

Du phénotype à la mutation causale : le cas des anomalies récessives bovines

A. DUCHESNE¹, C. GROHS¹, P. MICHOT^{1,2}, M. BERTAUD¹, D. BOICHARD¹, S. FLORIOT¹, A. CAPITAN^{1,2}

¹ GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350, Jouy-en-Josas, France

² Alice, Maison Nationale des Eleveurs, 75595, Paris, France

Courriel : amandine.duchesne@inra.fr

La prise en compte des anomalies dans la gestion des populations passe par la connaissance du génotype des reproducteurs qui nécessite un test ADN sur la mutation causale ou sur des marqueurs associés. La connaissance de la mutation causale permet également de mieux comprendre la fonction du gène impliqué ainsi que le mécanisme de l'anomalie. Les anomalies ont donc fait l'objet d'études importantes pour identifier les mutations responsables. Cet article décrit les approches utilisées dans ces études, pour les anomalies récessives chez les bovins.

Comme indiqué par Boichard *et al* (2016 ce numéro), une anomalie génétique est un phénotype déviant par rapport la population, le plus souvent délétère, rare et dont le déterminisme est dû principalement ou exclusivement à un seul facteur génétique. Dans le cadre d'anomalies récessives, le phénotype fournit une prédiction très partielle du génotype des individus. Les atteints portent l'allèle muté à l'état homozygote. On peut aussi en déduire que leurs parents phénotypiquement normaux sont hétérozygotes et parfois déduire que les individus de la lignée entre un atteint et l'ancêtre muté sont tous hétérozygotes. Toutefois, ces déductions restent très partielles et, dans la majorité des cas, on ne peut pas garantir qu'un individu soit homozygote non porteur ou hétérozygote porteur.

La seule possibilité repose sur l'analyse directe de l'ADN de l'individu. Les premiers outils techniques sont apparus au début des années 1990 mais ce n'est que vers la fin des années 2000 que ces outils de génotypage et de séquençage sont devenus très performants, ouvrant la possibilité d'identifier la mutation causale en quelques mois dans les cas les plus favorables. Le procédé le plus classique est constitué des étapes décrites dans cet article. Nous présentons la situation chez les bovins, mais la même évolution est observée dans de nombreuses espèces, dès lors que l'on peut disposer de données phénotypiques et des outils génomiques adaptés, en particulier une bonne séquence de référence du génome et d'outils de génotypage de nombreux marqueurs.

1 / Constitution d'une collection de cas et de témoins

Une analyse génétique repose sur une collection de cas. C'est l'objectif de l'Observatoire National des Anomalies Bovines (ONAB), décrit dans l'article précédent (Grohs *et al* 2016 ce numéro). Les cas n'ont pas besoin d'être très nombreux, quelques dizaines suffisent souvent et même parfois quelques unités. Par contre, il est essentiel que le diagnostic soit précis et que la maladie corresponde bien à l'anomalie décrite, car les méthodes utilisées supposent une cause génétique unique. Une hétérogénéité de déterminisme génétique, due par exemple à deux syndromes ressemblants mais différents, risque de conduire à l'échec de la recherche. Pour garantir cette homogénéité, on ne retient généralement que des individus atteints de la même race et dont les parents présentent tous un ancêtre commun, supposé être le fondateur ayant transmis l'anomalie. Le diagnostic repose alors sur la connaissance précise du pedigree sur un nombre suffisant de générations (5 à 15). Différents outils d'analyse de pedigree, comme Pedig (Boichard 2002) peuvent être mobilisés pour trouver l'ancêtre porteur et construire cette collection d'individus. Par exemple, le programme « anc_comm » de Pedig propose une liste d'ancêtres possibles et calcule leur contribution génétique à la collection de cas, en comparaison du reste de la population. La constitution de cette collection bien caractérisée maximise la probabilité de succès de l'étude et, de ce fait, doit être privilégiée chaque fois que c'est possible. L'autre

façon d'augmenter l'homogénéité de la collection est un diagnostic clinique plus précis, pratiqué par une personne compétente. Dans ce cadre, l'investissement des vétérinaires (écoles et vétérinaires libéraux) dans le phénotypage fin des animaux atteints est un atout pour l'ONAB, déterminant dans cette étape.

Bien entendu, l'analyse du génome nécessite un échantillon d'ADN des cas et d'un certain nombre de témoins non atteints. C'est un autre objectif de l'ONAB de collecter et centraliser des échantillons de tissu (sang, biopsie d'oreille, ou tout autre tissu nucléé) pour les rendre disponibles pour les analyses génétiques et génomiques.

Quant aux animaux utilisés comme « contrôles », il s'agit d'animaux représentatifs de la variabilité de la race et on utilise généralement ceux pour lesquels un génotype est disponible. Bien que certains soient susceptibles de porter la mutation délétère recherchée à l'état hétérozygote, leurs génotypes seront pour la plupart différents de ceux des animaux atteints. La puissance de l'analyse repose sur le nombre de ces « contrôles », et la collection de génotypes conservés dans les bases de données de l'INRA, en particulier dans l'espèce bovine, est à cet égard particulièrement précieuse.

2 / Cartographie de l'anomalie sur le génome

Les approches génétiques reposent sur l'utilisation de marqueurs génétiques. Un marqueur génétique correspond à une

position sur le génome présentant un polymorphisme, donc deux allèles au minimum, que l'on sait mettre en évidence. Un bon marqueur génétique a les propriétés suivantes : *i*) il est suffisamment polymorphe pour que l'on puisse facilement suivre les transmissions alléliques entre parent et produit ; *ii*) il est codominant, c'est-à-dire qu'on distingue bien les différents allèles ; *iii*) il est facile à typer, avec un taux d'erreur et un taux de génotypage indéterminé faibles, et un coût de génotypage bas ; *iv*) les marqueurs sont suffisamment nombreux pour couvrir la totalité du génome avec une densité suffisante. Entre le début des années 1990 et le milieu des années 2000, les marqueurs utilisés ont été majoritairement des microsatellites, très informatifs car polyalléliques, mais relativement coûteux à typer (Boichard *et al* 1998). Ils ont été supplantés par les marqueurs de type « *Single Nucleotide Polymorphism* » (SNP), moins informatifs car bialléliques, mais beaucoup plus nombreux et peu coûteux à typer en grand nombre. En effet, les puces à SNP permettent de typer des dizaines de milliers de SNP simultanément, avec un taux d'erreur très faible (< 0,1%), pour un coût inférieur au génotypage de 50 microsatellites. Le résultat du génotypage d'un individu est

la détermination de son génotype, AA, AB ou BB, à chaque marqueur (d'allèles A et B) testé sur le génome.

