

## Stratégie de recherche des gènes impliqués dans la croissance folliculaire

L'objectif de cette recherche est d'identifier les gènes responsables de la variabilité naturelle du taux d'ovulation, dans une perspective d'amélioration génétique. Nous n'envisageons ici que les gènes impliqués dans la croissance folliculaire, ou folliculogénèse, l'une des composantes du taux d'ovulation, l'autre composante étant l'atrésie folliculaire.

La stratégie du gène « candidat positionnel » proposée ici comprend quatre volets :

1 - l'identification des régions du génome où se trouvent des gènes impliqués dans la variabilité du caractère ;

2 - l'identification et la caractérisation des gènes impliqués dans la fonction responsable du caractère ;

3 - la recherche de gènes candidats par co-localisations entre les régions identifiées en 1 et les gènes identifiés en 2 ;

4 - la recherche d'un polymorphisme des gènes candidats et d'un lien entre ce polymorphisme et la variabilité du caractère.

### Identifier les régions du génome

Cette identification consiste en la recherche de locus quantitatifs (ou QTL pour Quantitative Trait Loci), c'est-à-dire de régions du génome impliquées dans l'expression d'un caractère quantitatif polygénique, ici le taux d'ovulation. Approche exclusivement génétique, elle ne nécessite aucune information préalable de type physiologique, biochimique ou pathologique sur la nature et le lieu d'expression des gènes impliqués.

Cette recherche nécessite :

- des animaux issus de familles convenablement constituées et pour lesquels des mesures de différents caractères sont effectuées ; ces familles doivent être informatives, c'est-à-dire présenter des variations de performances pour les caractères analysés, en particulier ici pour le taux d'ovulation ;

- des marqueurs polymorphes répartis sur l'ensemble du génome. Ces marqueurs sont essentiellement des microsatellites, c'est-à-dire des répétitions, en nombre variable, d'un même motif de 1 à 4 bases, par exemple (CA)<sub>n</sub>. Ce nombre de répétitions peut être déterminé facilement par amplification de l'ADN génomique par PCR (Polymerase Chain Reaction).

La recherche des locus quantitatifs consiste en une analyse de liaison entre d'une part les différents marqueurs et d'autre part le caractère « taux d'ovulation », c'est-à-dire que l'on recherche une coségrégation, association préférentielle entre certains allèles d'un marqueur donné et des valeurs faibles ou fortes du taux d'ovulation.

L'INRA a mis en place en 1992 un programme de recherches de locus quanti-

tatifs sur les descendants d'un croisement Meishan/Large White. L'analyse porte sur 1 000 animaux F2 (500 mâles et 500 femelles), un premier lot de 63 marqueurs microsatellites (CA) et une centaine de mesures couvrant des caractères de croissance, de qualité de la viande, de reproduction et de comportement. Pour la reproduction, sont mesurés l'âge à la puberté (dosage de progestérone), l'âge au premier œstrus, le taux d'ovulation à 9 mois, les taux plasmatiques de FSH (Folliculo Stimulating Hormone), le nombre d'embryons à 30 jours de gestation, les caractéristiques des tractus génitaux mâles (entre 160 et 180 jours) et femelles (à 30 jours de gestation). Les premières analyses de données indiquent la présence de QTL pour le taux d'ovulation et le nombre d'embryons sur les chromosomes 7 et 8 (Milan *et al* 1998).

Cette démarche d'identification de locus quantitatifs constitue aussi le point de départ du clonage positionnel qui vise à délimiter la zone la plus précisément possible pour y rechercher ensuite les gènes et, parmi ceux-ci, ceux qui présentent un polymorphisme associé au caractère. Basé uniquement sur des méthodes de biologie moléculaire, le clonage positionnel a permis d'identifier près d'une centaine de gènes, mais il demeure très long et complexe (Ballabio 1993).

### Identifier et caractériser les gènes impliqués dans la fonction

Cette approche, fonctionnelle, est basée sur les données de la physiologie, de la biochimie ou de la pathologie qui permettent de désigner le(s) lieu(x) d'expression des gènes recherchés et même, plus précisément, certains gènes importants dans la fonction considérée. Dans le cas de la composante folliculogénèse du taux d'ovulation, l'un des lieux d'expression est bien évidemment l'ovaire, et certains gènes comme ceux des enzymes de la stéroïdogénèse, des facteurs de croissance ou encore des récepteurs aux hormones gonadotropes, peuvent être considérés comme importants dans le développement du follicule ovarien. Il est évident qu'il est très difficile, à ce stade et sur ces seules bases, d'identifier les gènes responsables de la variabilité du taux d'ovulation.

