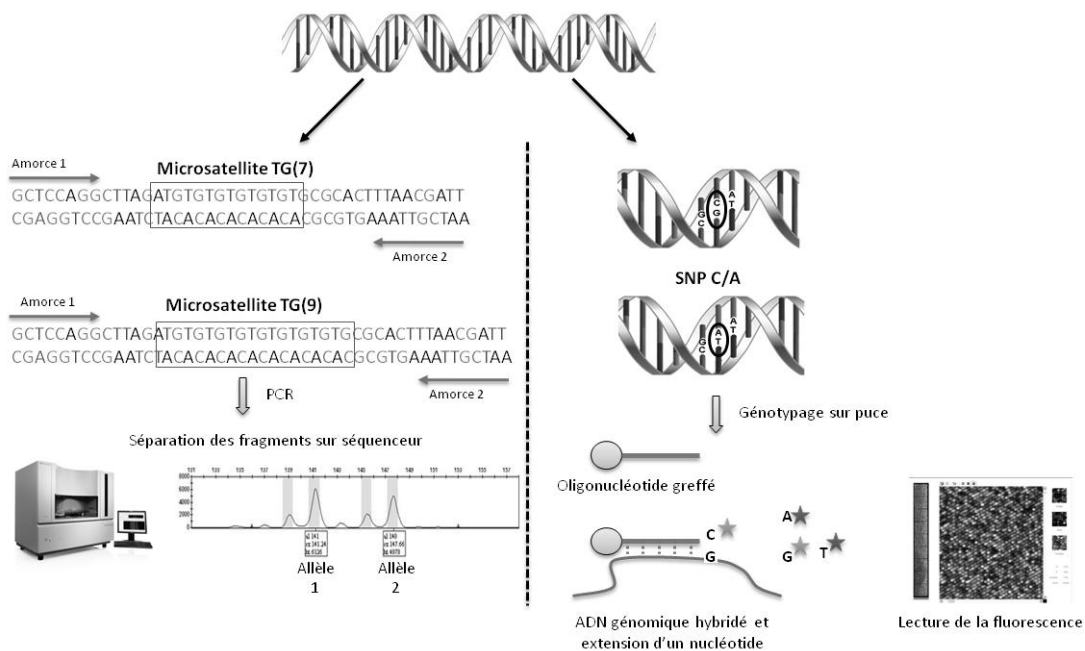
Figure V : Principe de détection de signature de sélection
 Figure V: Hitchhiking mapping



2.1.4 Les marqueurs et les technologies de génotypage

Les approches décrites précédemment nécessitent de déterminer l'origine des fragments chromosomiques ou de discriminer les différents variants d'un locus. Deux types de marqueurs polymorphes sont principalement utilisés à cette fin : les microsatellites et les SNPs (Figure VI).

Figure VI : Marqueurs microsatellites et SNPs.
 Figure VI: Microsatellite and SNP markers.



Les microsatellites sont des répétitions en tandem d'un motif de 1 à 10 nucléotides. Les dinucléotides (TG)_n sont les plus fréquents et les plus utilisés chez les mammifères car ils présentent une grande variabilité du nombre de répétitions. Il existe en moyenne 1 microsatellite tous les 25 kb, encadré par des séquences uniques permettant de les localiser sur le génome. Le génotypage est réalisé par amplification *in vitro* à l'aide d'amorces spécifiques (PCR) et les allèles sont distingués en fonction de leur taille par électrophorèse sur gel de polyacrylamide.

Les Single Nucleotide Polymorphisms (SNPs) correspondent à une substitution d'un seul nucléotide. Ils sont très abondants (tous les 0,5 à 1 kb) et distribués uniformément dans tout le génome. Des technologies de génotypage à haut débit sur puces à ADN (microarray) ont été développées, basées sur la capacité de l'ADN simple brin à s'hybrider avec un autre ADN simple brin complémentaire. Des oligonucléotides spécifiques des séquences adjacentes des SNPs sont ainsi greffés sur la puce et hybridés

avec l'ADN à génotyper. Le produit hybridé est ensuite allongé avec des nucléotides fluorescents. La fluorescence détectée par un scanner permet de connaître le type de nucléotide incorporé et donc le génotype. Chez le cheval une puce Illumina de 54 000 SNP est disponible et une nouvelle version comportant 74 000 SNP sera commercialisée prochainement. La technologie permet d'ores et déjà d'atteindre des densités de plus de 2 millions de SNPs par puce.

2.2. La transcriptomique

Les technologies de puce à ADN sont également employées pour mesurer l'expression de plusieurs milliers de gènes à la fois. Pour cela, les ARNm sont extraits des tissus, amplifiés et marqués avec un fluorophore avant d'être hybridés sur la puce qui contient de courtes séquences (oligonucléotides) complémentaires des différents ARN.

Il existe des puces d'expression standard, comportant quelques oligonucléotides par gène et couvrant de 30 000 à 50 000 gènes. Il existe également des puces « exons » couvrant l'intégralité des exons, soit près d'un million de sondes. Les puces exons permettent d'analyser non seulement les niveaux d'expression mais également les phénomènes d'épissage alternatif ou d'utilisation de promoteurs alternatifs. Chez le cheval, seule une puce standard Agilent est disponible, comportant 15 000 gènes.

Les puces de tiling comportent des oligonucléotides couvrant tout le génome de façon chevauchante (décalage de 10 nucléotides). L'hybridation d'ARN sur ces puces permet de détecter de nouveaux transcrits et d'analyser les différentes formes d'épissage alternatif des transcrits de ces gènes. Elles sont également employées pour identifier les sites de fixation de facteurs de transcription (ChIP on chip), l'état de condensation de la chromatine ou la méthylation des bases (MeDIP chip). Dans ces derniers cas, une immuno-précipitation de la chromatine est réalisée avec un anticorps reconnaissant le facteur de transcription ou l'ADN méthylé avant hybridation sur la puce. Il n'existe pas de telles puces commerciales chez le cheval, mais certains fournisseurs comme Roche-Nimblegen permettent d'en fabriquer à façon pour un coût raisonnable.

