

Intérêt du séquençage de génome complet pour la détection précoce des anomalies génétiques

MICHOT P. (1,2), CHAHORY S. (3), CHAMBRIAL M. (4), BARBEY S. (5), GROHS C. (2), DELOCHE M.C. (1), DANCHIN-BURGE C. (6), FRITZ S. (1), BOICHARD D. (2), CAPITAN A. (1,2)

(1) Alice, 149 rue de Bercy, 75595 Paris Cedex 12, France

(2) INRA, UMR 1313 Génétique Animale et Biologie Intégrative, 78352 Jouy-en-Josas, France,

(3) Université Paris-Est, Ecole Nationale Vétérinaire d'Alfort, Unité d'Ophtalmologie, Maisons-Alfort Cedex, France,

(4) ORIGENPLUS, 38 rue de la Méroillère, 61300 L'AIGLE

(5) INRA, UE 326 Domaine Expérimental du Pin. Le-Pin-Au-Haras, France

(6) IDELE, 149 rue de Bercy, 75595 Paris Cedex 12, France

RESUME - La disponibilité de la séquence du génome complet de nombreux individus, rendue possible par l'évolution très rapide des technologies de séquençage d'ADN, constitue un apport majeur pour l'identification des anomalies génétiques. L'approche conventionnelle par phénotypage des individus, cartographie de l'anomalie, et identification de la mutation, gagne en rapidité et puissance lorsque l'on dispose d'une grande base de référence pour filtrer les variants candidats. Mais les données de séquence permettent également d'envisager d'autres approches, en particulier pour les anomalies dont le phénotype est difficilement identifiable et qui passent encore inaperçues. Par exemple, la cartographie des déficits en homozygotes à partir des données de génotypage, suivie de l'analyse des données de séquence d'individus porteurs, a permis l'identification de plusieurs mutations responsables de mortalité embryonnaire dans différentes races. De façon plus prospective, on peut identifier des polymorphismes délétères dans les données de séquence à partir de leur annotation fonctionnelle, puis caractériser leur effet potentiel par le phénotypage d'animaux porteurs. Cette approche est illustrée avec une anomalie de la vision en race Normande. L'augmentation rapide du nombre d'anomalies génétiques résolues ces dernières années démontre la puissance apportée par le séquençage de génomes complets et ce phénomène s'amplifiera à l'avenir. L'approche comporte cependant des limites, l'annotation encore très incomplète des polymorphismes du génome ou la caractérisation phénotypique, souvent compliquée mais essentielle pour la compréhension des effets des mutations. Enfin, le nombre croissant d'anomalies découvertes dans chaque race nécessitera des stratégies de contre-sélection plus complexes que l'éradication directe.

Value of whole genome sequencing for early detection of genetic defects

MICHOT P. (1,2), CHAHORY S. (3), CHAMBRIAL M. (4), BARBEY S. (5), GROHS C. (2), DELOCHE M.C. (1), DANCHIN-BURGE C. (6), FRITZ S. (1), BOICHARD D. (2), CAPITAN A. (1,2)

(1) Alice, 149 rue de Bercy, 75595 Paris Cedex 12, France ; (2) INRA, UMR 1313 Génétique Animale et Biologie Intégrative, 78352 Jouy-en-Josas, France

SUMMARY - The availability of whole genome sequences (WGS) of many individuals, made possible by the rapid evolution of DNA sequencing technologies, is a major factor for genetic defect identification. The conventional approach consisting in phenotyping affected animals, mapping and identification of the causal mutation gains speed and power with the availability of large sequence databases to filter out candidate variants. WGS also open new avenues, especially for defects with non specific phenotypes which may remain unobserved. Mapping of regions showing homozygous deficiency with massive genotyping data, followed by analysis of the genome sequences of carriers, enables the identification of several causal mutations responsible for embryonic lethality in different breeds. In a more prospective way, one can identify deleterious polymorphisms in WGS on the basis of their functional annotation, and then characterize their potential effect on carrier animals. Here this approach is illustrated by a defect affecting vision in the Normandy cattle breed. The rapid increase in the number of resolved genetic defects during the last years illustrates the power brought by WGS and this will continue in the future. These approaches still suffer from important limitations, especially the incomplete annotation of the biological effects of variants on the genome, or the phenotypic characterization of the mutations, often complex but necessary for a full comprehension of their effects. Finally, the increasing number of known defects in each breed requires new selection strategies, more complex than a direct elimination.