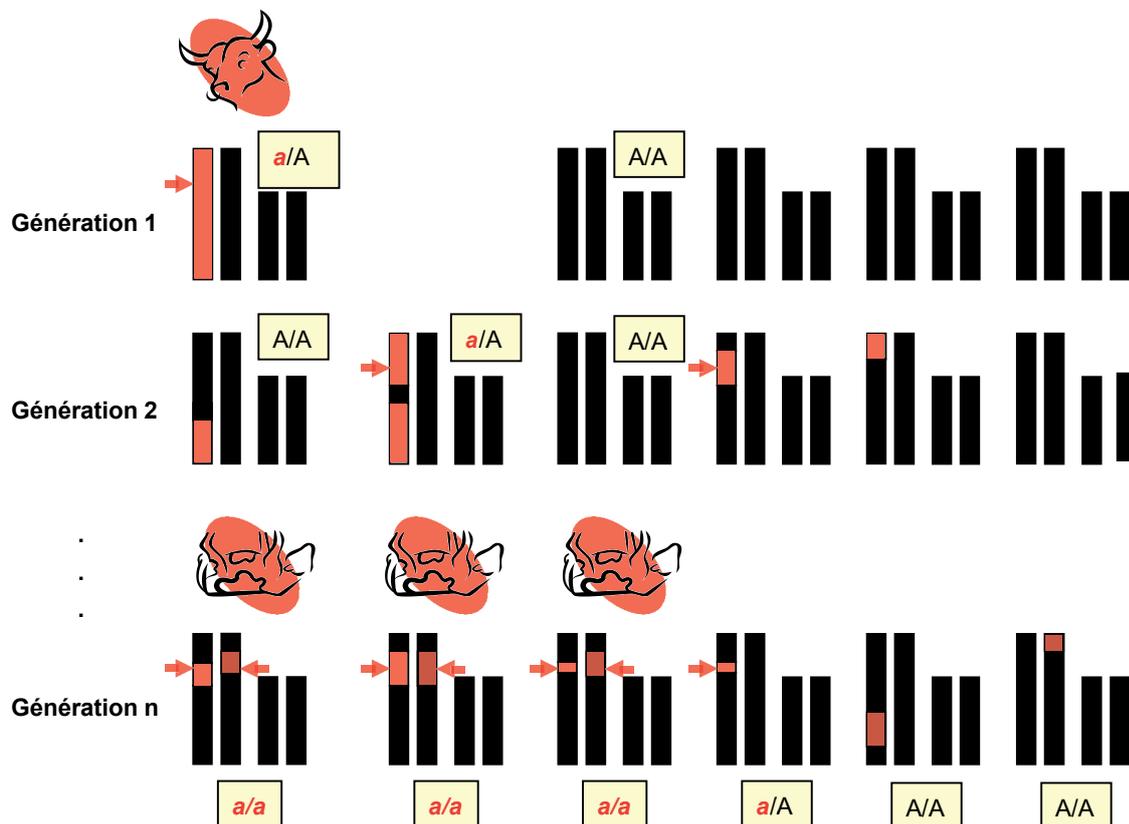
Au cours de la méiose, les chromosomes des gamètes sont produits en conservant de grands blocs chromosomiques parentaux définis par un nombre limité de crossing overs et donc de recombinaisons. Deux locus proches sur le même chromosome ont donc une probabilité élevée d'être transmis ensemble, car la probabilité de recombinaison entre eux est faible. Les méthodes usuelles de cartographie utilisent la forte probabilité de transmission conjointe de deux locus proches et *a contrario* l'indépendance de transmission de deux locus suffisamment distants ou sur des chromosomes différents. Si un gène et un marqueur sont transmis ensemble, on en déduit qu'ils sont proches et on dispose d'une localisation du gène sur le génome. Cette recherche peut être réalisée intra famille et on parle d'analyse de liaison, ou à l'échelle de la population et on parle d'analyse d'association.

Pour les anomalies récessives, on utilise un raisonnement plus poussé qui fait gagner en puissance. En conséquence, il est possible de conclure avec des

effectifs beaucoup plus faibles qu'en analyse de liaison ou d'association usuelle. On parle de cartographie par homozygotie, ou « *homozygosity mapping* ». Cette méthode initialement proposée par Lander et Botstein (1987) a été remise au goût du jour par Charlier *et al* (2008). Elle repose sur deux hypothèses fortes : une anomalie récessive et une mutation unique transmise à tous les cas. Le dispositif est présenté en figure 1. L'ancêtre initial, ayant subi la mutation ou l'ayant reçue d'un de ses parents, la transmet à ses descendants, phénotypiquement normaux comme lui car hétérozygotes. Chaque descendant qui a reçu la mutation a aussi reçu un segment chromosomique entourant cette mutation. Après un nombre suffisant de générations, les premiers individus homozygotes atteints commencent à apparaître, du fait d'accouplements consanguins. Ces individus ont reçu la mutation à la fois de leur père et de leur mère. Ils ont également reçu un fragment du segment chromosomique entourant initialement la mutation, ce fragment étant d'autant plus court qu'il a été plus raboté par les recombinaisons successives. On cartographie alors la mutation dans le plus petit segment chromosomique entièrement homozygote et identique chez tous les individus atteints,

Figure 1. Illustration de la cartographie par homozygotie.

Une mutation récessive apparaît chez l'ancêtre (emplacement de la flèche sur le chromosome rouge). Cette mutation se transmet à l'état hétérozygote à certains de ses descendants phénotypiquement normaux. Après quelques générations, les premiers individus consanguins apparaissent, dont certains, homozygotes à la mutation (les trois individus colorés à la génération n sur la figure), sont atteints et révèlent l'émergence. Autour de la mutation, on observe un segment chromosomique conservé provenant de l'ancêtre ayant transmis l'anomalie, que l'on utilise pour cartographier la mutation.



et rarement homozygote chez les témoins non atteints. Si le dispositif est de taille suffisante, ce segment est unique. Dans Michot *et al* (2015), le dispositif de cartographie de l'épidermolyse bulleuse jonctionnelle était réduit à un seul cas et plusieurs régions étaient homozygotes. Mais dès que l'on dispose de 5 à 10 cas, le nombre de régions se réduit à une seule. Ce segment est d'autant plus petit, donc la cartographie est d'autant plus précise, que le nombre de cas est élevé, que ces cas sont peu apparentés et que le nombre de générations les séparant de l'ancêtre commun est élevé.

En effet, toutes ces conditions augmentent le nombre de méioses et donc le nombre de recombinaisons potentielles diminuant la taille du segment conservé. Dans la majorité des cas étudiés dans notre équipe, le segment détecté était compris entre 1 et 6 Mb. On peut raisonner la recherche simplement en compa-