2.3. La protéomique

La protéomique étudie l'ensemble des protéines (protéome) d'un échantillon. Elle permet d'obtenir une image globale des synthèses biomoléculaires dans la mesure où le protéome reflète les événements cellulaires au niveau traductionnel et post-traductionnel. Classiquement, les protéines sont séparées par électrophorèse bidimensionnelle (2-DE) sur gel de polyacrylamide. Elle consiste en une séparation des protéines dénaturées selon leur charge (isoélectrofocalisation) puis selon leur poids moléculaire. Une analyse quantitative est ensuite réalisée après coloration (bleu colloïdal, nitrate d'argent, Sypro Ruby ou cyanines fluorescentes). Les protéines sont ensuite extraites des gels pour être finalement identifiées par spectrométrie de masse.

2.4. Le séquençage

Le séquençage correspond à la lecture nucléotide par nucléotide de la molécule d'ADN, ce qui permet de connaître la succession des 4 lettres de l'alphabet nucléique. La mise au point de séquenceurs automatiques, fondés sur la méthode de Sanger publiée en 1977, a permis l'essor du séquençage dans les années 90 et a stimulé les projets d'étude de génomes complets.

Depuis 2005 de nouvelles méthodes de séquençage massif ont été développées (Nouvelles Générations de Séquençage, **NGS**). Ces approches parallélisent et automatisent la production des fragments (banques) et leur séquençage. Différentes technologies (Roche 454, ABI SOLID, Illumina) permettent ainsi de séquencer plusieurs millions de fragments en parallèle. Le séquenceur de dernière génération Illumina HiSeq2000 peut ainsi produire 2 milliards de séquences de 100 bases et générer 200 Gb par semaine, soit l'équivalent de 66 génomes humains complets.

Les NGS s'appliquent à tous les domaines de la génomique qu'elles sont en train de révolutionner, tant par la sensibilité des techniques, les volumes de données et les difficultés de leur traitement que par les nouvelles applications qu'elles laissent entrevoir.

- En termes de structure des génomes, les NGS vont démocratiser le séquençage *de novo* de nouvelles espèces. Le projet « arbre de la vie » ambitionne ainsi de séquencer un millier d'espèces animales et végétales. Pour cela, l'institut de génomique de Pékin (Beijing Genomics Institute, BGI), qui s'est illustré en séquençant le génome du panda et l'ADN d'un homme préhistorique, emploie plus de 500 bioinformaticiens et s'est doté en 2010 de 140 HiSeq2000.
- Appliquées à l'expression des génomes, les NGS permettent d'analyser le transcriptome (RNA-seq) en couplant quantification et détection de nouveaux transcrits, y compris les petits ARN. Le

séquençage après immunoprécipitation permet d'identifier les emplacements de liaison des protéines sur l'ADN (ChIP-seq) ou de caractériser les régions méthylées (MeDIP-seq).

- Le réséquençage à grande échelle permet une évaluation sans précédent du polymorphisme existant au sein d'une espèce ou d'une race, ouvrant la voie à des analyses fines de biodiversité, de recherche de traces de sélection ou d'analyse de variants rares.

3. Les applications de la génomique pour un cheval sain et performant

3.1. Identification de gènes et conseil génétique

Il existe plusieurs stratégies d'identification de gènes d'intérêt: 1) l'approche « gène candidat » qui cherche à démontrer une association entre le caractère et une mutation détectée dans un gène connu pour son rôle dans un caractère similaire chez une autre espèce; 2) l'approche de clonage positionnel, qui détermine la région chromosomique d'intérêt par cartographie, souvent couplée à une approche de génomique fonctionnelle, permettant de cibler le gène d'intérêt au sein de la région. Les deux approches ont été utilisées chez le cheval pour mettre en évidence les bases moléculaires de caractères monogéniques tels la coloration et certaines pathologies ou anomalies, donnant lieu à des applications pratiques sous forme de tests de dépistage. La grande majorité des caractères d'intérêt sont cependant quantitatifs et sont gouvernés par l'interaction de plusieurs gènes et de l'environnement. Il est difficile de discerner la contribution de chacun des gènes dans le phénotype, ce qui pose des difficultés tant dans l'identification des gènes sous jacents que dans l'exploitation pratique des résultats.

3.1.1 Gènes responsables d'affections monogéniques chez le cheval

Peu d'anomalies génétiques sont décrites chez le cheval en comparaison des autres espèces. La base Online Mendelian Inheritance in Animals (OMIA) recense ainsi 23 anomalies, les gènes responsables ayant été identifiés dans la moitié des cas (FINNO *et al.* 2009) (Tableau 1). La sélection basée sur les performances sportives, privilégiant les individus présentant une santé et une conformation irréprochables, explique probablement en partie ces chiffres. Une méconnaissance du déterminisme génétique de nombreuses pathologies, liée aux difficultés de collecte des informations ainsi qu'au coût et à la durée des investigations génétiques, y contribuent également. Au niveau international, plusieurs programmes d'identification sont en cours (voir <http://dga.jouy.inra.fr/horse.genomics/hgpprojects.html>).

Les affections à déterminisme dominant sont faciles à étudier par clonage positionnel et nécessitent des dispositifs de quelques centaines d'animaux répartis en quelques grandes familles. Elles semblent peu nombreuses chez le cheval. L'exemple de l'Hyperkaliémie périodique paralysante (HYPP) observée en race Quarter Horse illustre cependant une situation où une telle pathologie a été sélectionnée car associée à une hypertrophie musculaire, caractère recherché par les éleveurs. Les chevaux malades sont porteurs d'une mutation dans le gène SCN4A codant pour une glycoprotéine jouant un rôle de canal sodium. Les cellules musculaires sont hyperexcitables, ce qui produit l'hypertrophie musculaire. En situation de stress, d'alimentation riche en potassium ou de froid, les animaux présentent cependant une faiblesse musculaire et des tremblements pouvant aller jusqu'à la paralysie.