INTRODUCTION

La source initiale de variabilité génétique est la mutation de l'ADN. Le taux de mutation en conditions naturelles est de l'ordre de 10^{-8} par base, ce qui signifie que lors de la production d'un gamète, quelques dizaines de néo-mutations sont créées. La grande majorité d'entre elles n'ont pas d'effet, certaines ont des effets bénéfiques et certaines ont des effets délétères. La plupart restent à fréquence très faible, voire sont perdues après quelques générations. Nous ne parlerons ici que des mutations à effet très délétère, qu'on appelle anomalies génétiques. Si l'effet délétère est dominant, la contre-sélection est rapide de sorte que ces mutations disparaissent souvent d'elles-mêmes. Si l'effet délétère est récessif, l'allèle muté, présent majoritairement à l'état hétérozygote, se comporte comme un variant neutre et la mutation n'est pas contre-sélectionnée. En conséquence, toute population porte des

anomalies génétiques. Plusieurs milliers sont décrites chez l'homme, plusieurs centaines chez le bovin, la différence n'étant due qu'à l'intensité variable des observations. Tout individu porte statistiquement, à l'état hétérozygote, plusieurs mutations récessives. Ce phénomène naturel et rare est amplifié chez le bovin du fait que les races sont des populations génétiquement petites. Par exemple, l'effectif génétique des races bovines laitières est de l'ordre de 50. Cette taille génétique réduite a deux conséquences : une augmentation inéluctable de la consanguinité et donc de l'homozygotie favorisant l'apparition d'homozygotes affectés (et donc révélant les anomalies récessives) ; d'autre part, une dérive génétique, c'est-à-dire une fluctuation aléatoire des fréquences alléliques d'une génération à l'autre imputable à l'échantillonnage réduit des reproducteurs. La fréquence d'une anomalie peut ainsi fortement augmenter d'une valeur initiale proche de 0 jusqu'à plusieurs pourcents du fait de la diffusion

de reproducteurs porteurs très influents et de leurs descendants. La politique usuelle est de réduire la fréquence, voire d'éliminer les anomalies, en raisonnant bien la politique à appliquer car elle peut être coûteuse et contreproductive. L'outil indispensable est la connaissance du génotype des reproducteurs qui permet de distinguer les homozygotes normaux des hétérozygotes porteurs. Cet article présente différentes approches pour identifier les mutations responsables d'anomalies génétiques en montrant l'apport majeur des nouvelles technologies de séquençage du génome.

