

Analyse du transcriptome ovarien chez le porc : identification de réseaux de gènes impliqués dans la folliculogénèse

Agnès BONNET (1), Kim-Anh LÊ CAO (2), Magali SANCRISTOBAL (1),
Gwenola TOSSER-KLOPP (1), François HATEY (1)

(1) INRA, Département de Génétique Animale, laboratoire de Génétique Cellulaire,
(2) INRA, Département de Génétique Animale, Station d'Amélioration Génétique des Animaux,
chemin de Borde Rouge, BP 52627, 31326 Castanet-Tolosan Cedex

agnes.bonnet@toulouse.inra.fr

INTRODUCTION

Basée sur l'analyse des ARN, cette étude a pour objectif d'établir un répertoire de gènes exprimés au cours du développement du follicule ovarien pour identifier et caractériser les gènes potentiellement impliqués dans le déterminisme génétique du taux d'ovulation et de la prolificité chez la truie.

1. MATÉRIELS ET MÉTHODES

1.1. Animaux

Les truies sont ovariectomisées 24 ou 96 h après l'arrêt du traitement Régumate (20 mg par jour pendant 18j). Les follicules sont disséqués, classés selon leur taille et triés selon leur état physiologique.

1.2. Extraction des ARN

Les ARN sont extraits des cellules de la granulosa par la méthode modifiée de Chomczynski et Sacchi (Hatey et al., 1995) en groupant les follicules par animal et par classe : petits (2 mm), moyens (3 mm) et gros follicules (5-6 mm). La qualité des ARN est contrôlée sur Bioanalyser Agilent 2100.

1.3. Hybridation Soustractive Suppressive (HSS)

Cette technique, qui allie soustraction et normalisation, permet d'isoler les gènes différentiellement exprimés entre deux situations biologiques. Les ADNc sont synthétisés et amplifiés à l'aide du kit SMARTTM PCR cDNA synthesis Kit (Clontech) et les HSS sont effectuées comme décrit par Bonnet et al. (2006).

1.4. Conception et hybridation des réseaux

L'expression des gènes est analysée avec un micro-réseau nylon d'ADNc. Les produits PCR déposés proviennent de

clones d'ADNc issus (i) de la technique HSS, (ii) de la banque multi-tissus de porc d'AGENAE et (iii) de gènes contrôle. Ce micro-réseau est décrit dans la base de données Gene Expression Omnibus (G.E.O.) sous le numéro GPL3978.

Les micro-réseaux sont hybridés avec les ADNc marqués provenant des ARN de cellules de granulosa des différentes classes de follicules selon le protocole décrit par le laboratoire des Technologies Avancées pour le Génome et la Clinique (INSERM-ERM).

1.5. Traitement des données

Les données issues du logiciel d'image BZSCAN (Lopez et al., 2004) sont transformées en logarithme puis normalisées par la moyenne de chaque membrane. Seuls les ADNc exprimés (> à la moyenne des vides plus 2 écarts types) sont conservés.

1.6. Analyses statistiques

Deux types d'analyses ont été menés avec le logiciel d'analyse R (R) :

Les ADNc différentiellement exprimés entre deux classes de follicules sont identifiés à l'aide d'un test de Fisher. Les p-values sont corrigées pour la multiplicité des tests et le nombre de faux positifs estimé (False Discovery Rate, FDR : Benjamini et Hochberg, 1995).

La recherche des gènes les plus pertinents dans la discrimination des différentes phases de la croissance folliculaire a été effectuée avec une analyse par les forêts aléatoires (Breiman, 2001 ; Baccini et al., 2005).

1.7. Analyse fonctionnelle

Cette analyse a été effectuée avec le logiciel Ingenuity Pathways Analysis (IPA) qui permet de mettre en évidence,

d'explorer et de visualiser les fonctions impliquées et les réseaux d'interactions biologiques d'une collection de gènes.

2. RÉSULTATS ET DISCUSSION

2.1. Construction des banques HSS

Quatre banques ont été construites pour identifier les ADNc spécifiquement surexprimés dans une classe de follicules. Pour éliminer les faux positifs, un tri a été effectué sur macro-réseaux de colonies et a permis de sélectionner 1697 ADNc. Leur séquençage a produit 1378 séquences valides publiées dans la banque de données EBI. L'assemblage des séquences avec toutes les séquences publiques effectué par Sigenae a identifié 345 contigs dont 65 nouveaux.

2.2. Analyse de l'expression par micro-réseaux

Pour étudier l'expression de ces gènes, nous avons construit un micro-réseau comprenant les 1697 ADNc précédents et 1097 ADNc d'une banque multi-tissus (AGENAE), ajoutés ici pour permettre une bonne normalisation.