Les affections récessives sont plus insidieuses : les animaux porteurs d'un seul allèle défavorable sont sains et seuls leurs descendants homozygotes pour l'allèle défavorable sont malades. L'identification des gènes et mutations responsables de ces affections récessives s'avère particulièrement efficace et requiert le génotypage sur puce SNP d'une dizaine d'individus atteints. Une délétion dans le gène de la myosine Va (MYO5A) a ainsi pu être identifiée en analysant 6 poulains arabes atteints du syndrome du poulain lavande (Lavender Foal Syndrom) et 30 collatéraux sains. Le syndrome léthal du poulain blanc (OLWFS, Overo Lethal White Foal Syndrome) est un autre exemple d'affection récessive, sélectionnée avec le patron de coloration blanc. Une substitution dans le gène codant le récepteur de l'endothéline β (EDNRB) est responsable du patron de coloration overo, transmis selon un mode dominant. Le croisement d'animaux overo donne naissance à 50 % de poulains overo, 25 % de poulains unicolores et 25 % de poulains blancs mourant rapidement de mégacôlon, une pathologie liée à une absence de motricité intestinale. La prévalence de ces pathologies est généralement faible mais peut augmenter à la faveur d'une utilisation intensive d'un reproducteur porteur. Alors qu'une large diffusion d'un nombre réduit de taureaux élités favorise l'émergence de ce type de pathologie chez les bovins, le nombre contingenté de cartes par étalons limite ce risque chez le cheval. *A contrario*, les nombreuses races de chevaux à petits effectifs risquent d'être de plus en plus souvent confrontées à ce type d'affections.

Loin de jeter le discrédit sur certaines filières ou certaines lignées, la génomique peut fournir des outils permettant de faire face en toute sérénité et sans « chasse aux sorcières » à l'émergence de certaines affections héréditaires : il devient possible de prévenir la naissance de poulains affectés, sans pour autant appauvrir le pool génétique. En effet, la connaissance de ces gènes et des mutations permet

un dépistage précoce des animaux porteurs d'affections génétiques, sans nécessité de validation sur descendance. Dans le cas des pathologies récessives, un simple conseil génétique peut être mis en œuvre de façon à éviter les accouplements à risque et à réduire l'incidence de la pathologie dans la population, sans pour autant perdre les qualités génétiques des animaux porteurs. La connaissance précoce d'une prédisposition génétique à une pathologie peut également permettre de prévenir sa survenue, en adaptant la conduite des animaux (régime alimentaire, type d'exercice...).

Tableau 1 : Gènes responsables de maladies génétiques
Table 1: Genes involved in genetic disease

Maladie génétique	Description	Gène Marqueur	Méthode	Source
HYPP (D)	Hyperkaliémie périodique paralysante	SCN4A	Patho comparée Liaison	RUDOLPH <i>et al.</i> 1992a; 1992b
OLWS (R)	Syndrome léthal du poulain blanc Overo	EDNRB	Patho comparée Association	METALLINOS <i>et al.</i> 1998; SANTSCHI <i>et al.</i> 1998; YANG <i>et al.</i> 1998
SCID (R)	Immunodéficiência sévère combinée	PRKDC	Liaison Candidats	BAILEY <i>et al.</i> 1997; BERNOCO and BAILEY 1998; SHIN <i>et al.</i> 1997
JEB-H (R)	Epidermolyse bulleuse jonctionnelle (Herlitz)	LAMC2	Patho comparée Association	GRAVES <i>et al.</i> 2009; MILENKOVIC <i>et al.</i> 2003; SPIRITO <i>et al.</i> 2002
EI (R)	Epitheliogenesis imperfecta	LAMA3	Patho comparée Liaison Association	GRAVES <i>et al.</i> 2009; LIETO and COTHRAN 2003
GBED (R)	Déficiência en enzyme branchante du glycogène	GBE1	Physiopathologie Association	VALBERG <i>et al.</i> 2001; WARD <i>et al.</i> 2004
MH (D)	Hyperthermie maligne	RYR1	Association Candidats	ALEMAN <i>et al.</i> 2004
HERDA (R)	Hyperélastose cutanée	PPIB	Homozygosity mapping	TRYON <i>et al.</i> 2007
PSSM (D)	Myopathie à stockage de polysaccharide	GYS1	Association pangénomique	HERSZBERG <i>et al.</i> 2009; MCCUE <i>et al.</i> 2008
Mélanome (D)	Susceptibilité au mélanome	STX17	Liaison Association	ROSENGREN PIELBERG <i>et al.</i> 2008
MCOA (C)	Anomalies oculaires congénitales multiples	Marqueurs	Liaison	ANDERSSON <i>et al.</i> 2008
CSNB (R)	Cécité nocturne congénitale non évolutive	(TRPM1)	Liaison Association	BELLONE <i>et al.</i> 2010
LVS (R)	Syndrome du poulain lavande	MYO5A	Association pangénomique	BROOKS <i>et al.</i> 2010
CA (R)	Abiotrophie cérébelleuse	MUTYH	Liaison Homozygosity mapping	BRAULT <i>et al.</i> 2010

Mode de transmission : dominant (D), récessif (R) ou codominant (C).

3.1.2 Gènes majeurs et aptitudes sportives

Plusieurs stratégies de valorisation peuvent être envisagées à partir de l'identification de gènes majeurs en lien avec des aptitudes sportives et la mise à disposition d'un test génétique :

- Une approche de sélection des individus portant les allèles favorables peut permettre aux éleveurs et évalonniés d'affiner la sélection des individus.
- Une approche de dépistage peut permettre de définir précocement les aptitudes d'un poulain et ainsi de mieux orienter et gérer sa carrière.
- D'un point de vue marketing, de tels tests peuvent aider à segmenter un marché et ainsi permettre de répondre de façon optimisée à des demandes variées voire contradictoires.
- L'introgression assistée par marqueurs peut permettre d'introduire les allèles favorables dans une race où ils sont absents, en réduisant le nombre de croisements nécessaires.