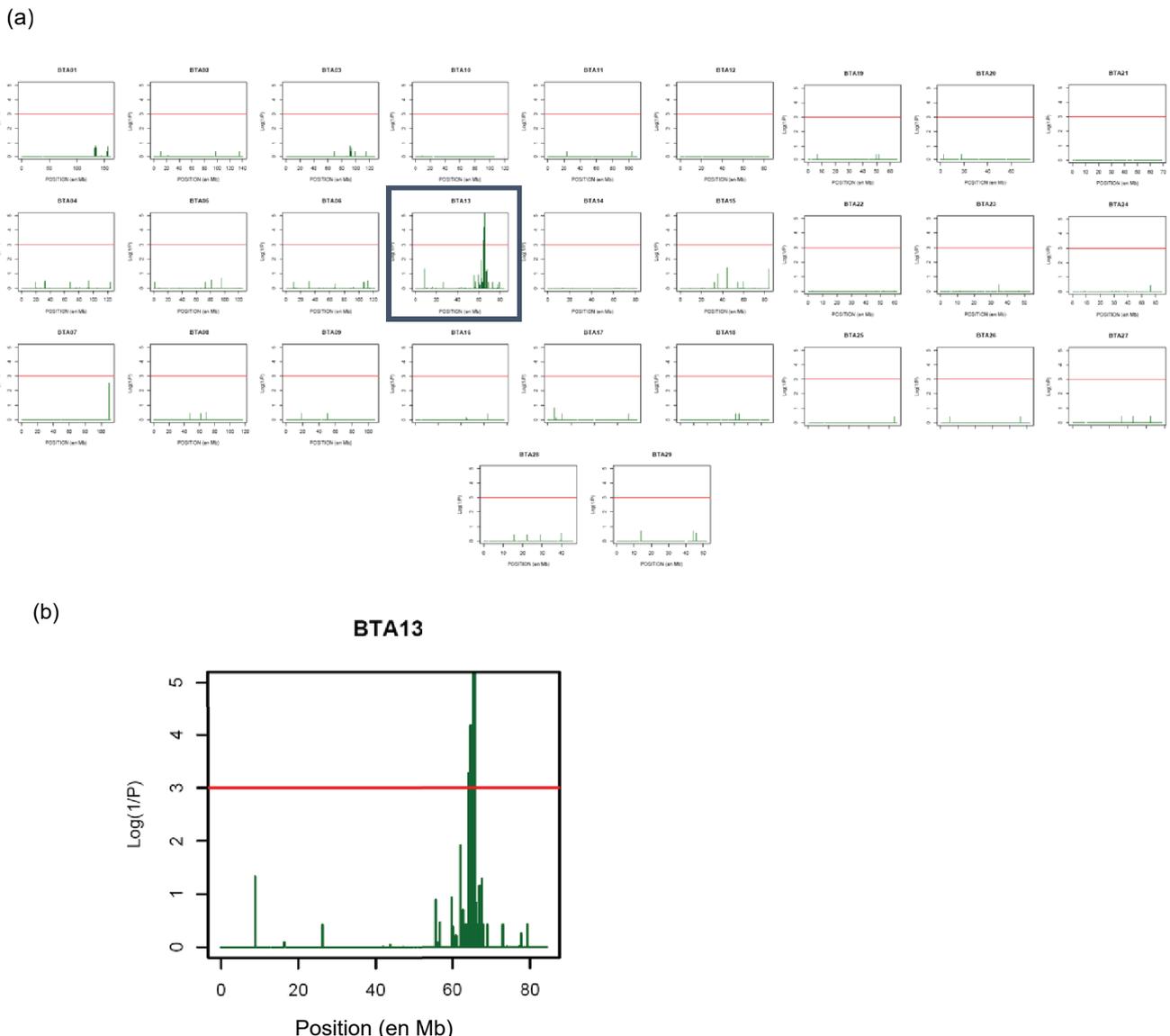
rant la fréquence d'homozygotes identiques chez les cas et les témoins pour des groupes de marqueurs adjacents, ou construire une statistique de test telle que proposé par Charlier *et al* (2008). Le test est alors d'autant plus puissant que les individus atteints sont homozygotes pour un allèle plus rare dans la population. La puissance du dispositif réside dans le fait que sous l'hypothèse nulle, donc du fait du hasard seul, la probabilité d'homozygotie pour les mêmes allèles des mêmes marqueurs décroît très rapidement avec le nombre d'individus. La figure 2 présente l'exemple de l'analyse du Syndrome d'Hypoplasie Généralisée Capréoliforme (SHGC), mettant en évidence une région de 2,5 Mb sur le chromosome 13 (Floriot *et al* 2015).

On comprend aisément que ce dispositif soit sensible aux hypothèses : si le déterminisme génétique est hétérogène avec plusieurs mutations de différentes

origines, il n'existe pas de segments communs à tous les individus atteints ; il en est de même, si tous les animaux ne sont pas atteints du même syndrome. En cas d'hétérogénéité, une solution consiste à restreindre l'analyse à un dispositif plus petit mais plus homogène (par exemple des individus avec des symptômes homogènes au même âge, ou des individus de la même famille), permettant éventuellement de conclure compte tenu de la puissance de l'approche.

La comparaison est relative à une population témoin. Généralement, on utilise un échantillon de la même race, déjà génotypé à d'autres fins, par exemple pour la sélection génomique ou pour l'étude d'un autre caractère, mais non atteint de la même anomalie. Cette population permet d'estimer les fréquences alléliques des marqueurs et donc de construire les tests statistiques. Elle permet aussi de vérifier que la région

Figure 2. Exemple de cartographie par homozygotie : l'anomalie SHGC en race Montbéliarde, localisée dans un intervalle de 2,5 Mb sur le chromosome 13. (a) statistique sur l'ensemble des 29 chromosomes ; (b) zoom sur le chromosome 13.



homozygote détectée ne l'est systématiquement que chez les cas et n'est donc pas un artefact technique. Bien sûr, cette approche perd un peu de puissance lorsque l'haplotype de marqueurs portant l'allèle muté existe sous la même forme à une forte fréquence dans la population générale mais avec l'allèle sauvage. Dans ce cas, une augmentation de la taille de la population contrôlée permet généralement de compenser cette perte de puissance.

Avant l'avènement des puces à SNP, cette approche était possible avec les microsatellites, mais elle était beaucoup plus longue. En effet, on ne disposait généralement dans un premier temps

que d'un marqueur tous les 10 à 20 Mb et on n'obtenait (au mieux) qu'un seul marqueur homozygote chez les individus atteints. Le taux d'échec était assez élevé. Même en cas de succès, il était nécessaire d'utiliser d'autres marqueurs pour confirmer le caractère homozygote et délimiter la région avec plus de précision. Un exemple est fourni dans Duchesne *et al* (2006) pour la cartographie de la syndactylie. Le gain de temps et de travail avec les puces à SNP a été considérable, accélérant fortement l'obtention des résultats. Le tableau 1 fournit une liste de résultats obtenus au cours des dernières années par cette approche dans les races bovines présentes en France.

Pour les anomalies dominantes, ou quand on ignore le déterminisme *a priori* de l'anomalie, les méthodes restent celles de la cartographie classique, avec recherche d'une liaison intra famille ou d'une association populationnelle (selon le dispositif disponible) entre le phénotype et les marqueurs génétiques. Ces méthodes ont toutefois progressé ces dernières années, s'adaptant au type de données génétiques à haut débit disponibles : prise en compte de la structuration des données dans les analyses d'association, analyse multi-marqueurs avec des méthodes bayésiennes (Zhou *et al* 2013).

Tableau 1. Liste d'anomalies bovines identifiées avec l'approche décrite dans cet article (cartographie puis séquençage) et publiées dans OMIA « Online Mendelian Inheritance in Animals - <http://omia.angis.org.au> ».

