

## **Le transcriptome épидидymaire du vertrat : étude de la régionalisation**

*Benoît GUYONNET (1,2), Jean-Louis DACHEUX (1), Florence JAFFREZIC (3), Anne LACOSTE (2), Guillemette MAROT (3),  
Marie-José MERCAT (2), Sandrine SCHWOB (2), Jean-Luc GATTI (1)*

*(1) UMR 6175 INRA-CNRS-Université de Tours-Haras Nationaux  
Physiologie de la Reproduction et des Comportements, 37380 Nouzilly*

*(2) IFIP-Institut du porc, La Motte au Vicomte, 35650 Le Rheu*

*(3) Unité de Génétique Quantitative et Appliquée, INRA, 78352 Jouy en Josas*

*gatti@tours.inra.fr*

### **Le transcriptome épидидymaire du vertrat : étude de la régionalisation**

Chez les mammifères, les spermatozoïdes acquièrent leur pouvoir fécondant au cours du passage dans l'épididyme, organe constitué d'un tube pelotonné et relié au testicule par le canal efférent. L'épididyme est divisé en 3 régions morphologiques (tête, corps, queue) subdivisées en 10 zones.

A l'aide d'une puce à ADN porcine (programme AGENAE, GEO : GPL3729), nous avons entrepris l'analyse du transcriptome épидидymaire chez 4 verrats Large White adultes, sélectionnés sur leur qualité de reproducteur. Pour chaque animal, nous avons hybridé les ADNc dérivés d'ARN extraits du testicule, des canaux efférents et déférents et de 9 zones épидидymaires. Les données obtenues ont été analysées selon deux méthodes statistiques non supervisées pour la recherche de classes : une méthode de partitionnement (PAM) et une méthode de classification hiérarchique (HCL). Seuls les gènes classés de la même manière par les deux types d'analyse ont été gardés pour les analyses fonctionnelles.

Cette approche a démontré l'existence de 5 régions transcriptomiques au niveau épидидymaire et de 3 classes de gènes ayant des expressions préférentielles au niveau de certaines de ces régions.

### **The boar epididymal transcriptome: clustering of intraregional segments**

In mammals, both testis and epididymis contribute to the formation of fertile spermatozoa. Gametes production takes place in the testis, and when they exit this organ spermatozoa have a specialized and distinct morphology but are infertile. It is only after their travel in the epididymis that sperm become fertile. Epididymis is an organ formed by a long and unique convoluted tubule that can be divided in 3 morphological regions (head, body and tail) subdivided in 10 segments.

In this study, tissues of four Large White adult boars including testis, efferent ducts, nine morphologically distinct segments of the epididymis and deferent duct were isolated and their mRNA extracted. The gene expression of each of these samples was analyzed using a pig generic microarray (AGENAE program; GEO accession number: GPL3729). Acquired data were analysed using two unsupervised statistical methods for class discovery: a method of partitioning (PAM) and a method of hierarchical clustering (HCL). Only genes classified in an identical manner by the two methods were kept for the functional analysis.

This study showed the existence of 5 distinct epididymal transcriptomic regions whereas 3 clusters of genes with an expression pattern related to the transcriptomic regions were also identified.

## INTRODUCTION

Chez les mammifères, la production de gamètes mâles fertiles fait intervenir deux organes qui sont le testicule et l'épididyme reliés par les canaux efférents. Au sein du testicule se déroule la spermatogenèse qui à partir d'une cellule souche indifférenciée aboutit à la formation de spermatozoïdes immatures. À la sortie du testicule, les spermatozoïdes, bien que morphologiquement formés, sont immobiles et incapables de féconder un ovocyte. C'est au cours de leur transit dans l'épididyme, tube pelotonné d'environ 70 m de long chez le verrat, que les spermatozoïdes vont acquérir leur mobilité et leur pouvoir fécondant. Au sein de l'épididyme, les gamètes sont environnés d'un fluide dont la composition varie en fonction des activités de sécrétion et de réabsorption de l'épithélium. La maturation des spermatozoïdes résulte de leurs interactions avec les composés de ce milieu et fait intervenir divers remaniements protéiques de surface (Dacheux et al., 2005; Gatti et al., 2004). L'épididyme intervient donc dans la qualité du gamète.

Des études d'analyse transcriptomique ont montré que chez la souris (Johnston et al., 2005) et chez le rat (Jelinsky et al., 2007) les différentes zones épididymaires pouvaient être regroupées en fonction des données d'expression de gènes. Ces associations de zones font apparaître des régions transcriptomiques au sein desquelles des gènes présentent des profils d'expression particuliers.