L'aptitude sportive, tout comme la performance à laquelle elle est liée, constitue un caractère synthétique éminemment complexe, abordé sous l'angle de l'adaptation à l'exercice. En caractérisant la régulation de l'expression du génome et en étudiant la physiologie de la réponse à l'entraînement, les

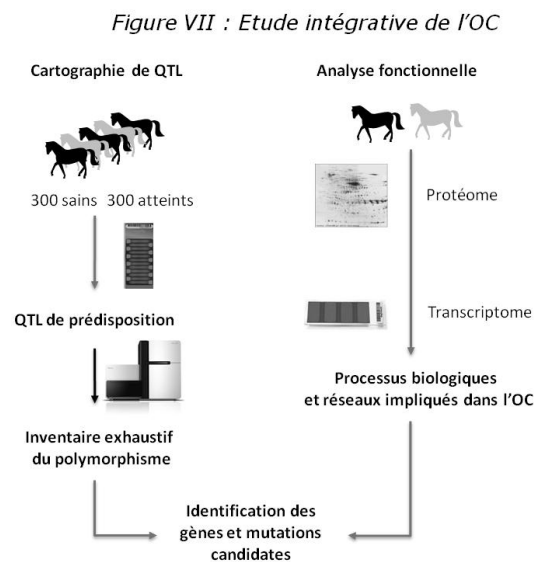
chercheurs espèrent identifier des gènes clés ayant un rôle crucial. Une étude transcriptomique a ainsi analysé les modifications d'expression induites précocement à la suite d'un exercice et durant la phase de récupération chez le pur sang anglais (McGIVNEY *et al.* 2009). Les résultats montrent une adaptation du métabolisme et la perturbation de multiples voies impliquées dans l'hypertrophie musculaire. Une analyse comparative du transcriptome musculaire avant et après 10 mois d'entraînement a également été réalisée en 2010, montrant une diminution de l'expression de 58 gènes, dont le gène de la myostatine (MSTN) (McGIVNEY *et al.* 2010). Or, cette protéine régule négativement la masse musculaire: des mutations « perte de fonction » produisent un phénotype hypermusclé (culard bovin par exemple). Une recherche de polymorphisme par séquençage du gène MSTN équin, ainsi qu'une analyse d'association pangénomique (187 chevaux) ont montré qu'un polymorphisme C>T dans l'intron 1 du gène était associé aux aptitudes sportives : les chevaux C/C sont plus doués pour les courses rapides (sprinters), les individus T/T ont plus d'endurance et gagnent les courses de fond (stayers) alors que les hétérozygotes C/T peuvent être d'excellents milers. Dans ce cas particulier, l'héritabilité de l'aptitude à la distance étant très élevée, le gène MSTN semble avoir un effet très fort. Equinome, une firme de biotechnologie irlandaise, commercialise d'ores et déjà un test génétique (Speed Gene test : le test du gène de la vitesse) permettant aux propriétaires et entraîneurs de chevaux de course de déterminer la distance optimale de course de leurs poulains et d'orienter en conséquence leur l'entraînement.

3.1.3 Les caractères complexes

Ces caractères sont déterminés par de nombreux facteurs génétiques et environnementaux qui pris isolément ont des effets modestes sur la variation phénotypique. L'analyse génétique de ces caractères met en évidence des loci à effets quantitatifs (QTL) qui expliquent une faible proportion de la variabilité du caractère. La principale difficulté des programmes de détection de QTL réside dans l'obtention d'effectifs suffisants pour assurer une bonne puissance de détection (à l'idéal plus de 1000 animaux). L'identification des gènes et mutations sous jacentes constitue également un véritable challenge et l'on compte actuellement peu de réussites, même dans des espèces modèles comme la souris.

Sauf exception, il serait probablement peu efficace de proposer une application directe de ces résultats sous forme de tests génétiques compte tenu de la faible part de chaque gène dans la variation du caractère. Par contre, la prise en compte de ces informations, même partielles, dans l'indexation des reproducteurs permettrait un progrès génétique.

Le pari fait actuellement par les chercheurs consiste à intégrer des informations positionnelles (QTL) et fonctionnelles (étude du transcriptome et du protéome), en espérant qu'une meilleure connaissance des processus biologiques impliqués permettra, d'une part, d'améliorer la précision du phénotype (et donc la localisation des QTL) et d'autre part, d'optimiser la recherche de gènes candidats au sein des régions QTL. C'est la stratégie suivie par exemple dans le cas du programme d'étude de l'ostéochondrose (OC) en trotteur français : une détection de QTL est effectuée grâce à une collaboration entre l'INRA, l'ENVA, le Cirale et l'IFCE (projet ANR GenEquin). Une recherche systématique de polymorphisme au sein des régions identifiées est réalisée par séquençage NGS (projet HiseqOC financé par le Département de Génétique Animale de l'INRA). Parallèlement, une analyse fonctionnelle est menée pour comparer le transcriptome et le protéome du cartilage et de l'os sous-chondral de chevaux sains et atteints (projet ANR Biocart, associant l'INRA, l'IFCE et l'ENVL).