Anomalie	Race	Chromosome	Nom du gène	Entrée OMIA
<i>Achondrogenesis, type II</i>	Holstein	BTA05	COL2A1	001926-9913
<i>Anhidrotic ectodermal dysplasia</i>	Holstein	BTA05	EDA	000543-9913
<i>Brachyspina</i>	Holstein	BTA21	FANCI	000151-9913
<i>Caprine-like Generalized Hypoplasia Syndrome</i>	Montbéliarde	BTA13	CEP250	001502-9913
<i>Cardiomyopathy, dilated</i>	Holstein	BTA18	OPA3	000162-9913
<i>Citrullinaemia</i>	Holstein	BTA11	ASS1	000194-9913
<i>Coat colour, Agouti</i>	Normande	BTA13	ASIP	000201-9913
<i>Coat colour, dominant red</i>	Holstein	BTA03	COPA	001529-9913
<i>Coat colour, extension</i>	Holstein	BTA18	MC1R	001199-9913
<i>Coat colour, white spotting</i>	Holstein	BTA22	MITF	000214-9913
<i>Complex vertebral malformation</i>	Holstein	BTA03	SLC35A3	000214-9913
<i>Curly hair, karakul-type</i>	Montbéliarde, Simmental	BTA19	KRT27	000246-9913
<i>Deficiency of uridine monophosphate synthase</i>	Holstein	BTA01	UMPS	000262-9913
<i>Depigmentation associated with microphthalmia</i>	Holstein	BTA22	MITF	001931-9913
<i>Epidermolysis bullosa, junctionalis</i>	Charolaise	BTA19	ITGB4	001948-9913
<i>Factor XI deficiency</i>	Holstein	BTA27	F11	000363-9913
<i>Glycogen storage disease V</i>	Charolaise	BTA29	PYGM	001139-9913
<i>Goitre, familial</i>	Salers	BTA14	TG	000424-9913
<i>Hypotrichosis, streaked</i>	Holstein	BTA05	TSR2	000542-9913
<i>Bovine Leucocyte adhesion deficiency</i>	Holstein	BTA01	ITGB2	000595-9913
<i>α-mannosidose</i>	Holstein	BTA06	MAN2B2	000625-9913
<i>β-mannosidose</i>	Holstein	BTA13	MANBA	000626-9913
<i>Muscular hypertrophy (double muscling)</i>	Diverses	BTA02	MSTN	000683-9913
<i>Myopathy of the diaphragmatic muscles</i>	Holstein	BTA23	HSPA1A	001319-9913
<i>Polled and multisystemic syndrome</i>	Charolaise	BTA02	ZEB2	001736-9913
<i>Polled/Horns</i>	Charolaise, Holstein...	BTA01	POLLED	000483-9913
<i>Retinitis pigmentosa 1</i>	Normande	BTA14	RP1	002029-9913
<i>Scurs, type 2</i>	Charolaise	BTA04	TWIST1	001593-9913
<i>Syndactyly (Mule foot)</i>	Holstein	BTA15	LRP4	000963-9913
<i>Vertebral and spinal dysplasia</i>	Holstein	BTA09	T	001951-9913

3 / Un premier diagnostic possible : le test sur haplotype

Tous les individus atteints sont homozygotes pour un même segment constitué de marqueurs génétiques de la puce de génotypage. Ce segment peut être utilisé pour déterminer le statut à la mutation, hétérozygote ou non porteur, d'individus phénotypiquement normaux. Plusieurs solutions sont possibles. On peut utiliser les typages bruts et classer porteur tout individu qui porte au moins un allèle associé à l'anomalie pour tous les marqueurs du segment. On peut aussi se focaliser sur les allèles les plus rares, donc les plus informatifs, et classer comme porteurs les individus portant tous les allèles rares de l'haplotype à risque. Certains auteurs (par exemple, Biffani *et al* 2015) ont proposé des méthodes statistiques, en particulier l'analyse discriminante, pour prédire le statut des animaux typés pour les marqueurs du segment incriminé ou d'un segment plus large.

Une solution plus fiable nécessite le « phasage » des génotypages, c'est-à-dire l'attribution des allèles de chaque marqueur aux chromosomes paternels et maternels de l'individu typé. Cette opération peut sembler lourde mais elle est réalisée de plus en plus souvent en routine en amont des prédictions génomiques, en particulier pour imputer les typages manquants. La présence de l'haplotype considéré, observé chez les cas et généralement très spécifique, est alors associée avec une forte probabilité à l'anomalie. Cette approche permet par exemple de déterminer le statut à l'anomalie des individus génotypés antérieurement avec une puce générique, sans avoir à les régénérer.

Ce diagnostic par haplotype n'est cependant pas sans erreur. On peut considérer différents types d'erreur ou d'indétermination :

i) les faux positifs : l'haplotype porteur de l'anomalie préexistait généralement dans la population avant que la mutation ne se produise. Le segment chromosomique portant cet haplotype de marqueurs existe donc encore sous deux formes, avec ou sans la mutation. Associer le diagnostic de l'anomalie à cet haplotype revient à considérer comme porteurs tous les individus portant l'haplotype, y compris ceux ne portant pas la mutation. Le taux d'erreur est d'autant plus important que l'haplotype utilisé pour le diagnostic est fréquent dans la population. Lorsqu'il est rare, il provient fréquemment de l'ancêtre muté et, dans ce cas, il porte la mutation et le taux d'erreur est faible ;

ii) les faux négatifs : la mutation n'est pas forcément apparue chez l'ancêtre

identifié mais pouvait préexister. Dans ce cas, elle peut être aussi associée à un autre haplotype partiellement différent, du fait de recombinaisons anciennes, de l'haplotype utilisé en diagnostic. Dans ce cas, on peut conclure à tort à un diagnostic non porteur. La même erreur peut se produire en cas de phasage incorrect ;

iii) les indéterminés : il existe différentes raisons conduisant à une indétermination. La plus fréquente est une recombinaison dans l'haplotype diagnostic. Dans ce cas, l'individu ne porte qu'une fraction de l'haplotype à risque et tant que la mutation n'est pas localisée plus précisément, il n'est pas possible de savoir si la mutation a été gardée ou si elle a été éliminée avec le morceau perdu ;

À ce stade, les résultats peuvent déjà être utilisés pour la gestion de l'anomalie dans la race et, en pratique, cela a été le cas dans de nombreuses situations avant la découverte de la mutation causale (« Syndrome d'Hypoplasie Généralisée Capréoliforme » (SHGC), veaux tourneurs, ataxie, déficience en cholestérol ou « Cholesterol Deficiency in Holstein » (CDH)...).

4 / Recherche de mutations candidates par séquençage du génome

La région identifiée fait généralement plusieurs mégabases, la localisation est donc précise puisque cette région ne représente qu'un millième environ du génome complet. En revanche, cette région est encore très grande et contient plusieurs dizaines de milliers de polymorphismes qu'il convient d'identifier puis de filtrer pour ne conserver que le ou les meilleurs candidats. Avant l'avènement du séquençage à haut débit, le séquençage était long et coûteux et il n'était pas envisageable de séquencer une région si grande. L'approche était donc la suivante :

i) par une approche de génomique comparée avec l'Homme ou la souris, identifier l'ensemble des gènes présents dans cette région, en général quelques dizaines, parfois quelques centaines ;

ii) parmi eux, sélectionner les meilleurs gènes candidats sur la base de leur annotation fonctionnelle, en particulier leur rôle dans des anomalies similaires chez l'Homme ou la souris ;

iii) séquencer la partie codante de ces gènes en espérant cibler la bonne région.

Par exemple, dans l'étude de la syndactylie (Duchesne *et al* 2006), après une étape de cartographie, le choix du gène

candidat LRP4 a été suggéré à la fois par sa position dans l'intervalle étudié et par l'existence de deux lignées de souris polysyndactyles, du fait d'une mutation dans ce même gène (Simon-Chazottes *et al* 2006). Mais en l'absence de gène candidat fort, cette approche pouvait se révéler très aléatoire.