Une étude réalisée au laboratoire portant sur l'analyse protéomique du sécrétome et de la composition protéique du fluide épididymaire chez le verrat a montré l'existence d'une régionalisation de la sécrétion (Syntin et al., 1996).

A l'aide d'une puce à ADN porcine (GEO : GPL3729), nous avons cherché à identifier les gènes présentant des expressions différentielles importantes au niveau de l'épididyme. En adaptant des méthodes d'analyse statistique non supervisée, les données obtenues pour chacune des zones épididymaires nous ont permis d'associer des zones ayant des profils d'expression similaires et de mettre en évidence des groupes de gènes présentant des expressions préférentielles au niveau de ces regroupements de zones.

## 1. MATÉRIELS ET MÉTHODES

### 1.1. Animaux

Cette étude porte sur 4 verrats Large White, âgés de plus d'un an et demi provenant de centres d'insémination artificielle et sélectionnés sur leur fertilité (recherche d'animaux de fertilité élevée). Pour chaque animal, différents critères de mesure de la qualité de la semence (volume des éjaculats, concentration en spermatozoïdes...) ont été suivis tout au long de leur carrière de reproducteur. Les verrats sélectionnés présentaient au moment de leur abattage une qualité de semence normale (semence commercialisable par les centres d'insémination artificielle). La fertilité de ces animaux, calculée à partir des résultats d'au moins 150 inséminations artificielles, est comprise entre 88 et 94%. Après abattage, des échantillons de tissu testiculaire, des canaux efférents, de 9 différentes zones épididymaires (0,1,2,3,4,5,6,7,8/9) (Figure 1) et des canaux déférents sont prélevés et congelés à -80°C. Au niveau épididymaire, des critères morphologiques

permettent d'isoler la tête constituée des zones 0 à 4, le corps formé des zones 5 à 7 et la queue regroupant les zones 8 et 9.



**Figure 1 - Epididyme de verrat (taille réelle : 30 cm).**

VE : vas efférent ;  
E0 à E4 : tête de l'épididyme ;  
E5 à E7 : corps de l'épididyme ;  
E8/9 : queue de l'épididyme ;  
VD : vas déférent.

### 1.2. Extraction d'ARN

Les ARN messagers (ARNm) des tissus sont isolés après broyage (Ultra-Turax) par une double extraction phénol chloroforme (TRI Reagent, Sigma). La qualité des ARNm obtenus est estimée par le rapport 260/280 nm et séparation sur gel d'électrophorèse (Bioanalyseur, Agilent technologies). 5 µg d'ARN totaux ont servi à la synthèse des ADNc marqués au <sup>33</sup>P par rétro-transcription.

### 1.3. Hybridation des puces

Les membranes de nylon utilisées (7,5 x 2,5 cm) comportent 9216 dépôts dont 8959 proviennent de l'amplification des clones obtenus lors du programme AGENAE. Pour chaque clone, le fragment d'ADN porcine qu'il contient est amplifié par PCR avant d'être déposé sur les membranes. Les 257 dépôts restants correspondent à des témoins positifs (luciférase) et négatifs (eau).

Une première hybridation (hybridation vecteur) est réalisée en utilisant un oligonucléotide d'une séquence présente aux extrémités de chacun des produits de PCR déposés et marqué au <sup>33</sup>P, ceci afin d'estimer la quantité d'ADN présente pour chaque dépôt. Après déshybridation, les ADNc tissulaires marqués sont hybridés pendant 16 heures. Ensuite, les membranes sont rincées et exposées sur écran « phosphore » 24 heures avant d'être digitalisées (Fuji-Bas 5000). Le niveau d'intensité de chaque dépôt est quantifié par le logiciel AGScan (Cathelin et al., 2007).

### 1.4. Analyses des données

Les étapes de nettoyage, filtrage, normalisation et analyses statistiques sont réalisées en utilisant le logiciel R (R Development Core Team, 2006).

Les gènes différentiellement exprimés sont déterminés à l'aide d'un test F modifié issu de la librairie Limma (Smyth et al., 2003). Dans l'analyse, l'effet animal est pris en compte afin de corriger la dépendance entre les données issues d'un même animal. La méthode de Benjamini et Hochberg (1995) est utilisée afin de corriger pour les tests multiples la liste de gènes différentiels.

Pour toutes les analyses de classification (clustering), la distance basée sur la valeur «1-corrélation» est utilisée afin de séparer

les gènes corrélés négativement et d'associer ceux corrélés positivement. Nos résultats ont été analysés par deux méthodes : (1) une méthode de partitionnement (PAM : Partitioning Around Medoids) et (2) une méthode de classification hiérarchique (CAH : Classification Hiérarchique Ascendante et HCL : Hierarchical Clustering).