Cette identification passe par l'établissement d'un répertoire plus ou moins ciblé des gènes exprimés dans un tissu ou un organe, ici le follicule, c'est-à-dire l'isolement et l'identification des différents transcrits présents dans les cellules folliculaires. Parmi les différentes stratégies d'isolement, certaines font

F. HATEY

INRA Laboratoire  
de Génétique  
Cellulaire, BP 27,  
31326 Castanet-  
Tolosan Cedex

appel à des génothèques d'ADN complémentaires (ADNc), d'autres ne passent pas par une étape de clonage préalable ; les unes analysent indistinctement tous les gènes, les autres utilisent différentes techniques pour sélectionner certains gènes en fonction d'hypothèses biologiques.

L'identification des gènes par séquençage et comparaison de séquences se fait sans chercher, au moins dans un premier temps, à obtenir la séquence complète de tout le transcrit. Au contraire, en particulier avec les génothèques, on ne déterminera que la séquence de l'extrémité des ADNc ; c'est la stratégie des « étiquettes » (Expressed Sequence Tags ou EST, Adams *et al* 1991).

### Analyse de l'ensemble des gènes

L'utilisation de génothèques d'ADNc est un moyen privilégié puisque, obtenues par transcription inverse des ARN messagers, elles sont un reflet du tissu d'origine ; il y a donc une bonne correspondance, qualitative et quantitative, entre les ARN messagers du tissu et les ADNc de la génothèque : l'analyse de clones prélevés au hasard constitue alors un moyen simple et efficace pour identifier les gènes exprimés dans un tissu. Cette approche va cependant se trouver limitée par le phénomène de redondance, c'est-à-dire la présence de très nombreuses copies des mêmes ARN messagers. En effet, au sein d'une cellule, tous les ARN messagers ne sont pas présents en un même nombre de copies, et si certains ARN messagers sont très abondants, d'autres peuvent être moins abondants ou rares. Des protocoles de « normalisation » ont été proposés pour réduire ces écarts d'abondance des ADNc, l'idéal étant d'obtenir une population dans laquelle tous les ADNc seraient présents en une seule copie. Ils sont basés sur des cycles de dénaturation/renaturation au cours desquels les espèces les plus abondantes se renaturent préférentiellement, ce qui conduit à un enrichissement de la fraction simple brin en ADNc rares.

D'autres techniques permettent d'obtenir directement un profil d'expression plus ou moins facile à analyser, comme la SAGE (Serial Analysis of Gene Expression, Velculescu *et al* 1995), le mRNA Differential Display (Liang et Pardee 1992) ou l'AFLP-cDNA (Amplified restriction Fragment Length Polymorphism, Bachem *et al* 1996). Ces deux dernières techniques sont bien adaptées à la recherche ciblée de gènes et seront envisagées dans le paragraphe suivant. Dans la SAGE, un court fragment est prélevé à un endroit précis des ADNc, les différents fragments sont assemblés entre eux, clonés et séquencés. La séquence de ces fragments et leur fréquence permettent de définir un profil d'expression.

### Recherche ciblée

En fonction d'hypothèses biologiques, différentes stratégies peuvent être utilisées pour rechercher des gènes spécifiques : par

exemple, compte tenu du rôle primordial de la FSH dans le développement folliculaire, les gènes régulés par cette hormone.

Pour sélectionner les ADNc régulés, différentes stratégies de tri peuvent être utilisées, comme le tri différentiel d'une génothèque, le mRNA Differential Display et l'AFLP-cDNA, ou encore différents protocoles de soustraction, tels le cDNA-RDA (Representational Difference Analysis, Hubank et Schatz 1994) et la PCR suppressive (Diatchenko *et al* 1996).

Dans le tri différentiel, deux répliques de la génothèque sont hybridées avec des sondes correspondant aux deux situations que l'on veut comparer. La différence d'intensité des signaux d'hybridation observés pour un clone donné indique qu'il contient un ADNc correspondant à un ARN messager régulé.