1. LES DONNEES DE SEQUENCE

Les technologies de séquençage évoluent très rapidement. On pourra se référer à Nielsen et al (2011) pour une description détaillée. Schématiquement, les méthodes actuelles et usuelles reposent sur le découpage de la molécule d'ADN en très nombreux petits fragments, chaque fragment étant séquencé sur quelques centaines de bases dans des dispositifs miniaturisés. Chaque petite séquence est ensuite « alignée », c'est-à-dire replacée sur la séquence de référence de l'espèce (supposée disponible, donc préalablement construite). Chaque point du génome étant séquencé statistiquement plusieurs fois, les segments sont généralement chevauchants et assurent une couverture dense du génome. Une fois tous les segments alignés, on récupère la séquence (tout au moins une grande partie) de l'individu analysé. Les différences entre la séquence de l'individu analysé et la séquence de référence peuvent avoir deux origines : des erreurs de séquençage qu'il faut éliminer, ou des différences vraies de séquences qu'il convient d'identifier. Le génotype de l'individu analysé est déduit des allèles observés et du nombre de leurs copies. Supposons, par exemple, qu'à une base donnée, la référence indique l'allèle A. Si cette base est séquencée 20 fois chez l'individu analysé et qu'on observe 9 fois l'allèle A et 11 fois l'allèle G, on déduira, avec une très grande fiabilité, que l'individu est hétérozygote A/G. On peut mettre ainsi en évidence des substitutions (SNP) ou de petites insertions ou délétions (InDel). La séquence est donc équivalente à un génotypage exhaustif (en théorie) de l'individu. Les polymorphismes structuraux, comme les grandes insertions/délétions ou les inversions, demandent d'autres méthodes pour être mis en évidence. Chez un bovin typique, on observe généralement 3 à 6 millions de polymorphismes fiables pour une vingtaine de millions observés à l'échelle d'une race donnée et près de 40 millions pour l'espèce (Daetwyler et al, 2014). Afin d'établir des relations avec un phénotype, les polymorphismes sont annotés. C'est-à-dire qu'on leur donne un sens fonctionnel compte tenu des connaissances actuelles du génome. Ces informations reposant principalement sur la position des gènes et la structure de leur(s) transcrit(s), sont accessibles dans des bases de données publiques telles que *RefSeq*, *UCSC* ou *Ensembl* et forment la base principale de l'annotation. Par superposition entre ces bases et la position des variants identifiés, on leur attribue une classe fonctionnelle définie par leur localisation par rapport aux gènes (exon, intron, partie en amont ou aval du gène, intergénique) et une prédiction de l'effet sur la protéine traduite (par exemple, changement ou non d'acide aminé, apparition ou suppression d'un codon stop, séquence déterminant un épissage alternatif, décalage du cadre de lecture) (McCarthy et al, 2014). Ainsi, sur les millions de polymorphismes identifiés chez le bovin, environ 66% sont dans une région intergénique (et dans ce cas, l'annotation est mal connue), 26% dans un intron et seulement 1% se situent dans la séquence codante. Moins de 0.5 % sont dans des classes fonctionnelles considérées comme délétères, ce qui représente tout de même quelques centaines de milliers de variants (www.1000bullgenomes.com). Les annotations jouent un rôle essentiel dans le filtrage des variants, toujours trop nombreux après la phase de séquençage. On fait en effet l'hypothèse que puisqu'une anomalie a un effet fort, la

mutation responsable doit avoir une prédiction d'effet biologique marqué. Les évolutions dans le domaine du séquençage sont très rapides et les principaux changements pour l'utilisateur sont le prix (en diminution) et le débit (en augmentation), sans oublier la qualité (le taux d'erreur) et la longueur des lectures. Dans les centres les plus performants, le prix marginal d'une séquence de mammifère de couverture 10 à 15 x (c.à.d. pour laquelle chaque base du génome est lue en moyenne 10 à 15 fois) est de l'ordre de 1000€. On comprend ainsi qu'il devient souvent préférable de séquencer un génome complet plutôt que de passer des mois à caractériser une région particulière du génome.