À nouveau, l'évolution des technologies a permis d'avancer beaucoup plus vite et beaucoup plus sûrement. Dans un premier temps a été développée une technologie de capture de région du génome associée à du séquençage. Le principe est le suivant :

i) la région étant définie et sa séquence de référence connue, on fait synthétiser des oligonucléotides complémentaires de cette région, en nombre suffisant pour recouvrir totalement la région (hors séquences répétées) avec un certain chevauchement ;

ii) ces oligonucléotides sont utilisés pour capturer les séquences complémentaires du ou des individus étudiés, dont la région génomique est suffisamment peu différente pour hybrider avec cette matrice, sur support solide ou en phase liquide selon les procédés ;

iii) les fragments d'ADN capturés sont ensuite séparés et séquencés.

Le gain de cette approche réside surtout dans la petite fraction à séquencer par rapport au génome total et donc au coût plus limité par échantillon, autorisant à travailler sur un nombre bien plus élevé d'échantillons. Cette technologie a été utilisée en 2010 pour la caractérisation de l'anomalie SHGC en race Montbéliarde (Floriot *et al* 2015). Elle a comme difficulté essentielle le temps et le travail nécessaires pour dessiner les oligonucléotides de capture, ainsi que le coût de leur synthèse qui ne se justifie que si le nombre d'échantillons à traiter est important. En pratique, cette approche de capture est très largement utilisée chez l'Homme pour l'exome (l'ensemble des exons de tous les gènes) car ceci peut être réalisé avec une puce de capture générique, donc moins chère (Warr *et al* 2015). En revanche, la capture d'une région spécifique devient de plus en plus rare.

Aujourd'hui, le coût du séquençage a diminué, son débit a augmenté et il paraît souvent plus simple, beaucoup plus rapide et pas plus coûteux de séquencer l'ensemble du génome, même si on n'est intéressé que par une petite région représentant un millième du total. On séquence généralement le génome d'au moins deux cas. Deux des individus choisis sont ceux qui déterminent les bornes gauche et droite du segment homozygote conservé.

Choisir plusieurs individus permet également de s'affranchir d'éventuels artefacts techniques de séquençage et de distinguer et donc d'éliminer les néo-mutations apparues chez l'un d'entre eux mais pas chez les autres. Le traitement bioinformatique des résultats est le suivant :

i) les séquences sont d'abord alignées sur le génome de référence et l'on ne s'intéresse ensuite qu'aux séquences de la région ciblée ;

ii) on détecte et retient tous les variants différents de la référence et, à ce stade, on observe de l'ordre de quelques milliers de variants ;

iii) on filtre ensuite les variants avec différents critères, avec l'objectif d'en éliminer le maximum : élimination des variants à l'état hétérozygote chez les atteints ; élimination des variants présents chez l'un des individus atteints mais pas chez les autres ; élimination des variants présents à l'état homozygote dans une population contrôlée connue pour être non affectée ; si on pense que l'anomalie est limitée à la race étudiée (car récente), on peut éliminer tous les variants présents dans une autre race.

Ces deux derniers filtres peuvent être particulièrement puissants si cette population contrôlée est très grande. Nous utilisons dans ce cadre la population du projet « 1000 bull genomes » (Daetwyler *et al* 2014) qui dépasse 2500 génomes séquencés en octobre 2016. Parce que c'est techniquement plus simple et plus rapide, on tend à étudier prioritairement les polymorphismes de petite taille (SNP et petites insertions ou délétions). Mais il convient également de rechercher les variants structuraux qui sont moins nombreux mais qui ont plus fréquemment un effet biologique. Pour cela, un outil de choix est le logiciel « Integrative Genomics Viewer » (IGV). Ainsi, par cette approche, Michot *et al* (2015) ont mis en évidence une délétion de 4,8 kbases dans le gène ITGB4 responsable du syndrome d'épidermolyse bulleuse jonctionnelle en race Charolaise.

Toute cette approche repose sur la qualité de l'alignement des séquences produites sur la séquence de référence. Une difficulté importante est la qualité encore imparfaite des séquences de référence dans les espèces d'élevage, avec de nombreuses régions absentes de la référence et des régions mal positionnées sur le génome. Dans le premier cas, les séquences produites ne sont pas alignées et sont perdues, dans le second, elles peuvent être exclues de l'analyse car pas dans la région ciblée. Dans ces deux situations, un défaut local de la séquence de référence peut conduire à

rater la mutation recherchée. Dans ce cas, un effort spécifique doit être réalisé pour reconstruire localement une séquence de référence. On peut supposer cependant que cette période n'est que transitoire et que la qualité des séquences de référence deviendra rapidement suffisante pour que ce genre de difficulté se rencontre de plus en plus rarement.

Dans la population, il peut exister des individus atteints et sains homozygotes pour le même haplotype de marqueurs, mais avec ou sans la mutation responsable de l'anomalie, les individus atteints étant alors des descendants de l'ancêtre chez qui la mutation s'est produite. Séquencer ces deux types d'individus est une approche très efficace, car le nombre de variants différents entre ces deux haplotypes est très réduit (voire limité à la mutation causale), ce qui facilite énormément la phase de filtrage et de ciblage du variant recherché.

Dans la phase suivante, on peut souvent sélectionner les variants résiduels en fonction des connaissances existantes sur les gènes impliqués. Si l'orientation vers un gène candidat est souvent hasardeuse à l'échelle du génome entier et n'est donc pas recommandée, la région ciblée contient suffisamment peu de gènes (quelques dizaines) pour que le risque d'en étudier seulement un en particulier soit raisonnable. En principe, le choix se porte sur des gènes « candidats fonctionnels », car impliqués dans des anomalies comparables dans d'autres espèces, en particulier l'Homme et la souris. Grohs *et al* (2016 ce numéro) décrivent les ressources disponibles dans ces deux espèces, beaucoup plus complètes que celles des espèces d'élevage.

De façon générale, les mutations qui affectent la région codante d'un gène et entraînent une modification de la protéine ont souvent des conséquences biologiques plus graves que celles qui se situent en dehors du codant. Cette affirmation compte beaucoup d'exceptions, mais elle est suffisamment vérifiée pour que les variants dans la partie codante du génome soient étudiés en priorité. Ainsi la stratégie choisie repose fréquemment sur le séquençage de l'exome en génétique humaine. Comme indiqué dans le premier article (Boichard *et al* 2016), on cible prioritairement les mutations *i*) ajoutant ou supprimant des codons stop, *ii*) entraînant un décalage du cadre de lecture, *iii*) modifiant le site d'initiation ou un site d'épissage, *iv*) induisant une substitution d'acides aminés avec des propriétés très différentes, ou bien encore *v*) affectant une région très conservée de la protéine entre espèces. Les logiciels de prédiction des fonctions des protéines comme

« SIFT » (Kumar *et al* 2009) ou « Polyphen » (Adzhubei *et al* 2010) permettent d'orienter les choix avec efficacité. Restreindre la liste des variants à ces types de mutations conduit souvent à ne conserver qu'une seule ou un tout petit nombre de mutations candidates.

Les arguments évolutifs apportent également beaucoup d'information. L'évolution est un processus qui combine l'apparition de polymorphismes par mutation et la fluctuation de fréquence de ces variants sous diverses forces évolutives. Lorsqu'une région du génome est essentielle au bon fonctionnement de l'organisme, les mutations qui s'y produisent naturellement et au hasard sont fortement contre-sélectionnées et on constate une forte conservation de cette région, en particulier entre espèces. Plus le mécanisme impliqué est fondamental, plus large est la gamme d'espèces chez lesquelles on observe cette conservation, au niveau de la protéine codée, voire parfois au niveau de l'ADN. Le fait qu'une mutation touche une région du génome très conservée ou modifie un acide aminé très conservé d'une protéine dans une large gamme d'espèces constitue un argument fort pour la retenir comme candidate probable. À noter que les logiciels SIFT et Polyphen prennent également en compte les homologies entre gènes orthologues et s'appuient donc aussi sur ces arguments évolutifs.