Dans le mRNA Differential Display et l'AFLP-cDNA, les ADNc sont séparés en différents sous-groupes de telle sorte que l'on puisse les individualiser par électrophorèse après amplification par PCR. Ici encore, la comparaison de l'intensité des bandes obtenues pour les ADNc correspondants dans les différentes situations analysées permet d'identifier les ARN messagers régulés.

Dans les techniques de soustraction, la population d'ARN messager « cible », contenant les transcrits spécifiques (ou les ADNc correspondants) est épuisée par hybridation avec un excès d'ARN messagers (ou d'ADNc) ne contenant pas ces transcrits. Par exemple, si l'on cherche les gènes dont l'expression est stimulée par la FSH dans les cellules de la granulosa en culture, les ARN extraits des cellules traitées par la FSH seront épuisés par ceux extraits de cellules non traitées. Différents protocoles permettent ensuite de séparer les séquences spécifiques et l'ARN entraîneur. Ainsi, dans le cas de la RDA et de la PCR suppressive, des adaptateurs sont ajoutés aux ADNc « cible » de sorte que, après soustraction, seuls les ADNc spécifiques recherchés peuvent être amplifiés par PCR.

La mise en œuvre de banques organisées permet une meilleure gestion des informations concernant les différents clones : ceux-ci sont déposés individuellement dans les puits de microplaques qui servent pour la conservation des clones et pour la réalisation de répliques à haute ou très haute densité, par le dépôt décalé de l'ensemble des clones de plusieurs plaques sur une même membrane. Ces membranes sont ensuite hybridées avec différentes sondes, de telle sorte qu'il est possible d'établir un profil d'hybridation pour chacun des clones. L'acquisition des données d'hybridation et leur traitement requièrent des moyens d'analyse particuliers (Lennon et Lehach 1991, Auffray *et al* 1995, Nguyen *et al* 1995). Grâce au développement de différentes technologies, cette approche peut être miniaturisée : des milliers ou des dizaines de milliers de séquences peuvent être déposées sur des surfaces de quelques cm<sup>2</sup>, permettant ainsi l'acquisition simultanée d'un très grand nombre de profils d'expression ; ce sont les «

microarrays », « puces ADN » ou « DNA chips » (Daignan-Fornier et Aigle 1998), en plein développement, qui vont certainement bouleverser l'étude de l'expression des gènes.

En utilisant comme modèle de folliculogénèse des cellules de granulosa en culture, nous avons recherché les gènes régulés par la FSH, en utilisant les techniques de tri différentiel et de Differential Display. Nous avons ainsi isolé près de 200 séquences uniques dont 70 correspondent à des séquences identifiées. Parmi celles-ci, le gène de la glutathion S-transférase dont l'expression présente des caractéristiques intéressantes et que nous prendrons comme modèle.

## Rechercher les co-localisations entre les régions et les gènes

La localisation chromosomique des gènes du répertoire est obtenue par cartographie cytogénétique ou par cartographie physique.

La cartographie cytogénétique est réalisée soit par hybridation *in situ* sur les chromosomes en métaphase, soit, de plus en plus souvent, en utilisant un panel d'hybrides somatiques. La mise en œuvre d'un tel panel est en effet beaucoup plus simple : elle consiste à rechercher une coségrégation entre les chromosomes ou les fragments de chromosome du porc et les produits d'amplification PCR spécifiques du gène que l'on cherche à localiser (Wilcox *et al* 1991). La résolution de la carte cytogénétique est de l'ordre de 10 mégabases (Mb).

En utilisant le panel d'hybrides développé à l'INRA (Yerle *et al* 1996), nous avons localisé une quarantaine de gènes, identifiés ou non, parmi les 200 isolés. A titre d'exemple, le gène codant pour la glutathion S-transférase de type alpha a été localisé sur le bras long du chromosome 7 porcin.

Une localisation régionale est généralement obtenue, mais peut ne pas être suffisante. Pour obtenir une localisation plus fine, l'utilisation d'un panel d'hybrides irradiés dans lequel les chromosomes du porc ont été fragmentés en morceaux plus petits par irradiation avant la fusion cellulaire est nécessaire. La résolution obtenue en utilisant les hybrides irradiés est de l'ordre de la centaine de kilobases (kb).