2. DE L'ANOMALIE A LA MUTATION CAUSALE

2.1 L'APPORT DE LA SEQUENCE DANS L'APPROCHE CONVENTIONNELLE

La première étape de l'étude d'une anomalie génétique est de repérer son émergence dans la population, d'identifier les individus atteints et d'organiser leur phénotypage. En Europe, plusieurs pays se sont dotés d'observatoires dont le but est d'assurer l'épidémiologie des anomalies génétiques. Les réseaux mis en place facilitent les échanges entre les professionnels au contact des animaux (éleveurs, techniciens, vétérinaires) et les équipes de recherche. En France, l'Observatoire National des Anomalies Bovines (ONAB, www.onab.fr) assure cette veille par le recensement des cas observés. En plus des phénotypes, il assure aussi la collecte d'information de pedigree et de matériel biologique. L'étude du phénotype et du pedigree des individus atteints permet en général de confirmer un déterminisme génétique et d'émettre une hypothèse sur le mode de transmission. L'étape suivante consiste à cartographier l'anomalie sur le génome. Le développement de puces permettant le génotypage de milliers de marqueurs a considérablement facilité la recherche d'association entre marqueurs et phénotype. Dans le cas d'une anomalie récessive, la méthode couramment utilisée est la cartographie par homozygotie dont le principe est le suivant. Lorsqu'une mutation récessive, résultant d'un événement de mutation unique, apparaît chez un individu, elle est transmise à sa descendance, ainsi qu'aux générations suivantes, avec une partie du chromosome qui l'entoure. Les porteurs de la mutation partagent donc un fragment identique de chromosome, d'autant plus petit que le nombre de recombinaisons s'est accumulé, donc que ces porteurs sont nombreux et que le nombre de générations les séparant de l'ancêtre est important. Un individu atteint, homozygote pour la mutation, est donc aussi homozygote pour un segment qui l'entoure. Ce segment peut être identifié par les marqueurs moléculaires, comme homozygotes et identiques chez tous les atteints. Les pedigrees usuels, constitués de quelques cas, permettent une localisation de la région associée à l'anomalie dans quelques mégabases (Charlier et al, 2008). La dernière étape est celle de la recherche de la mutation causale. En l'absence de données de séquence de génome complet, on cherche à identifier un gène candidat parmi les quelques dizaines ou centaines de gènes de l'intervalle étudié, en s'appuyant sur la comparaison du phénotype observé avec les données d'annotation disponibles chez l'homme et la souris. Le gène est ensuite séquencé pour rechercher les polymorphismes délétères et identifier celui à l'origine du phénotype. Cette recherche est longue et peut ne pas aboutir, en particulier lorsqu'aucun gène évident n'est présent dans la région étudiée. Aujourd'hui, il est possible de séquencer l'intégralité de la région étudiée ou même l'ensemble du génome de quelques individus atteints. La recherche de la mutation candidate gagne en puissance et en rapidité en comparant directement les polymorphismes présents chez les individus atteints à ceux présents chez des individus que l'on sait non-porteurs. Lorsque l'on dispose d'une grande base de référence comme le projet « 1000 génomes bovins », ce filtre est particulièrement puissant, comme l'illustre le grand nombre de mutations identifiées ces dernières années chez le bovin.

2.2 EXEMPLE DE L'ÉPIDERMOLYSE BULLEUSE JONCTIONNELLE

L'exemple présenté ci-dessous décrit comment à partir de l'ADN d'un seul animal atteint et sans cartographie génétique préalable, il est dorénavant possible d'identifier des mutations récessives responsables de syndromes très rares tels l'épidermolyse bulleuse jonctionnelle (EBJ) en race Charolaise (Michot et al, 2015). Quatre veaux charolais issus d'accouplements consanguins et présentant une anomalie congénitale de la peau ont été signalés dans le même élevage. Les analyses cliniques et histopathologiques ont révélé un syndrome d'EBJ. L'ADN d'un seul cas étant disponible au moment de l'étude, la recherche de régions homozygotes a été effectuée directement à partir des données de séquençage du génome. 54 régions homozygotes de taille supérieure à une mégabase ont été identifiées. Les variants présents dans ces régions ont ensuite été filtrés à l'aide des données de séquence de 8 Charolais sains et de 234 individus d'autres races. Aucun SNP ou petit InDel candidat n'ayant été identifié, nous avons procédé à une recherche de variants structuraux. Ainsi, nous avons mis en évidence une grande délétion affectant les exons 17 à 23 du gène codant pour l'intégrine beta 4, une protéine essentielle au maintien de l'intégrité structurale de la peau et dont la perte de fonction entraîne un syndrome similaire chez l'homme et de nombreuses espèces. Le génotypage des parents et d'un autre cas né au cours de l'étude a permis de confirmer la parfaite association entre cette mutation candidate et les génotypes attendus.

2.3 LIMITES DE L'APPROCHE

Avec le développement du génotypage et du séquençage de génome complet, l'identification de la mutation responsable d'une anomalie est plus facile et rapide. Cependant, la stratégie repose sur l'observation et la description d'un phénotype anormal. Pour les anomalies récessives, cela suppose l'existence d'individus homozygotes dans la population. Par conséquent, l'allèle délétère a déjà diffusé suffisamment dans la population pour permettre la naissance des premiers veaux homozygotes à partir de leurs deux parents porteurs, soit généralement 4 à 6 générations au minimum après l'évènement de mutation originel. Même avec un bon test de diagnostic, il faut un effort d'éradication d'autant plus important que l'allèle délétère a plus diffusé. D'autre part, cela nécessite que l'anomalie présente des caractéristiques cliniques bien identifiables pour être rapportée. Certaines anomalies passent souvent inaperçues comme les anomalies responsables de mortalité embryonnaire, ou de pathologies peu spécifiques (mortalités dues à une plus forte sensibilité aux infections, par exemple) ou de troubles apparaissant tardivement dans la vie ou impactant peu la carrière de l'animal. Ces anomalies ont donc peu de chances d'être découvertes avec cette approche.