La méthode de classification hiérarchique est une méthode non supervisée (c'est-à-dire que l'analyse des données se fait sans a priori sur le nombre de classes attendues) alors que la PAM est plus connue comme une méthode supervisée. Dans notre cas, la PAM est utilisée comme méthode non supervisée en lui attribuant une gamme de nombre de classes au lieu de lui imposer un nombre fixe. De cette manière, la PAM nous permet de définir le nombre de classes approprié et ce, sans a priori.

Pour les méthodes de classification hiérarchique, le critère d'agglomération Ward communément admis par les statisticiens (Seber, 1984) est utilisé. La méthode HCL réalise une double classification sur les gènes et les zones, en construisant deux dendrogrammes dépendants, un pour les gènes et un pour les zones ; seul celui des gènes sera étudié.

## 2. RESULTATS

### 2.1. Traitement des données

Les données sont filtrées pour éviter les artefacts et les dépôts de mauvaise qualité sont éliminés lorsque les intensités mesurées pour l'hybridation avec le vecteur sont inférieures au seuil défini par la moyenne des intensités des dépôts témoins négatifs plus trois fois l'écart type (calculé individuellement pour chaque membrane). Seuls les dépôts avec au minimum 3 réplicats sur 4 utilisables sont considérés afin d'estimer la variabilité intra groupe, soit 7676 dépôts pour chacun des 12 tissus étudiés.

### 2.2. Recherche de gènes différentiels

2115 transcrits ont été trouvés différentiellement exprimés entre nos 12 prélèvements (testicule, canal éférent, zones épididymaires 0 à 8/9, canal déférent) en utilisant un taux de faux positif (FDR, pour « false discovery rate ») inférieur à 1/100000.

### 2.3. Regroupement de zones : régionalisation

Un regroupement des zones par utilisation combinée d'une PAM et d'une CAH a été réalisé à partir de cette liste de gènes différentiels. Cette démarche permet d'obtenir un graphique issu de la PAM, aidant à choisir le nombre de classes, étape critique dans les approches classiques des méthodes non supervisées (Figure 2).

Sur la figure 2, la qualité de classification des zones (mesurée par la silhouette moyenne) est représentée en fonction du nombre de classes choisies. Les valeurs de silhouette moyenne élevées correspondent à des classifications de bonne qualité. Le nombre de classes retenu est celui correspondant à une classification de bonne qualité permettant d'obtenir le plus d'informations physiologiques.

Le regroupement donnant le plus fort score (silhouette moyenne) correspond à la distinction du testicule par rapport aux

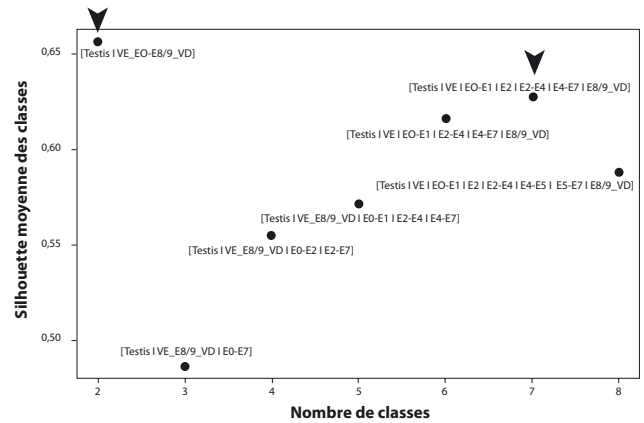


Figure 2 - Représentation de la qualité de la classification des zones en fonction du nombre de régions [Répartition des zones au sein des classes ou régions]

autres tissus mais cette classification est peu informative. Par contre l'association correspondant au second meilleur score de qualité permet de distinguer sept régions tissulaires. Cette classification a été retenue car elle donne beaucoup plus d'informations physiologiques.

L'interprétation du dendrogramme obtenu par la CAH permet également de classer en sept groupes les différents tissus, ce classement étant identique à celui obtenu par la PAM.

Par ces deux approches, les sept associations correspondent (i) au testicule, (ii) aux canaux éférents, (iii) à la tête antérieure de l'épididyme (zones 0-1), (iv) à la tête médiane (zone 2), (v) à la tête postérieure (zones 3-4), (vi) au corps (zones 4 à 7) et (vii) à la région caudale (zone 8/9) regroupée avec le canal déférent.

### 2.4. Recherche de gènes coréglés

La classification des gènes présentant des profils d'expression similaires est réalisée en utilisant une méthode de partitionnement (PAM) et une méthode de classification hiérarchique (HCL).