La recherche de signatures de sélection pourrait constituer une nouvelle approche particulièrement efficace chez le cheval, permettant d'aborder l'identification de gènes importants, ayant probablement contribué à la différenciation des races. Au sein d'une race, certains variants de ces gènes pourraient être à l'origine d'une partie de la variabilité phénotypique. Une analyse a ainsi été réalisée en 2009 en pur sang anglais, à l'aide de 384 microsatellites, mettant en évidence 19 régions (Gu *et al.* 2009). Plusieurs gènes impliqués dans le développement musculaire, le métabolisme lipidique et la voie de signalisation de l'insuline sont présents au sein de ces régions. Une étude d'association à l'aide de 68 SNP identifiés dans 17 de ces gènes a par la suite été réalisée sur un panel de 148 chevaux. Une association avec les performances a pu être mise en évidence pour 13 de ces SNPs (HILL *et al.* 2010). De même, l'INRA et l'IFCE ont engagé fin 2010 un projet d'analyse de la diversité génomique de 4 races de chevaux français : le Selle Français, l'Anglo Arabe (type course et sport), le Cob et le Percheron. Les objectifs de cette étude

sont multiples. Il s'agit tout d'abord de participer à un consortium international s'intéressant aux aspects de biodiversité. De plus, le dispositif choisi permettra d'étudier les flux de gènes entre ces races et d'analyser finement leur structure génétique (quelle part du génome des SF provient du Cob, quelle sont les régions Ar et PS retenues dans les différents types d'AA...). Enfin, il devrait être possible de mettre en évidence des régions du génome sélectionnées en lien avec les aptitudes particulières de chaque race. Une telle étude pilote permettra d'évaluer les potentialités de l'approche « signature de sélection » en termes d'identification de gènes d'intérêt et d'amélioration génétique.

3.2. Recherche de biomarqueurs

3.2.1 Notion de biomarqueurs

Un biomarqueur est une caractéristique biologique mesurable liée à un processus normal ou non. En médecine, l'intérêt porte sur des molécules biologiques présentes dans le sang, dans les liquides corporels ou certains tissus qui peuvent témoigner d'un trouble, d'une maladie ou de l'exposition à des produits chimiques. Un biomarqueur génomique *strico sensu* désigne une caractéristique mesurable de l'ADN ou de l'ARN qui soit un indicateur des processus biologiques normaux, des processus pathogènes ou de la réponse à des médicaments (thérapeutiques ou dopantes). Des biomarqueurs protéiques peuvent également être mis en évidence à l'aide des techniques de génomique.

On peut distinguer deux catégories de biomarqueurs : les biomarqueurs prédictifs, présents avant l'apparition des symptômes d'une pathologie, et les biomarqueurs pronostiques utilisés pour déterminer l'évolution clinique d'une maladie (probabilité de survie, de récurrence, réponse au traitement). Ils sont aujourd'hui considérés comme des outils essentiels à la gestion de nombreux risques et pathologies.

3.2.2 Biomarqueurs protéiques

La recherche de biomarqueurs protéiques permettant de détecter précocement des pathologies musculo-squelettiques est réalisée depuis de nombreuses années, en se focalisant principalement sur quelques protéines « candidates » connues. Des produits de synthèse ou de dégradation de la matrice cartilagineuse ou de l'os (propeptides du collagène II, chondroïtine sulfate, glycosaminoglycans, acide hyaluronique, ICTP, CTX-1, hydroxyproline) ont par exemple été étudiés dans le sang et le liquide synovial afin de mettre en évidence un lien avec des atteintes ostéo-articulaires. De même pour des protéines impliquées dans l'homéostasie de ces tissus (métalloprotéinases matricielles, IGF-1, ostéocalcine, phosphatase alcaline, bone sialoprotein, RANKL) ou la réponse inflammatoire (interleukine-6, Myeloperoxidase).

Les techniques de protéomique peuvent être employées afin d'identifier des biomarqueurs sanguins ou tissulaires, sans *a priori* et de façon exhaustive. Une recherche de biomarqueurs en lien avec les rhabdomyolyses d'efforts récurrents a ainsi été réalisée par 2-DE, montrant 9 protéines sur-exprimées et 4 sous-exprimées dans des biopsies de muscle de chevaux atteints de rhabdomyolyses d'effort récurrents (RER) (BOUWMAN *et al.* 2010). Des analyses par spectrométrie de masse sont également utilisées afin d'effectuer un suivi longitudinal et détecter les effets d'un éventuel dopage (BARTON *et al.* 2009). Des profils plasmatiques obtenus par spectrométrie de masse pourraient également être utilisés afin de confirmer ou préciser un diagnostic d'empoisonnement (sénéçon jacobée (MOORE *et al.* 2008)) ou d'infection (CHAUSSABEL *et al.* 2010).

3.2.3 Biomarqueurs génomiques

Les marqueurs génétiques sont évidemment des biomarqueurs ADN permettant de prédire certaines caractéristiques des animaux, comme exposé précédemment.

Les séquences nucléotidiques spécifiques de virus ou de bactéries constituent également autant de biomarqueurs génomiques utilisables à des fins diagnostiques. L'utilisation de mini-arrays ou biosensors permet de détecter par hybridation plusieurs dizaines de séquences spécifiques de pathogènes et de démontrer la présence d'une contamination ou d'une infection, même très faible.

Les profils d'expression génique obtenus par transcriptomique permettent également d'identifier des groupes de gènes présentant des signatures spécifiques d'un état physiologique. La première étude démontrant les potentialités de cette approche chez l'Homme a été publiée en 1999 (GOLUB *et al.* 1999), montrant qu'il était possible de distinguer les patients atteints de deux sous-types particuliers de lymphome. De telles études se sont multipliées, dans le but de classifier et stratifier les types de tumeurs pour mieux prédire leurs évolutions, adapter les traitements et prédire les risques de rechute. Des applications similaires sont envisageables chez le cheval afin d'identifier des signatures en lien avec la réponse à l'entraînement, la récupération après exercice, le dopage ... Une analyse transcriptomique comparée des leucocytes mononucléaires de 10 chevaux d'endurance de haut niveau avant et après la

course a ainsi mis en évidence une réponse inflammatoire aiguë durant 24 heures après l'effort (CAPOMACCIO *et al.* 2010). Des analyses complémentaires sont requises pour vérifier si ce retour rapide à l'homéostasie est une réponse adaptative à l'entraînement ou une caractéristique particulière des animaux de haut niveau.