3. DU GENOTYPE AU PHENOTYPE

3.1 DEFICIT EN HOMOZYGOTES

Le même principe de cartographie par homozygotie peut être utilisé à partir des bases de données de génotypage de sélection génomique. Supposons qu'une anomalie récessive ségrège dans la population. Elle est entourée d'un segment chromosomique caractérisé par un haplotype (c'est-à-dire une succession précise d'allèles à différents marqueurs adjacents). Si cet haplotype est assez fréquent à l'état hétérozygote dans la population mais absent à l'état homozygote, cela signifie que les animaux correspondants n'ont jamais été candidats à la sélection, soit parce qu'ils n'existent pas du fait d'une mortalité au cours de la gestation, au vêlage ou peu après, soit parce qu'ils présentent des caractéristiques éliminatoires pour la sélection. Cette approche, initialement proposée par VanRaden et al. (2011) a été reprise par Fritz et al. (2013) qui ont analysé 76000 animaux typés avec la puce 50k. Une dizaine de régions ont été mises en évidence dans chaque race (Holstein, Normande, Montbéliarde). Pour chaque région, l'étude des haplotypes a permis de définir le statut (porteur/non

porteur) des taureaux d'insémination et de détecter les accouplements à risque (c'est-à-dire taureau porteur sur fille de porteur) dans la base nationale. Dans 9 cas, ces accouplements ont montré une baisse de fertilité compatible avec une mortalité des homozygotes durant la gestation. Ainsi les locus HH1, HH3, CVM, Brachyspina déjà connus ont été retrouvés et quatre nouveaux loci (HH4, MH1, MH2 et NH1) ont été décrits. L'analyse de la séquence de taureaux porteurs et leur comparaison avec le reste de la base « 1000 génomes bovins », a ensuite permis l'identification des mutations causales pour les loci HH3, HH4 et MH2. L'annotation fonctionnelle de ces variants (prédits comme très délétères), l'absence d'homozygotes à ces mutations dans un large panel d'animaux génotypés pour ces mutations, ainsi que le très haut niveau de conservation des séquences entre espèces ont permis de supposer que ces mutations étaient causales, même en l'absence de validation fonctionnelle stricte. A nouveau, on montre ici la puissance des données de séquences, en utilisant à la fois les porteurs et les nombreux contrôles. D'autres analyses sont en cours afin de caractériser les mutations responsables de déficits en homozygotes après la naissance.

3.2 LA DETECTION DES ANOMALIES GENETIQUES PAR UNE ANALYSE DES DONNEES DE SEQUENCE ET UNE APPROCHE DE GENETIQUE INVERSE :

3.2.1 Le principe général

Avec l'accès aux données de séquences de génomes, des approches dites de « génétique inverse » se sont développées (McArthur et al. 2012; Sulem et al. 2015). Le principe est de sélectionner des polymorphismes candidats sur la séquence, puis de rechercher leur effet éventuel sur le phénotype des individus porteurs. Cette stratégie se base sur l'exploitation de l'annotation des données de séquences et la prédiction de l'effet des variants observés. On s'intéresse aux polymorphismes très délétères qui modifient la structure et le fonctionnement de la protéine et à qui on attribue un fort potentiel pathologique. Ce sont principalement le gain d'un codon stop, le décalage du cadre de lecture, la modification des sites d'épissage mais également le changement d'acide aminé lorsque les propriétés sont très différentes ou qu'il impacte un domaine fonctionnel de la protéine (comme un site de liaison par exemple). D'autres données d'annotation comme la fonction du gène ou l'association à des maladies génétiques dans diverses espèces (OMIM, OMIA, MGI mouse) sont prises en compte et donnent une première appréciation du phénotype potentiellement attendu. La phase de vérification du phénotype nécessite des individus homozygotes aux variants sélectionnés, ainsi que des autres génotypes comme contrôles. Ils peuvent être recherchés directement par un dépistage dans la population à l'aide d'un test sur la mutation causale. Souvent, cette méthode est peu efficace si les individus sont rares ou éliminés rapidement. On cherche alors des parents porteurs (génotypés sur la mutation causale ou prédits comme porteurs à partir de marqueurs liés) et on identifie les accouplements à risque. Ceci permet de planifier l'étude à partir des veaux à naître, en particulier leur génotypage et surtout leur phénotypage qui est en partie guidé par la fonction du gène, les voies métaboliques auxquelles il participe et par les phénotypes observés chez d'autres espèces (le phénotype d'un KO du gène chez la souris par exemple). Il est important de noter que la plupart des annotations délétères ne se traduisent pas par des anomalies, du fait de mécanismes compensateurs fréquents au niveau du génome. Le rendement de cette approche est donc assez faible.