La cartographie physique utilise des fragments d'ADN génomique de plusieurs centaines de kilobases, clonés dans des chromosomes artificiels de levure (Yeast Artificial Chromosomes ou YAC) ou de bactéries (Bacterial Artificial Chromosomes ou BAC). Ces fragments de grande taille sont aisément localisés directement par hybridation *in situ* ; la résolution est de 150 kb à 1 Mb. La présence du gène analysé est recherchée dans ces fragments par amplification PCR à l'aide d'amorces spécifiques.

En absence de gènes localisés dans la zone d'intérêt, les données de cartographie compa-

rée peuvent permettre de proposer des candidats en recherchant, dans les espèces pour lesquelles les cartes génétiques sont bien développées, comme l'homme et la souris, quels sont les gènes qui sont localisés dans la région correspondante du génome. En effet, pour la quasi totalité des chromosomes du porc, les régions correspondantes sur les chromosomes humains ont été identifiées. Dans un deuxième temps, il faut vérifier si les gènes candidats ainsi désignés correspondent à une fonction qui peut rendre compte des différences dans l'expression du caractère.

Comme démonstration de l'intérêt de la cartographie comparée, rappelons que l'identification du gène de l'hyperthermie maligne au gène du récepteur de la ryanodine a d'abord été établie chez l'homme puis, par homologie, chez le porc. Pour apprécier le potentiel de cette cartographie comparée, les chiffres parlent d'eux-mêmes : le nombre de gènes localisés chez le porc était de 158 début 1997 (Yerle *et al* 1997) ; chez l'homme, les résultats accumulés par un consortium international ont permis la localisation de plus de 16 000 transcrits (Schuler *et al* 1996).

Au terme de cette étape, nous disposons donc de gènes candidats qui sont à la fois localisés dans la région du génome identifiée au cours de la recherche de QTL et impliqués dans l'expression du caractère.

## Rechercher le polymorphisme et le lien entre ce polymorphisme et la variabilité du caractère

La variabilité du caractère résultant de celle d'un ou plusieurs gènes, la dernière étape est la recherche du polymorphisme des gènes candidats, polymorphisme qui résulte de plusieurs phénomènes différents, tels que mutations ponctuelles, insertions ou délétions.

Différentes méthodes permettent de révéler ce polymorphisme, en particulier la recherche de polymorphisme de conformation et le séquençage.

Le polymorphisme de conformation est basé sur la différence de mobilité des fragments d'ADN en fonction de leur conformation tridimensionnelle, et donc de leur séquence. Les systèmes qui ont été développés permettent de détecter des différences minimales telles que les substitutions. Ce sont essentiellement la SSCA (Single Strand Conformation Analysis, Orita *et al* 1989) et la DGGE (Denaturing Gradient Gel Electrophoresis, Myers *et al* 1985). Dans l'une et l'autre technique, les différences de séquence entraînent des différences de conformation des brins d'ADN, différences qui sont révélées par électrophorèse où elles se traduisent par des différences de migration.

Ainsi, la simple analyse par SSCA d'un fragment de 220 paires de bases du gène de la glutathion S-transférase, amplifié par PCR à partir de l'ADN de 12 animaux non apparentés, révèle 6 profils différents. Il reste à déter-



miner l'existence d'un lien entre ce polymorphisme et le taux d'ovulation.

Par ailleurs, le séquençage direct de la zone concernée permet d'identifier la mutation à l'origine du polymorphisme et de pressentir le caractère « causal » de la mutation. Ceci est important puisque seule la mutation « causale » présente une liaison absolue avec le caractère.

Il reste enfin à établir sur des familles le lien entre ce polymorphisme et la variabilité du caractère, aussi bien au niveau de la séquence qu'à celui de l'expression du gène correspondant.

## Conclusion

Cette approche de candidat positionnel permet aujourd'hui d'identifier avec certitude le ou les gènes impliqués dans une fonction complexe telle que le taux d'ovulation. Elle n'en reste pas moins une approche longue, nécessitant plusieurs années de recherches, aussi bien dans les aspects génétiques que fonctionnels. Elle a fait ses preuves en pathologie humaine (Collins 1995), il est raisonnable d'espérer qu'il en sera de même pour l'amélioration génétique des espèces domestiques.

## Contributions

Les données expérimentales présentées ici ont été obtenues au laboratoire de Génétique Cellulaire de l'INRA (Toulouse), par les chercheurs des groupes « Carte du porc » : Joël Gellin, André Goureau, Yvette Lahbib-Mansais, Denis Milan, Annie Robic et Martine Yerle, « Génétique et Différenciation ovarienne » : Catherine Clouscard-Martinato, François Gasser, François Hatey, Philippe Mulsant et Gwenola Tosser-Klopp et « Biomathématiques » : Claude Chevalet et Magali San Cristobal-Gaudy.