Les miARN régulent l'expression des génomes et sont impliqués dans de nombreux processus biologiques. La mise en évidence de leur expression différentielle dans les cellules tumorales et leur spécificité tissulaire ont stimulé leur utilisation comme biomarqueurs pronostiques chez l'Homme, non seulement dans les cas de cancer mais également d'autres pathologies comme l'arthrose, le diabète ou les myopathies. De telles applications sont également envisageables chez le cheval. Une étude récente de 10 miARN a ainsi été réalisée sur des biopsies de muscle de cob normand (sains ou atteint de myopathie par rétention de polysaccharide, PSSM) et de trotteur français (sains ou atteints de RER). Elle a montré des différences d'expression entre races ainsi que des profils d'expression caractéristiques de chaque myopathie (BARREY *et al.* 2010). De façon surprenante, il a été montré dans diverses espèces dont le cheval que le sérum et le plasma contenaient de grandes quantités de miARN dérivés de différents tissus ou organes (CHEN *et al.* 2008). Les miARN pourraient ainsi constituer d'excellent biomarqueurs non invasifs.

3.2.4 Marques épigénétiques

Un projet financé par le COST en 2010 se propose d'analyser la méthylation de l'ADN du sperme d'étalon, en lien avec la capacité fécondante. En effet, diverses études suggèrent qu'une programmation épigénétique inadéquate des spermatozoïdes serait une cause majeure d'infertilité. L'analyse porte sur des gènes candidats soumis à empreinte parentale mais comporte également une analyse pangénomique par MeDIP-seq. Afin d'éliminer tout risque de contamination par des cellules somatiques, des préparations de spermatozoïdes seront obtenues à l'aide d'une nouvelle technologie de séparation des cellules (fractionnement par couplage flux force). Une collection de semences variées sera étudiée afin d'analyser les éventuelles corrélations entre niveau de méthylation et caractéristiques physiques, morphologiques et biologiques du sperme.

3.3. Les applications diagnostiques des NGS

Le séquençage individuel ouvre la voie à la médecine personnalisée, à la fois en termes de prédiction de risque, mais également en vue de développer des approches de diagnostic *in vitro* pour orienter le traitement (théranostique) ou adapter le traitement au génotype de l'individu (pharmacogénomique). En oncologie humaine par exemple, la caractérisation génomique des tumeurs contribue à définir des stratégies thérapeutiques et à mesurer leur efficacité.

Les NGS pourraient également trouver des applications dans le suivi global de l'état sanitaire des animaux. Dans le domaine des maladies infectieuses, il est possible de détecter et d'identifier rapidement des microbes responsables d'infections et d'analyser les mécanismes de résistance aux médicaments. Des approches de métagénomique, réalisées à partir d'un échantillon de fèces, pourraient être mises à profit pour caractériser le niveau et le type d'infestation par des nématodes gastrointestinaux. De même, une étude métagénomique des écosystèmes bactériens de la flore intestinale en fonction de l'âge, de l'alimentation ou de la maladie pourrait permettre d'étudier l'effet des modifications de régime alimentaire ou aboutir au développement de nouvelles approches diagnostiques, basées sur l'altération des communautés microbiennes associées à des pathologies.

Conclusion

La génomique est un domaine en plein développement, avec tout ce que cela comporte d'incertitudes, d'espoirs, voire de fantasmes, et il est important d'en mesurer les limites. Il en est ainsi des études d'association pangénomique: chez l'homme, plus de 250 loci associés à des pathologies complexes ont été identifiés, mais leurs effets phénotypiques sont faibles et rendent compte généralement d'à peine 10 % de l'héritabilité. Ces variants ont donc une valeur prédictive très limitée, contrairement aux idées véhiculées par certaines entreprises qui prétendent pouvoir prévoir notre destin à partir de notre séquence d'ADN. L'augmentation des effectifs analysés (plusieurs dizaines de milliers d'individus actuellement) permet d'identifier des variants ayant des effets toujours plus faibles, sans pour autant améliorer l'évaluation du risque. Ces résultats remettent en cause l'hypothèse admise depuis une dizaine d'année selon laquelle les maladies multigéniques fréquentes résulteraient de la conjonction de variants communs. Quelques variants fréquents ont probablement des effets notables, mais l'essentiel des effets serait dû finalement à des variants rares, des effets d'interaction ou d'autres mécanismes encore inconnus (GOLDSTEIN 2009). Or ces aspects ne sont pas étudiés actuellement, en particulier car les SNPs

présents sur les puces de génotypage ont été sélectionnés en fonction d'une fréquence minimale de l'allèle mineur de 5 %. La démocratisation et la baisse de coût du séquençage permettra de contourner cette difficulté.

De même en ce qui concerne les études de génomique fonctionnelle. Les technologies étant encore onéreuses, la plupart des études sont des études pilotes, réalisées sur des effectifs réduits. Le compromis économique se fait au détriment de la puissance de l'étude et de la pertinence globale de ses conclusions, ce qui hypothèque la plupart des applications pratiques. La recherche de biomarqueurs pronostiques ARN ou miARN chez l'homme nous montre que les résultats d'études limitées s'avèrent pour le moins optimistes, si ce n'est invalides, à l'échelle de la population générale.

Mettre au point les outils de génomique, les valider et les optimiser pour qu'ils soient économiquement viables requiert donc d'avoir accès à des populations larges et bien caractérisées. L'éclatement et la diversité des filières ainsi que le manque de phénotypes précis constituent de ce fait des freins importants au développement de ces approches chez le cheval en France. De plus, la politique d'excellence scientifique risque également de détourner les équipes de sujets appliqués difficiles à valoriser en termes de publications et dont le financement par l'Agence Nationale de la Recherche est incertain.