5 / Confirmation des résultats

Toute analyse bioinformatique ne fournit que des variants candidats qu'il faut vérifier et valider. Tout d'abord, il faut vérifier, par exemple par amplification par PCR d'une petite région génomique entourant la mutation et séquençage Sanger dans les deux sens, que la mutation est bien présente chez les atteints et n'est pas un artefact.

Ensuite, il faut analyser une population assez large pour vérifier qu'aucun homozygote normal n'est observé (ce qui invaliderait le variant candidat), que le variant est bien spécifique de la race (ou bien rechercher dans quelles races il est présent), confirmer le génotype d'individus connus comme atteints, sains et porteurs, estimer le niveau de concordance avec la prédiction par haplotype, et estimer la fréquence de l'anomalie.

Ce travail peut être réalisé avec un test spécifique de cette mutation. C'est ce qui a été réalisé dans un premier temps avec l'anomalie SHGC en race Montbéliarde. Un test a été mis en œuvre sur 1 140 animaux de race Montbéliarde (dont 190 atteints) et sur un panel de 316 animaux de 10 autres races (Floriot *et al* 2015).

Mais la solution la plus souvent privilégiée aujourd'hui est l'ajout du test de cette mutation sur une puce de génotypage existante (pour l'INRA, la puce EuroG10k d'Eurogenomics), permettant de réaliser le génotypage pour la mutation à un coût marginal et sur de très grands effectifs. Cette option permet ensuite une mise en production instantanée du test avec diffusion des résultats, puisque c'est le même outil qui est utilisé pour sa validation puis la production. Ainsi, par exemple, le test de l'anomalie de dégénérescence rétinienne liée au gène RP1 a été validé sur 110 652 animaux de 9 races avec la puce EuroG10k avant d'être mis en production par simple diffusion des résultats déjà obtenus en routine (Michot *et al* 2016). Cette stratégie n'a qu'un seul inconvénient, elle nécessite d'attendre le développement et la mise en production d'une nouvelle version de puce, soit fréquemment entre 6 et 12 mois. En revanche, dès qu'elle est opérationnelle, cette solution donne des résultats très rapidement, compte tenu du grand nombre de typages réalisés en routine.

6 / Validation fonctionnelle

La validation fonctionnelle consiste à prouver l'impact de la mutation, sur l'espèce hôte, ou sur une espèce modèle. Alors que l'évolution des technologies a fortement changé les méthodes et accéléré l'identification de la mutation candidate responsable d'une anomalie, la validation fonctionnelle reste une tâche difficile, nécessitant de mettre en œuvre des techniques et des compétences très variées. La lourdeur de la tâche empêche de conduire ces travaux de validation fonctionnelle de façon systématique et, dans bien des cas, on est obligé de se contenter des résultats précédents. Il est ainsi fréquent que l'anomalie soit gérée dans une race avant que la mutation identifiée ne soit formellement validée et publiée. Plusieurs approches peuvent être mises en œuvre.

Par un screening important de la population, on peut trouver des individus homozygotes pour des haplotypes identiques dans la région étudiée, à l'exception de la mutation candidate. Cette option est d'autant plus réaliste que l'haplotype est fréquent et que la mutation est récente. Dans ce cas, si les individus non porteurs de la mutation sont normaux et que les porteurs sont atteints, ils constituent en eux-mêmes une preuve de causalité, mais ils n'apportent pas d'explication du mécanisme impliqué.

Lorsque ce screening ne permet pas de conclure, il faut prouver la causalité expérimentalement. Le choix de la méthode de validation dépend essentiellement de

la nature de la mutation causale identifiée, et des tissus biologiques disponibles. Il est judicieux de prélever des tissus au moment où les premiers animaux atteints sont décrits et phénotypés finement. En effet, on constate que lorsque les tests de dépistage deviennent disponibles, les mesures prises dans les populations (éradication de l'anomalie, orientation des accouplements) portent leurs fruits très rapidement et le nombre de cas peut devenir limitant pour la recherche. Par ailleurs, il faut également collecter les mêmes tissus chez des animaux contrôles (phénotypiquement sains, de même race et de même âge idéalement). Plusieurs cas de figures peuvent se présenter :

i) si la mutation a un effet potentiel sur la transcription du gène (modification du site d'initiation, d'un site d'épissage, introduction d'un codon stop prématuré qui induit la dégradation des ARNm non sens par le mécanisme de « *Nonsense-Mediated Decay* » ou NMD...), il convient de vérifier le niveau d'expression du gène. Le choix du ou des tissus à étudier est crucial, ainsi que l'âge de l'animal au moment du prélèvement. Certains gènes, par exemple, sont exprimés majoritairement pendant le développement embryonnaire, et l'étude de leur niveau d'expression après la naissance de l'animal n'a plus de sens. Idéalement, le niveau d'expression sera quantifié par PCR quantitative, et les transcrits seront séquencés pour identifier des transcrits alternatifs ;

ii) si la mutation a un effet supposé sur l'expression de la protéine (quantité plus faible de transcrits, protéine tronquée...), le niveau d'expression de la protéine peut être évalué par western blot, à partir de tissus sains et atteints sélectionnés de la même manière que ci-dessus. La difficulté principale de cette approche est l'accès à des anticorps qui puissent fonctionner sur l'espèce bovine. En effet, les anticorps sont majoritairement mis au point pour fonctionner chez l'Homme et/ou chez la souris. Leur utilisation chez le bovin, en hybridation hétérologue, conduit souvent à des hybridations croisées aspécifiques qui peuvent masquer le vrai signal, ou à une absence d'hybridation lorsque les séquences protéiques entre le bovin et l'espèce pour laquelle a été produit cet anticorps sont trop différentes. Par ailleurs, le western blot ne permet pas de vérifier l'impact d'une simple substitution d'acide aminé, ou bien d'une différence de taille minimale entre protéines normale et mutée ;

iii) la localisation de la protéine peut également être étudiée par immunohistochimie ou immunofluorescence sur des coupes de tissus, une mutation pouvant

affecter la présence ou les conditions d'adressage, de transport, d'ancrage de la protéine à d'autres molécules, et ainsi affecter sa fonction ;

iv) l'effet de la mutation peut aussi être testé à plus grande échelle, avec l'hypothèse suivante : elle affecte globalement le niveau d'expression des transcrits et/ou des protéines. Dans ce cas, l'étude du transcriptome (par puce à expression ou bien par séquençage des ARN) et/ou du protéome (par spectrométrie de masse MS-MS) peut permettre de conclure sur la causalité. Là encore, du choix du tissu et du stade à étudier dépend la conclusion de l'étude ;

v) Les cultures de cellules sont un cadre approprié pour tester des mécanismes cellulaires, par exemple une modification d'expression d'un gène analysable au niveau de l'ARN ou de la protéine, une modification de la morphologie cellulaire, de l'activité enzymatique, de la concentration en un produit de chaîne métabolique, globalement ou dans un compartiment de la cellule, ou de toute autre activité intracellulaire. Les cellules sont préalablement transformées par mutagenèse dirigée, de sorte que les deux types cellulaires peuvent être comparés ;