La PAM permet de définir le nombre de classes au sein desquelles les gènes sont le mieux représentés (Figure 3).

La meilleure classification est constituée des gènes testiculaires versus les gènes épididymaires. La classification avec le deuxième meilleur score permet d'obtenir quatre classes et d'avoir

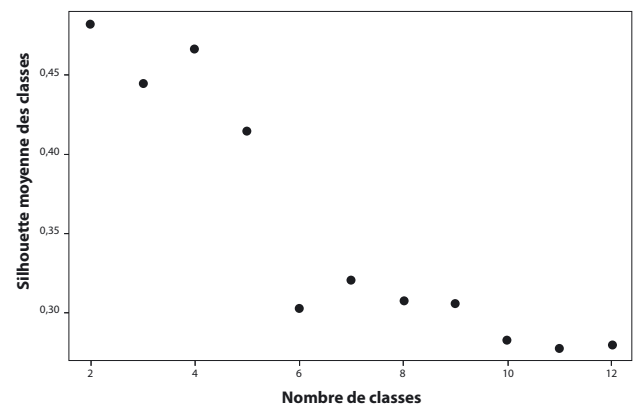


Figure 3 - Représentation de la qualité de la classification des gènes en fonction du nombre de classes

plus d'informations sur les groupes de gènes. Ces quatre classes regroupent des gènes qui ont une expression préférentielle au niveau a) du testicule, b) de la tête, c) du corps, d) du canal éférent et de la queue de l'épididyme (E8/9 et VD). Pour chacun des groupes de gènes, les identifiants des transcrits sont récupérés.

L'interprétation du dendrogramme obtenu par l'HCL permet également de classer en quatre groupes les gènes pour lesquels les identifiants sont recueillis.

Une table de contingence construite à partir des différentes listes de gènes a permis d'établir la liste des gènes regroupés au sein d'une même classe par les deux approches statistiques (Tableau 1).

**Tableau 1 - Table de contingence des identifiants contenus dans chacune des classes de gènes obtenues par les méthodes PAM et HCL**

		Classes PAM			
		1	2	3	4
Classes HCL	1	122	62	65	453
	2	45	319	2	3
	3	5	14	774	1
	4	226	9	9	6

**Tableau 2 - Dépouillement de la table de contingence (numéros et effectifs des classes)**

	HCL	PAM	Transcrits communs
<b>Classe «Testicule»</b>	3 (794)	3 (850)	774
<b>Classe «Tête»</b>	4 (250)	1 (398)	226
<b>Classe «Corps»</b>	1 (702)	4 (463)	453
<b>Classe «VE/Queue»</b>	2 (369)	2 (404)	319

Sur les 2115 transcrits différenciellement exprimés entre les 12 conditions, 1772 se regroupent au sein de quatre classes avec les deux approches statistiques utilisées. Parmi eux, 774 transcrits présentent une expression préférentielle au niveau du testicule, 226 au niveau de la tête de l'épididyme, 453 au niveau du corps et 319 au niveau des canaux éférents et de la queue de l'épididyme (zone 8/9 et canal déférent).

### 3. DISCUSSION

Nous avons développé une approche statistique consistant en l'utilisation combinée de méthodes non supervisées, une méthode de partitionnement (la PAM) et deux méthodes de classification hiérarchique (la CAH et la HCL). Cette dernière a permis d'obtenir des regroupements de zones et des classes de gènes avec un degré de confiance important. Cette approche permet d'éliminer les objets (zones ou gènes) qui sont mal représentés dans les groupes ou classes définis et qui, par la suite, sont néfastes aux étapes d'interprétation et de validation des résultats. Les objets retenus, les mieux classés, serviront pour les études fonctionnelles.

L'utilisation de puces génériques porcines a permis de mettre en évidence 2115 transcrits présentant une expression différentielle entre les 12 régions du tractus génital mâle étudiées avec un taux de faux positif extrêmement faible (FDR=1/100000). Les tissus des testicules, des canaux éférents et de l'épididyme sont très différents au niveau transcriptionnel. Au niveau épididymaire, les 9 zones épididymaires étudiées se regroupent en 5 régions transcriptomiques et le canal déférent s'associe avec la zone 8/9 de l'épididyme. Si l'on compare cette régionalisation à celle de Syntin et al. (1996) obtenue par protéomique chez le porc, on note des différences au niveau de la tête et du corps de l'épididyme. En effet, ils montrent l'existence de cinq régions avec deux régions au niveau de la tête et du corps et une région au niveau de la queue. La comparaison de nos résultats avec ceux obtenus chez la souris et le rat montre que, dans ces trois espèces de mammifères, la tête de l'épididyme est la région morphologique au sein de laquelle les profils d'expression des gènes induisent la création du plus grand nombre de régions transcriptomiques. En d'autres termes, c'est au niveau de la tête de l'épididyme que les cellules épithéliales présentent le plus de variations au niveau du nombre et de l'intensité des gènes exprimés. En effet, chez le verrot, les 4 zones de la tête sont regroupées en 3 régions transcriptomiques. Johnston et al. (2005) montrent chez la souris que les 5 zones de la tête forment 4 régions et, chez le rat, Jelinski et al. (2007) regroupent en 4 régions les 9 zones de la tête de l'épididyme.