La situation n'est cependant pas désespérée, bien au contraire. Si chez l'homme la prédiction individuelle est l'objectif appliqué, chez les espèces d'élevage la sélection permet une amélioration génétique globale de la qualité de la population. La génomique offre dans ce domaine des perspectives séduisantes qui ont amené la filière bovine à s'engager dans la voie de l'évaluation génomique. Chez le cheval, le développement d'approches de caractérisation génomique permettrait aux professionnels de disposer d'une information de base plus riche, leur permettant d'exercer leur savoir-faire et leur expertise en toute liberté. De tels projets pourraient constituer une opportunité permettant de fédérer tous les acteurs de la filière autour de programmes communs, dans une logique gagnant-gagnant. Un phénotypage précis et diversifié (analyses morphométrique, biochimie du sang, mesure d'expression ARN et miARN plasmatique et dans les tissus d'intérêt, examens cliniques, bilan radiologique, mesure de comportement, mesure détaillée des performances...) d'une même population génotypée (> 1000 animaux) permettrait de coupler caractérisation génomique, identification et validation de biomarqueurs et recherche fondamentale dans diverses disciplines scientifiques. Les retombées appliquées pourraient consister en une évaluation plus précise de la valeur génétique des animaux, une caractérisation précoce des potentialités des poulains permettant une optimisation de la conduite des animaux, ainsi que la validation de quelques biomarqueurs pronostiques en lien avec la réponse à l'entraînement ou certaines pathologies. Encore faut-il savoir comment exploiter ces informations sur le plan zootechnique ou vétérinaire, d'où la nécessité d'un dialogue plus intense entre tous les partenaires.

La génomique permet l'acquisition de connaissances à une échelle et à une vitesse qui révolutionne les approches scientifiques classiques. Ces connaissances et les avancées technologiques concomitantes offrent de multiples perspectives tant en amélioration génétique des animaux qu'en termes de gestion des populations, de zootechnie ou de médecine vétérinaire. Mais, loin des idées naïves «du tout génomique» des années 1990, en aucun cas elle ne constitue une solution miracle. La complexité du vivant fait que nous sommes encore très loin de comprendre comment la séquence d'ADN dirige la construction et le fonctionnement d'un être vivant. En attendant, le monde du cheval bénéficie déjà, et pourrait bénéficier encore plus à terme, des avancées de la génomique et de ses technologies. L'enjeu est capital dans un contexte de compétition économique internationale.

Références

- ALEMAN, M., J. RIEHL, B. M. ALDRIDGE, R. A. LECOUEUR, J. L. STOTT *et al.*, 2004 Association of a mutation in the ryanodine receptor 1 gene with equine malignant hyperthermia. *Muscle Nerve* **30**: 356-365.
- ANDERSSON, L. S., R. JURAS, D. T. RAMSEY, J. EASON-BUTLER, S. EWART *et al.*, 2008 Equine Multiple Congenital Ocular Anomalies maps to a 4.9 megabase interval on horse chromosome 6. *BMC Genet* **9**: 88.
- BAILEY, E., R. C. REID, L. C. SKOW, K. MATHIASON, T. L. LEAR *et al.*, 1997 Linkage of the gene for equine combined immunodeficiency disease to microsatellite markers HTG8 and HTG4; synteny and FISH mapping to ECA9. *Anim Genet* **28**: 268-273.
- BARREY, E., B. BONNAMY, E. J. BARREY, X. MATA, S. CHAFFAUX *et al.*, 2010 Muscular microRNA expressions in healthy and myopathic horses suffering from polysaccharide storage myopathy or recurrent exertional rhabdomyolysis. *Equine Veterinary Journal* **42**: 303-310.
- BARTON, C., P. BECK, R. KAY, P. TEALE and J. ROBERTS, 2009 Multiplexed LC-MS/MS analysis of horse plasma proteins to study doping in sport. *Proteomics* **9**: 3058-3065.
- BELLONE, R. R., G. FORSYTH, T. LEEB, S. ARCHER, S. SIGURDSSON *et al.*, 2010 Fine-mapping and mutation analysis of TRPM1: a candidate gene for leopard complex (LP) spotting and congenital stationary night blindness in horses. *Brief Funct Genomics* **9**: 193-207.