2.3. Analyse statistique

Une analyse statistique utilisant le modèle linéaire mixte a permis de montrer que 82 % des variations peuvent s'expliquer par l'effet des gènes.

Nous avons sélectionné 643 ADNc différenciellement exprimés dans au moins une classe de taille de follicule avec le test de Fisher (FDR < 0,2 %). Le faible taux de faux positifs peut s'expliquer par la présence de nombreux gènes régulés sur le micro-réseau (technique HSS). Ces ADNc correspondent à 79 gènes et 8 contigs dont 3 spécifiques des banques HSS. La classification ascendante hiérarchique des ADNc et des hybridations montre une grande proportion de gènes surexprimés chez les gros follicules mais la sélection de gènes (test de Fisher) ne permet pas de différencier correctement les petits follicules des moyens.

L'analyse par les forêts aléatoires (RF), méthode non paramétrique qui prend en compte les corrélations entre gènes, nous a permis de sélectionner et de classer selon leur perti-

nence 120 ADNc donnant une bonne prédiction des classes de follicules. Ces ADNc correspondent à 25 gènes connus, 4 contigs dont l'un est spécifique des banques HSS. Quinze gènes ont été contrôlés en RT-PCR quantitative et leur régulation a été confirmée (p-value < 5 %).

2.4. Analyse fonctionnelle

Les gènes sélectionnés par la méthode des RF se répartissent dans deux réseaux de régulation. Le premier réseau regroupe des gènes réprimés au cours du développement. Ils interviennent essentiellement dans l'organisation cellulaire (assemblage et maintien). Le second réseau comprend à la fois des gènes réprimés et surexprimés au cours de ce processus et intervenant dans le métabolisme stéroïdien, la reproduction (développement et fonction) et la synthèse des protéines. Ces réseaux impliquent des gènes connus pour leur rôle important dans le développement folliculaire comme l'insuline, le récepteur d'IGF1 et l'aromatase.

Cette analyse permet d'interpréter l'implication de gènes nouveaux tels que l'ITM2A (integral membrane protein 2A) la calumenine (fixation du calcium), la smootheline (liaison à l'actine).

CONCLUSION

Ces résultats permettent de valider l'utilisation de réseaux d'ADNc comme approche globale du transcriptome. Les analyses ont identifié un nombre important de gènes dont l'expression varie au cours du développement du follicule. L'analyse des données par les forêts aléatoires a permis de mettre en évidence des gènes dont la régulation est importante pour ce processus. Le travail d'interprétation de ces informations est en cours.

REMERCIEMENTS

Les auteurs remercient le Centre de Ressources Génotypage Séquençage (Génopôle Midi-pyrénées) pour son aide dans le développement de l'outil d'analyse du transcriptome, Francis Benne, Janine Rallières et Katia Fève pour leur soutien technique et l'unité SENAH (Saint-Gilles, INRA) pour les animaux.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Baccini A., Besse P., Déjean S., Martin P.G.P., Robert-Granié C., SanCristobal M., 2005. Stratégies pour l'analyse statistique de données transcriptomiques. *J. Soc. Franç. Statist.*, 146, 5-44.
- Benjamini V., Hochberg V., 1995. Controlling the false discovery rate : a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B.*, 57, 289-300.
- Bonnet A., Frappart P.O., Dehais P., Tosser-Klopp G., Hatey F., 2006. Identification of differential gene expression in *in vitro* FSH treated pig granulosa cells using suppression subtractive hybridization. *Reprod. Biol. Endocrinol.*, 4, 35.
- Breiman L., 2001. Random forests. *Machine Learning*, 45, 5-22.
- G.E.O., www.ncbi.nlm.nih.gov/geo
- Hatey F., Mulsant P., Bonnet A., Benne F., Gasser F., 1995. Protein kinase C inhibition of *in vitro* FSH-induced differentiation in pig granulosa cells. *Mol. Cell. Endocrinol.*, 107, 9-16.
- INSERM-ERM, http://tagc.univ-mrs.fr/oncogenomics/Nylon_microarrays.php.
- IPA, www.ingenuity.com
- Lopez F., Rougemont J., Lloriod B., Bourgeois A., Loi L., Bertucci F., Hingamp P., Houlgatte R., Granjeaud S., 2004. Feature extraction and signal processing for nylon DNA microarrays. *BMC Genomics*, 5, 38.
- R, a language and environment for statistical computing, R Foundation for statistical computing, Vienna, Austria., <http://www.cran.r-project.org>.