vi) la dernière option, la plus lourde, pour valider la mutation consiste à reproduire le phénotype dans une espèce modèle telle que la souris ou le poisson-zèbre, le choix de l'espèce modèle dépendant du gène étudié, et de sa fonction dans la biologie de l'animal. Cette option est aujourd'hui largement facilitée par les outils de modification du génome (CRISPR/Cas9). Il est donc possible d'invalider un gène, ou bien d'introduire une mutation spécifique, avec des taux de succès bien supérieurs à ceux de la recombinaison homologue, et l'obtention de lignées transgéniques peut se faire en quelques mois dans les cas les plus favorables. Le processus passe par la création de la mutation dans une lignée de souris, la reproduction de la lignée pour obtenir des homozygotes, puis leur analyse clinique complexe, à différentes échelles (individu, organe, cellule) et différents stades physiologiques. L'utilisation des espèces modèles permet de disposer plus facilement de l'ensemble des tissus, ainsi que de plusieurs stades pour étudier finement l'impact de la mutation sur l'individu. En revanche, des phénotypes spécifiques d'espèces ne peuvent pas être reproduits dans une autre espèce, comme c'est le cas de la présence de corne chez la vache, et cette approche ne peut alors être utilisée. De même, il est possible que le gène étudié s'exprime différemment dans l'espèce modèle et dans ce cas, les résultats sont difficilement exploitables.

7 / Les limites de l'approche « du phénotype à la mutation »

L'approche présentée précédemment est maintenant très efficace avec les technologies actuelles de génotypage et de séquençage. En revanche, elle repose sur une collection de cas homogènes de taille suffisante. Du coup, elle est inopérante quand cette condition n'est pas remplie et plusieurs exemples sont décrits ci-dessous, avec les moyens éventuels à mettre en œuvre pour contourner la difficulté :

i) l'anomalie se traduit par un syndrome non spécifique et souvent banal. Par exemple, le BLAD se traduit par des infections banales répétées, digestives et respiratoires. Ce genre d'anomalie peut donc passer inaperçue, car trop ressemblante aux pathologies usuelles, généralement d'origine environnementale et bien plus nombreuses. Pour surmonter ce handicap, il faut mieux cibler l'anomalie. Un observatoire des mortalités peut mettre en évidence les pères porteurs quand ils ont un nombre suffisant de descendants, et surtout, on constate que certains accouplements, interprétés *a posteriori* comme entre hétérozygotes, induisent une forte augmentation de la pathologie. Cette approche n'est efficace que si la fréquence de l'anomalie est déjà élevée, permettant l'apparition d'un nombre significatif de cas. Mais il convient surtout de raffiner le phénotype, pour le rendre plus spécifique de l'anomalie. Par exemple, le BLAD est caractérisé par une numération sanguine des globules blancs extraordinairement élevée (ils se multiplient sans pouvoir combattre les infections) et une numération de centaines de milliers de leucocytes par μl est un signe d'appel fort (Kehrli *et al* 1990). Autre signe, les infections ne sont pas purulentes puisque les leucocytes privés d'adhésines restent pri-

sonniers des vaisseaux sanguins et ne peuvent pas être mobilisés dans le foyer infectieux. Ce diagnostic plus précis permet de constituer des collections homogènes. Encore faut-il le développer, ce qui nécessite des études préalables sur des cas identifiés et on revient à la difficulté de départ, avant de le mettre en œuvre sur les cas suspects. C'est dans cet esprit que l'ONAB combine deux types d'information, des déclarations par des éleveurs ou des techniciens, plus nombreuses mais souvent peu précises, et des diagnostics cliniques plus complets réalisés par des vétérinaires praticiens ou dans les écoles vétérinaires après hospitalisation de cas ;

ii) une autre situation peut échapper à l'observation. Elle correspond à un phénotype compatible avec l'élevage, l'animal pouvant être élevé et produire comme ses congénères. La domestication a diminué la pression de sélection naturelle et a permis l'apparition d'une forme nouvelle de variabilité qui ne serait pas tolérée à l'état sauvage. La variabilité de couleur de la robe entre clairement dans cette catégorie. Mais c'est aussi le cas de syndromes invalidants tels que la dégénérescence rétinienne due à une mutation dans le gène RPI. Les animaux perdent progressivement la vision à partir de 4 ans, d'abord dans l'obscurité puis en pleine lumière, et deviennent complètement aveugles vers 5 à 6 ans. Dans un milieu d'élevage, ces animaux restent intégrés au troupeau, suivent leurs congénères et continuent de produire, de sorte que ce type d'anomalie peut rester inaperçu ou sous-estimé en importance. La forte fréquence de cette anomalie en race bovine Normande reste inexplicée, on sait seulement qu'elle est ancienne et plutôt en réduction progressive. Ce type d'anomalie est davantage développé dans Fritz *et al* (2016 ce numéro) ;

iii) une anomalie conduisant à une mortalité précoce durant la gestation est

un cas encore plus complexe. Dans ce cas, l'anomalie se traduit par une baisse de fertilité et passe généralement totalement inaperçue. Cette catégorie d'anomalie est sans doute assez fréquente car de nombreux gènes, critiques pour le développement, peuvent en être responsables. Aucun échantillon biologique n'est récupérable et l'approche décrite dans cet article ne peut pas être mise en œuvre. Une approche spécifique, reposant sur la fouille de grandes bases de données de génotypage, est décrite dans l'article de Fritz *et al* (2016 ce numéro).

Conclusion

L'approche traditionnelle de caractérisation des anomalies récessives du phénotype au génotype a atteint un niveau d'efficacité remarquable grâce aux évolutions technologiques récentes (puce de génotypage, séquençage, bioinformatique, annotation du génome) et aux stratégies qui s'y sont adaptées. Le goulot d'étranglement a toujours été la disponibilité des cas, mais le constat est encore plus vrai aujourd'hui, alors que la plupart des autres étapes sont moins limitantes. Ce gain d'efficacité est cependant un argument de plus en plus convaincant : alors qu'il y a quelques années encore, la déclaration de cas était considérée comme un risque pour le déclarant (risque financier, commercial, juridique, ou d'image) sans garantie de solution à court terme, il constate aujourd'hui qu'une réponse peut lui être apportée en un temps court, avec une solution technique mais aussi des conséquences favorables au niveau de sa protection juridique et commerciale. Pour le chercheur, si les progrès ont été spectaculaires, il reste encore d'importantes marges de progrès, par exemple dans la prédiction bioinformatique de l'effet des variants et surtout dans la validation fonctionnelle des résultats.