Le programme de recherche de locus quantitatifs implique de nombreuses autres équipes de l'INRA, en particulier la Station de Génétique Quantitative et Appliquée (Jouy-en-Josas) et le domaine du Magneraud.

## Références

- Adams M.D., Kelley J.M., Gocayne J.D., Dubnick M., Polymeropoulos M.H., Xiao H., Merril C.R., Wu A., Olde B., Moreno R.F., *et al.*, 1991. Complementary DNA sequencing : Expressed Sequence Tags and human genome project. *Science*, 252, 1651-1656.
- Auffray C., Behar G., Bois F., Bouchier C., Da Silva C., Devignes M.D., Duprat S., Houlgatte R., Jumeau M.N., Lamy B., *et al.*, 1995. IMAGE : molecular integration of the analysis of the human genome and its expression. *C.R. Acad. Sci. III*, 318, 263-272.
- Bachem C.W., van der Hoeven R.S., de Bruijn S.M., Vreugdenhil D., Zabeau M., Visser R.G., 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP : analysis of gene expression during potato tuber development. *Plant J.*, 9, 745-753.
- Ballabio A., 1993. The rise and fall of positional cloning. *Nature Genet.*, 3, 277-279.
- Collins F.S., 1995. Positional cloning moves from perdictional to traditional. *Nature Genet.*, 9, 347-350.
- Daignan-Fornier B., Aigle M., 1998. Du génome au « transcriptome » : un saut de puce ! *Médecine/Sciences*, 14, 225-226.
- Diatchenko L., Lau Y.F., Campbell A.P., Chenchik A., Moqadam F., Huang B., Lukyanov S., Lukyanov K., Gurskaya N., Sverdlov E.D., *et al.*, 1996. Suppression subtractive hybridization : a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA*, 93, 6025-6030.
- Hubank M., Schatz D.G., 1994. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res.*, 22, 5640-5648.
- Lennon G.G., Lehrach H., 1991. Hybridization analyses of arrayed cDNA libraries. *Trends Genet.*, 7, 314-317.
- Liang P., Pardee A.B., 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 257, 967-971.
- Milan D., Bidanel J-P., Le Roy P., Chevalet C., Woloszyn N., Caritez J-C., Gruand J., Lagant H., Bonneau M., Le Faucheur L., *et al.*, 1998. Current status of QTL detection in Large White X Meishan crosses in France. 6th World Congress on Genetics Applied to Livestock Production, January 1998, Armidale, Australia.
- Myers R.M., Lumelsky N., Lerman L.S., Maniatis T., 1985. Detection of single base substitutions in total genomic DNA. *Nature*, 313, 495-498.
- Nguyen C., Rocha D., Granjeaud S., Baldit M., Bernard K., Naquet P., Jordan B.R., 1995. Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics*, 29, 207-216.
- Orita M., Suzuki Y., Sekiya T., Hayashi K., 1989. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics*, 5, 874-879.
- Schuler G.D., Boguski M.S., Stewart E.A., Stein L.D., Gyapay G., Rice K., White R.E., Rodriguez-Tome P., Aggarwal A., Bajorek E., *et al.*, 1996. A gene map of the human genome. *Science*, 274, 540-546.
- Velculescu V.E., Zhang L., Vogelstein B., Kinzler K.W., 1995. Serial analysis of gene expression. *Science*, 270, 484-487.
- Wilcox A.S., Khan A.S., Hopkins J.A., Sikela J.M., 1991. Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs : implications for an expression map of the genome. *Nucleic Acids Res.*, 19, 1837-1843.
- Yerle M., Echard G., Robic A., Mairal A., Dubut-Fontana C., Riquet J., Pinton P., Milan D., Lahbib-Mansais Y., Gellin J., 1996. A somatic cell hybrid panel for pig regional gene mapping characterized by molecular cytogenetics. *Cytogenet. Cell Genet.*, 73, 194-202.
- Yerle M., Lahbib-Mansais Y., Pinton P., Robic A., Goureau A., Milan D., Gellin J., 1997. The cytogenetic map of the domestic pig (*Sus scrofa domestica*). *Mamm. Genome*, 8, 592-607.