- BERNOCO, D., and E. BAILEY, 1998 Frequency of the SCID gene among Arabian horses in the USA. *Anim Genet* **29**: 41-42.
- BOUWMAN, F. G., M. M. VAN GINNEKEN, J. P. NOBEN, E. ROYACKERS, E. DE GRAAF-ROELFSEMA *et al.*, 2010 Differential expression of equine muscle biopsy proteins during normal training and intensified training in young standardbred horses using proteomics technology. *Comp Biochem Physiol Part D Genomics Proteomics* **5**: 55-64.
- BRAULT, L. S., C. A. COOPER, T. R. FAMULA, J. D. MURRAY and M. C. PENEDO, 2010 Mapping of equine cerebellar abiotrophy to ECA2 and identification of a potential causative mutation affecting expression of MUTYH. *Genomics*.
- BROOKS, S. A., N. GABRESKI, D. MILLER, A. BRISBIN, H. E. BROWN *et al.*, 2010 Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS Genet* **6**: e1000909.
- CAPOMACCIO, S., K. CAPPELLI, E. BARREY, M. FELICETTI, M. SILVESTRELLI *et al.*, 2010 Microarray analysis after strenuous exercise in peripheral blood mononuclear cells of endurance horses. *Anim Genet* **41 Suppl 2**: 166-175.
- CHAUSSABEL, D., V. PASCUAL and J. BANCHEREAU, 2010 Assessing the human immune system through blood transcriptomics. *BMC Biol* **8**: 84.
- CHEN, X., Y. BA, L. MA, X. CAI, Y. YIN *et al.*, 2008 Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res* **18**: 997-1006.
- FINNO, C. J., S. J. SPIER and S. J. VALBERG, 2009 Equine diseases caused by known genetic mutations. *Vet J* **179**: 336-347.
- GOLDSTEIN, D. B., 2009 Common genetic variation and human traits. *N Engl J Med* **360**: 1696-1698.
- GOLUB, T. R., D. K. SLONIM, P. TAMAYO, C. HUARD, M. GAASENBEEK *et al.*, 1999 Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**: 531-537.
- GRAVES, K. T., P. J. HENNEY and R. B. ENNIS, 2009 Partial deletion of the LAMA3 gene is responsible for hereditary junctional epidermolysis bullosa in the American Saddlebred Horse. *Anim Genet* **40**: 35-41.
- GU, J., N. ORR, S. D. PARK, L. M. KATZ, G. SULIMOVA *et al.*, 2009 A genome scan for positive selection in thoroughbred horses. *PLoS One* **4**: e5767.
- HERSZBERG, B., M. E. MCCUE, T. LARCHER, X. MATA, A. VAIMAN *et al.*, 2009 A GYS1 gene mutation is highly associated with polysaccharide storage myopathy in Cob Normand draught horses. *Anim Genet* **40**: 94-96.
- HILL, E. W., J. GU, B. A. MCGIVNEY and D. E. MACHUGH, 2010 Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Anim Genet* **41 Suppl 2**: 56-63.
- LIETO, L. D., and E. G. COTHRAN, 2003 The epitheliogenesis imperfecta locus maps to equine chromosome 8 in American Saddlebred horses. *Cytogenet Genome Res* **102**: 207-210.
- MCCUE, M. E., S. J. VALBERG, M. B. MILLER, C. WADE, S. DIMAURO *et al.*, 2008 Glycogen synthase (GYS1) mutation causes a novel skeletal muscle glycogenosis. *Genomics* **91**: 458-466.
- MCGIVNEY, B. A., S. S. EIVERS, D. E. MACHUGH, J. N. MACLEOD, G. M. O'GORMAN *et al.*, 2009 Transcriptional adaptations following exercise in thoroughbred horse skeletal muscle highlights molecular mechanisms that lead to muscle hypertrophy. *BMC Genomics* **10**: 638.
- MCGIVNEY, B. A., P. A. MCGETTIGAN, J. A. BROWNE, A. C. EVANS, R. G. FONSECA *et al.*, 2010 Characterization of the equine skeletal muscle transcriptome identifies novel functional responses to exercise training. *BMC Genomics* **11**: 398.
- METALLINOS, D. L., A. T. BOWLING and J. RINE, 1998 A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung disease. *Mamm Genome* **9**: 426-431.
- MILENKOVIC, D., S. CHAFFAUX, S. TAOURIT and G. GUÉRIN, 2003 A mutation in the LAMC2 gene causes the Herlitz junctional epidermolysis bullosa (H-JEB) in two French draft horse breeds. *Genet Sel Evol* **35**: 249-256.
- MOORE, R. E., D. KNOTTENBELT, J. B. MATTHEWS, R. J. BEYNON and P. D. WHITFIELD, 2008 Biomarkers for ragwort poisoning in horses: identification of protein targets. *BMC Vet Res* **4**: 30.
- ROSENGREN PIELBERG, G., A. GOLOVKO, E. SUNDSTRÖM, I. CURIK, J. LENNARTSSON *et al.*, 2008 A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nat Genet* **40**: 1004-1009.
- RUDOLPH, J. A., S. J. SPIER, G. BYRNS and E. P. HOFFMAN, 1992a Linkage of hyperkalaemic periodic paralysis in quarter horses to the horse adult skeletal muscle sodium channel gene. *Anim Genet* **23**: 241-250.
- RUDOLPH, J. A., S. J. SPIER, G. BYRNS, C. V. ROJAS, D. BERNOCO *et al.*, 1992b Periodic paralysis in quarter horses: a sodium channel mutation disseminated by selective breeding. *Nat Genet* **2**: 144-147.
- SANTSCHI, E. M., A. K. PURDY, S. J. VALBERG, P. D. VROTSOS, H. KAESE *et al.*, 1998 Endothelin receptor B polymorphism associated with lethal white foal syndrome in horses. *Mamm Genome* **9**: 306-309.
- SHIN, E. K., L. E. PERRYMAN and K. MEEK, 1997 A kinase-negative mutation of DNA-PK(CS) in equine SCID results in defective coding and signal joint formation. *J Immunol* **158**: 3565-3569.
- SPIRITO, F., A. CHARLESWORTH, K. LINDER, J. P. ORTONNE, J. BAIRD *et al.*, 2002 Animal models for skin blistering conditions: absence of laminin 5 causes hereditary junctional mechanobullous disease in the Belgian horse. *J Invest Dermatol* **119**: 684-691.
- TRYON, R. C., S. D. WHITE and D. L. BANNASCH, 2007 Homozygosity mapping approach identifies a missense mutation in equine cyclophilin B (PPIB) associated with HERDA in the American Quarter Horse. *Genomics* **90**: 93-102.
- VALBERG, S. J., T. L. WARD, B. RUSH, H. KINDE, H. HIRARAGI *et al.*, 2001 Glycogen branching enzyme deficiency in quarter horse foals. *J Vet Intern Med* **15**: 572-580.
- WARD, T. L., S. J. VALBERG, D. L. ADELSON, C. A. ABBEY, M. M. BINNS *et al.*, 2004 Glycogen branching enzyme (GBE1) mutation causing equine glycogen storage disease IV. *Mamm Genome* **15**: 570-577.
- YANG, G. C., D. CROAKER, A. L. ZHANG, P. MANGLICK, T. CARTMILL *et al.*, 1998 A dinucleotide mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome (LWFS); a horse variant of Hirschsprung disease. *Hum Mol Genet* **7**: 1047-1052.