Références

- Adzhubei I.A., Schmidt S., Peshkin L., Ramensky V.E., Gerasimova A., Bork P., Kondrashov A.S., Sunyaev S.R., 2010. A method and server for predicting damaging missense mutations. *Nat. Methods*, 7, 248-249.
- Biffani S., Dimauro C., Macciotta N., Rossoni A., Stella A., Biscarini F., 2015. Predicting haplotype carriers from 709 snp genotypes in *bos taurus* through linear discriminant analysis. *Genet. Select. Evol.*, 47, 4.
- Boichard D., 2002. Pedig : a fortran package for pedigree analysis suited to large populations. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France, 28-13.
- Boichard D., Le Roy P., Levezuel H., Elsen J.M., 1998. Utilisation des marqueurs moléculaires en génétique animale. In : Numéro spécial, Biotechnologies animales. Gosclaud F. (Ed). INRA Prod. Anim., 11, 67-80.
- Boichard D., Grohs C., Danchin-Burge C., Capitan A., 2016. Les anomalies génétiques : définition, origine, transmission et évolution, mode d'action. In : Anomalies génétiques. Boichard D. (Ed). Dossier, INRA Prod. Anim., 29, 297-306.
- Charlier C., Coppeters W., Rollin F., Desmecht D., Agerholm J.S., Cambisano N., Carta E., Dardano S., Dive M., Fasquelle C., Frennet J.C., Hanset R., Hubin X., Jorgensen C., Karim L., Kent M., Harvey K., Pearce B.R., Simon P., Tama N., Nie H., Vandeputte S., Lien S., Longeri M., Fredholm M., Harvey R.J., Georges M., 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet.*, 40, 449-454.
- 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., Coote D., Chamberlain A.J., Anderson C., Van Tassell C.P., Hulsege I., Goddard M.E., Gulbrandsen B., Lund M.S., Veerkamp R.F., Boichard D., Fries R., Hayes B.J., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.*, 46, 858-867.
- Duchesne A., Gautier M., Chadi S., Grohs C., Floriot S., Gallard Y., Caste G., Ducos A.,

- Eggen A., 2006. Identification of a doublet missense substitution in the bovine LRP4 gene as a candidate causal mutation for syndactyly in Holstein cattle. *Genomics*, 88, 610-621.
- Floriot S., Vesque C., Rodriguez S., Bourgain-Guglielmetti F., Karaïskou A., Gautier M., Duchesne A., Barbey S., Fritz S., Vasilescu A., Bertaud M., Moudjou M., Halliez S., Cormier-Daire V., El Hokayem J., Nigg E.A., Manciaux L., Guatteo R., Cesbron N., Toutirais G., Eggen A., Schneider-Maunoury S., Boichard D., Sobczak-Thépot J., Schibler L., 2015. C-Nap1 mutation affects centriole cohesion and is associated with a Seckel-like syndrome in cattle. *Nat. Comm.*, 6, 6894.
- Fritz S., Michot P., Hozé C., Grohs C., Barbat A., Boussaha M., Boichard D., Capitan A., 2016. Anticiper l'émergence d'anomalies génétiques grâce aux données génomiques. In : *Anomalies génétiques*. Boichard D. (Ed). Dossier, INRA Prod. Anim., 29, 339-350.
- Grohs C., Duchesne A., Floriot S., Deloche M.C., Boichard D., Ducos A., Danchin-Burge C., 2016. L'Observatoire National des Anomalies Bovines, son action et ses résultats pour une aide efficace à la gestion des anomalies génétiques. In : *Anomalies génétiques*. Boichard D. (Ed). Dossier, INRA Prod. Anim., 29, 307-318.
- Kehrli M.E., Schmalstieg F.C., Anderson D.C., Van der Maaten M.J., Hughes B.J., Ackermann M.R., Wilhelmson C.L., Brown G.B., Stevens M.G., Whetstone C.A., 1990. Molecular definition of the bovine granulocytopeny syndrome - identification of deficiency of the Mac1 (CD11b/CD18) glycoprotein. *Am. J. Vet. Res.*, 51, 1826-1836.
- Kumar P., Henikoff S., Ng P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, 4, 1073-1081.
- Lander E.S., Botstein D., 1987. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*, 236, 1567-1570.
- Michot P., Chahory S., Marete A., Grohs C., Dagios D., Donzel E., Aboukadi A., Deloche M.C., Allais-Bonnet A., Chambrial M., Barbey S., Boussaha M., Danchin-Burge C., Fritz S., Boichard D., Capitan A., 2016. A reverse genetic approach identifies an ancient frameshift mutation in RPI causing recessive progressive retinal degeneration in European cattle breeds. *Genet. Select. Evol.*, 48, 56.
- Michot P., Fantini O., Braques R., Allais-Bonnet A., Saintilan R., Grohs C., Barbieri J., Genestout L., Danchin-Burge C., Gourreau J.M., Boichard D., Pin D., Capitan A., 2015. Whole-genome sequencing identifies a homozygous deletion encompassing exons 17 to 22 of the Integrin Beta 4 Gene in a Charolais calf with Junctional Epidermolysis Bullosa. *Genet. Select. Evol.*, 47, 37.
- Simon-Chazottes D., Tutois S., Kuehn M., Evans M., Bourgade F., Cook S., Davisson M.T., Guénet J.L., 2006. Mutations in the gene encoding the low-density lipoprotein receptor LRP4 cause abnormal limb development in the mouse. *Genomics*, 87, 673-677.
- Warr A., Robert C., Hume D., Archibald A., Deeb N., Watson M., 2015. Exome sequencing: current and future perspectives. *G3 (Bethesda)*, 5, 1543-1550.
- Zhou X., Carbonetto P., Stephens M. 2013. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.*, 9, e1003264.

Résumé

Cet article présente la méthodologie utilisée pour identifier la mutation responsable d'une anomalie génétique à partir de cas d'animaux affectés. Dans un premier temps, une collection de cas aussi homogènes que possible est constituée, de la même race et avec les mêmes signes cliniques, complétée par une population témoin apparentée mais non atteinte. Une analyse de pedigree est possible pour rechercher l'ancêtre commun qui a pu transmettre l'anomalie à chacun des cas. Le génotypage par puce permet de mettre en évidence très rapidement une petite région du génome homozygote et identique à tous les marqueurs qui contient la mutation recherchée. La mutation est ensuite identifiée par séquençage du génome de quelques cas, filtrage des variants observés sur la base d'une part, de leur présence chez d'autres animaux, et d'autre part, de leur annotation fonctionnelle. Une validation statistique est ensuite pratiquée par génotypage à grande échelle, pour vérifier l'association totale entre génotype et phénotype. Enfin, la causalité de la mutation est étudiée par analyse fonctionnelle, incluant l'analyse des ARN et des protéines, l'imagerie cellulaire, voire la création de modèles transgéniques.

Abstract

From phenotype to the causal mutation: recessive defects

This article presents the methodology used to identify the mutation responsible for a genetic defect from the observation of cases. In the first step, a set of homogeneous cases are collected, from the same breed and with the same clinical signs. This collection is completed by a related but unaffected control population. A pedigree analysis is possible to point towards a common ancestor who may have transmitted the defect to all cases. A genotyping step using a SNP chip is used to display a small genomic region homozygous and identical at all markers and including the mutation. The mutation is then identified by genome sequencing of a few cases followed by the filtering of the variants against a large sequence database of unaffected animals. The best candidate variants are retained on the basis of their functional annotation. The mutation is then statistically confirmed on the basis of large scale genotyping, to verify the complete association between the mutation and the phenotype. Finally, the causality of the mutation is proven by functional analysis, including RNA and protein analysis, cellular imaging, and even through transgenic models carrying the mutation.

DUCHESNE A., GROHS C., MICHOT P., BERTAUD M., BOICHARD D., FLORIOT S., CAPITAN A., 2016. Du phénotype à la mutation causale : le cas des anomalies récessives bovines. In : *Anomalies génétiques*. Boichard D. (Ed). Dossier, INRA Prod. Anim., 29, 319-328.