La recherche de classes de gènes à partir des 2115 transcrits différentiels a permis d'identifier 4 classes regroupant 1772 transcrits. Ces gènes présentent des profils d'expression préférentielle au niveau du testicule, de la tête, du corps, du canal éférent et de la queue de l'épididyme. L'annotation en cours grâce au logiciel Blast2GO (Conesa et al., 2005) va nous permettre d'acquérir des informations concernant les processus biologiques, les fonctions moléculaires et la localisation cellulaire des produits de ces gènes. Ces informations seront ensuite utilisées afin de mettre en évidence des fonctions majoritaires de chacune des régions transcriptomiques.

### CONCLUSION

Ces premiers résultats permettent de valider notre approche statistique, laquelle sera utilisée pour l'analyse des données issues de l'étude du transcriptome épididymaire au cours du développement post-natal. La synthèse de ces analyses devrait nous permettre de cibler des régions et des gènes d'intérêt pour l'acquisition de la fertilité par les spermatozoïdes. Ces hypothèses seront testées lors de notre étude du transcriptome épididymaire chez des verrats de fertilité contrastée.

### REMERCIEMENTS

Ce travail bénéficie d'un soutien du programme ANR GENANIMAL (CNRG A 02085 et FNS 03 2 109) et d'une bourse CIFRE dans le cadre du projet BIOPORC-IFIP. Le CRB GADIE (INRA, Jouy-en-Josas) est remercié pour la fourniture des puces à ADN ainsi qu'A. Le Cam et J. Monfort du plateau technique « puce à ADN » de Rennes. L'équipe SIGENAE (INRA, Toulouse) est remerciée pour le support bioinformatique. Nous remercions aussi fortement les Centres d'Insémination Artificielle et les Organisations de Sélection Porcine pour la fourniture des verrats pour cette étude.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- Benjamini Y., Hochberg Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc. B*, 85, 289-300.
- Cathelin R., Lopez F., Klopp C., 2007. AGScan: a pluggable microarray image quantification software based on the ImageJ library. *Bioinformatics*, 23, 247-248.
- Conesa A., Gotz S., Garcia-Gomez J. M., Terol J., Talon M., Robles M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21, 3674-3676.
- Dacheux J. L., Castella S., Gatti J. L., Dacheux F., 2005. Epididymal cell secretory activities and the role of proteins in boar sperm maturation. *Theriogenology*, 63, 319-341.
- Gatti J. L., Castella S., Dacheux F., Ecroyd H., Metayer S., Thimon V., Dacheux J. L., 2004. Post-testicular sperm environment and fertility. *Anim. Reprod. Sci.*, 82-83, 321-339.
- Jelinsky S. A., Turner T. T., Bang H. J., Finger J. N., Solarz M. K., Wilson E., Brown E. L., Kopf G. S., Johnston D. S., 2007. The rat epididymal transcriptome: comparison of segmental gene expression in the rat and mouse epididymides. *Biol. Reprod.*, 76, 561-570.
- Johnston D. S., Jelinsky S. A., Bang H. J., Dicaneloro P., Wilson E., Kopf G. S., Turner T. T., 2005. The mouse epididymal transcriptome: transcriptional profiling of segmental gene expression in the epididymis. *Biol. Reprod.*, 73, 404-413.
- R Development Core Team, 2006. R: A language and environment for statistical computing., R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Seber G.A.F., 1984. *Multivariate observations*, John Wiley & Sons Inc, New-York.
- Smyth G. K., Yang Y. H., Speed T., 2003. Statistical issues in cDNA microarray data analysis. *Methods Mol. Biol.*, 224, 111-136.
- Syntin P., Dacheux F., Druart X., Gatti J. L., Okamura N., Dacheux, J. L., 1996. Characterization and identification of proteins secreted in the various regions of the adult boar epididymis. *Biol. Reprod.*, 55, 956-974.