3.2.2 Applications pour la détection d'anomalies génétiques chez le bovin

Les populations bovines se prêtent bien à l'application de ce type de stratégie. De par leur structure (grand nombre d'individus vs taille génétique faible) il est possible de capter dans les données de séquençage la majorité des anomalies

ségrégant dans une population, même à faible fréquence. Par le biais de la sélection génomique, un grand nombre de phénotypes et génotypes sont collectés en routine et accessibles, ce qui facilite l'étude de l'effet des variants et l'identification des individus porteurs. Une étude belge récente (Charlier et al., 2014) de 331 variants « perte de fonction » en race Blanc Bleu Belge et 43 variants en race Holstein a identifié un déficit en homozygotes pour 112 et 11 d'entre eux, soit environ un tiers. Pour trois de ces variants, un phénotype de mortalité embryonnaire et juvénile a été confirmé par le suivi des accouplements à risque. Un programme similaire est en cours à l'INRA dans le cadre du projet BOVANO. Environ 1000 SNP et petits InDel ont été sélectionnés à partir des données de séquence de génomes entières générées dans le projet « 1000 bull genomes » et de différents projets de recherches conduits à l'INRA. Ces SNP ont été intégrés à la puce de génotypage EuroG10K utilisée en routine dans le cadre de la sélection génomique, permettant ainsi la collecte des génotypes de très nombreux individus. C'est dans ce cadre qu'un variant a été associé à une anomalie de la vision en race Normande. Le polymorphisme identifié est une insertion d'une base dans la séquence du gène *Retinitis Pigmentosa-1 (RP1)*, situé sur le chromosome 14, ayant pour conséquence l'apparition d'un décalage du cadre de lecture et la terminaison prématurée de la protéine. De nombreuses mutations dans *RP1* sont associées aux formes dominantes et récessives de la rétinite pigmentaire chez l'homme. Cette dégénérescence héréditaire de la rétine se caractérise par une perte progressive de la vision, nocturne puis diurne, causée par la désorganisation et la mort progressive des photorécepteurs (Pierce et al., 1999). Le génotypage a révélé une fréquence de l'allèle délétère particulièrement élevée en race Normande (27%, correspondant à 7% d'homozygotes). Par ailleurs, dans cette même race, un phénotype de perte de vision s'aggravant avec l'âge très similaire à celui observé chez l'homme et qui pourrait être expliqué par ce variant a été rapporté à l'ONAB ces dernières années. Pour le confirmer, 39 vaches Normandes des trois génotypes et d'âge compris entre 1 et 9 ans ont été recrutées pour phénotypage. Les examens oculaires (réponse à la menace, réflex pupillaire direct et indirect, examen du segment antérieur et du fond d'œil après dilatation de la pupille) ont révélé chez les individus homozygotes âgés de plus de 5 ans (5/8) une dégénérescence bilatérale de la rétine accompagnée d'un déficit visuel marqué. Les 3 plus jeunes ne semblaient pas encore présenter de symptôme, allant dans le sens d'une apparition progressive de la maladie. En complément, l'analyse histologique des rétines de 2 vaches homozygotes, réformées après le phénotypage, montre une disparition des segments externes des photorécepteurs. L'ensemble de ces résultats confirme l'effet délétère de la mutation à l'état homozygote. Un suivi clinique des individus plus jeunes est envisagé pour estimer précisément l'âge d'apparition des symptômes. Peu de cas de dégénérescence héréditaire de la rétine ont été décrits chez le bovin et l'origine génétique est souvent difficile à démontrer du fait de la complexité du phénotypage et des possibilités de confusion avec des causes environnementales. Malgré son apparition tardive et son caractère moins débilitant pour la production que la plupart des anomalies génétiques, les conséquences de cette anomalie sont loin d'être négligeables. En effet, elle impacte sensiblement le bien-être des animaux atteints et touche chaque élevage avec un homozygote pour 13 vaches. Grâce à sa puissance, l'approche de génétique inverse nous a permis de découvrir la partie immergée de cet iceberg et elle se révélera sans conteste un atout majeur pour l'identification de nouvelles anomalies du même type.

CONCLUSION

Ces quelques exemples montrent la puissance apportée par les données de séquence de génome complet, tant pour les animaux atteints que pour les animaux contrôles, dont le nombre doit être relativement élevé pour garantir une bonne

efficacité du filtrage des variants. De ce nombre dépend également la précision d'imputation des variants chez les animaux génotypés et non séquencés. Les anomalies étant relativement rares, l'imputation n'est précise que si le nombre d'animaux séquencés est élevé. Or l'imputation est une phase essentielle pour l'analyse de grandes populations que l'on ne peut pas séquencer ni régénérer. L'annotation est également un facteur essentiel. Avec les méthodes actuelles, elle est très informative pour les variants dans les séquences codantes pour des protéines, mais elle est encore partielle dans les séquences non codantes des gènes et très réduite dans les séquences intergéniques qui peuvent pourtant avoir des fonctions régulatrices. Il est donc probable que l'on rate encore de nombreux variants d'intérêt, faute de savoir les interpréter au niveau fonctionnel. Alors que la génomique progresse très rapidement, la phase la plus limitante reste et restera la caractérisation phénotypique, qu'elle soit lors des émergences ou guidée par les génotypages. Les progrès de la génomique ne remplacent pas le phénotypage, ils le complètent et la plus-value dépend surtout du rapprochement des informations. Les nouvelles approches vont permettre de détecter un nombre croissant d'anomalies. Les résultats obtenus les 3-4 dernières années sont déjà plus nombreux que tous ceux cumulés auparavant et le phénomène va s'amplifier. Il concernera aussi les QTL, c'est-à-dire des variants à effets plus réduits. Un nombre grandissant de gènes individuels devra être intégré dans la gestion des populations qui devra être plus réfléchie que par le passé : tous les reproducteurs risquent d'être porteurs de l'une ou l'autre des anomalies décrites, de sorte qu'une élimination brutale des porteurs ne sera plus possible. Il faudra inclure ces informations dans un index de synthèse pour viser une réduction progressive de la fréquence de porteurs, du moins chez les mâles.

Les auteurs remercient le GAEC de l'Araucaria pour son accueil, M. Philippe et EVOLUTION pour les échantillons transmis, A. Boukadiri (INRA GABI) pour son assistance technique dans la réalisation des analyses histologiques, ainsi que les partenaires du consortium "1000 génomes bovins". Le projet BOVANO (ANR-14-CE19-0011) est financé par l'ANR et Apisgene.

- Charlier C., Coppeters W., Rollin F. et al., 2008.** Nat genet, 40, 449-453
- Charlier C., Li W., Harland C. et al, 2014.** Proc10th WCGALP.
- Daetwyler H., Capitan A., Pausch H., et al, 2014.** Nat Genet, 46, 858-867
- Fritz S., Capitan A., Djari A. et al., 2013.** PLoS One, 8, e65550.
- MacArthur DG., Balasubramanian S., Frankish A. et al., 2012.** Science, 335, 823-828.
- McCarthy DJ., Humburg P., Kanapin A., et al., 2014.** Genome Med, 6(3):26.
- Michot P., Fantini O., Braques R. et al., 2015.** Genet Sel Evol, 47:37.
- Nielsen R., Paul J.S., Albrechtsen A. and Song Y. S. 2011.** Nat Rev, 12, 443-451
- Pierce EA., Quinn T., Meehan T. et al. 1999.** Nat Genet, 22,248-254.
- Sulem P. Helgason H., Oddson A., et al. 2015.** Nat Genet, 47, 548-552.
- VanRaden PM., Olson KM., Null DJ. and Hutchison JL., 2011.** J Dairy Sci, 94, 6153-61.