

Construction d'un modèle de prédiction de la force de cisaillement de la viande de porc cuite à partir de données d'expression génique

Grégory GUERNEC (1), Marie DAMON (2), Annie VINCENT (2), Frédéric HERAULT (3), Pierre CHEREL (4) et Pascale LE ROY (3)

(1) INRA, UR1037 SCRIBE, 35000 Rennes, France

(2) INRA, UMR SENAH, 35590 Saint-Gilles, France

(3) INRA, UMR Génétique Animale, 65 rue de St Brieuc, 35000 Rennes, France

(4) Hendrix Genetics, 100 avenue Denis Papin, 45800 St Jean de Braye, France

Gregory.Guernec@rennes.inra.fr

Avec la collaboration technique de Jérôme GLENISSON (4) et José PIRES (4)

Construction of a predictive model of shear force of cooked pig meat using high-throughput expression data

Pork quality is highly variable and the availability of relevant predictors of fresh meat quality at slaughter is critical to optimize carcass use and improve its valuation. High throughput gene expression studies have been widely used to describe biological mechanisms underlying variation of several traits but data are scarce regarding their use as tools for the development of biomarkers. In this study, exhaustive transcriptomics analysis of 300 Longissimus muscles (collected at slaughter) was achieved by microarray hybridization, using a custom repertoire of 15,000 60-mer DNA probes, targeting transcripts expressed in pig skeletal muscle. This study describes the method applied for selecting informative and robust predictors of shear force of cooked meat using mRNA levels. In particular, the minimal number of predictors included in the model was defined using a statistical framework for prediction. Finally, an optimal model with 7 predictors was established but will have to be validated using qPCR before it can be considered an effective tool to predict pork tenderness.

INTRODUCTION

Le tri et un classement des carcasses de porc demeurent nécessaires pour mieux maîtriser la variabilité de la qualité de viande et optimiser l'utilisation des carcasses et des pièces de découpe par les abattoirs et les transformateurs. Cela implique (1) une évaluation précoce de défauts qui sont visibles uniquement après découpe ou transformation et (2) de disposer d'outils de prédiction faciles à mettre en œuvre.

La tendreté est un critère important pour les consommateurs (Maltin *et al.*, 2003).

Or, si des relations ont été établies entre des profils d'expression génique (Lobjois *et al.*, 2008) ou protéique (Laville *et al.*, 2007) et la tendreté estimée par la mesure expérimentale de la force de cisaillement de la viande cuite (WBcuit), elle demeure difficile à prédire du fait d'interactions complexes entre les différentes composantes du tissu musculaire.

La méthodologie présentée vise à prédire au mieux le WBcuit en tirant profit de la mesure simultanée de l'expression d'un grand nombre de gènes (technologie microarray).

1. MATERIEL ET METHODES

1.1. Animaux, prélèvements et mesure du WBcuit

Les 325 animaux de cette étude sont issus d'une population F2 constituée à partir de deux lignées mâles de FRANCE-HYBRIDES (FH016, lignée de type Piétrain, FH019, lignée synthétique de type Hampshire, Duroc, et Large White) et indemnes des allèles HAL n ou RN⁺.

Dix grammes de muscle *Longissimus* sont prélevés 15 minutes après l'abattage puis congelés. Ces échantillons sont cuits pendant 30 min (chaleur sèche, au four) à 240°C. La force de cisaillement (WB) est estimée par la force maximale à la rupture avec une lame de Warner-Bratzler (Lobjois *et al.*, 2008).

1.2. Acquisition et traitement des données d'expression

Les ARN totaux sont extraits au Trizol puis les échantillons sont marqués au Cy3 avec le kit Low RNA Input Linear Amplification (Agilent) et hybridés une nuit à 65°C sur une puce à ADN 15K muscle spécifique (Agilent). L'acquisition des images est réalisée en utilisant le scanner Agilent G2505B.

Les quantifications des signaux brutes sont obtenues à l'aide du logiciel Feature Extraction Software (Agilent v9.5).

Un filtrage sur l'intensité des spots et une détection des valeurs aberrantes par la méthode des écart-réduits sont effectués. Les données manquantes sont imputées par la méthode des k plus proches voisins. Après élimination des gènes de faible variance (agrégation par k-means) et des gènes sans annotation valide (homologie de séquence $\leq 85\%$ et/ou moins de 18 pb consécutives), 3371 gènes sont conservés. Après vérification de la symétrie de la distribution du WBcuit, 276 échantillons sont conservés.

1.3. Construction du modèle de prédiction

Un modèle prédictif sans biais est établi au moyen d'une validation croisée par k découpages aléatoires du fichier de départ en échantillons d'apprentissage (participant à la construction du modèle) et de test (participant à sa validation). La recherche du modèle optimal est effectuée par comparaison de plusieurs modèles statistiques.

Le problème de la grande dimensionnalité (p gènes variables >> n lames individus) spécifique aux données de transcriptomique exclut l'utilisation directe des techniques classiques de régression. Deux autres approches ont donc été choisies : Partial Least Square (PLS, Wold *et al.*, 1983) et Random Forest (RF, Breiman 2001). La méthode PLS conserve l'hypothèse de linéarité et vise à maximiser conjointement la variance des prédicteurs et la corrélation existant entre ces derniers et la variable à prédire.

Cette approche a de plus été choisie pour déterminer dans un premier temps le nombre optimal de prédicteurs par minimisation de l'erreur test. L'analyse RF est basée sur l'agrégation d'arbres binaires de décision et privilégie l'hypothèse de non linéarité. Les gènes candidats résultant de la réunion des prédicteurs sélectionnés par les 2 approches sont inclus dans un modèle linéaire.

Trois modèles sont ensuite établis par minimisation de 3 critères d'optimalité distincts : AIC, BIC et erreur Test. Ces modèles sont évalués en fonction de leur capacité à minimiser les erreurs Test, d'Apprentissage et relative aux lames atypiques et comparés à des modèles faisant intervenir des mesures de qualité de viande classiquement déterminées à l'abattoir (modèles phénotypiques).

2. RESULTATS ET DISCUSSION

2.1. Nombre de variables prédictives

Seize échantillons sont identifiés comme atypiques. Les 260 lames restantes sont aléatoirement réparties en 2 sous-groupes : test (35%) et apprentissage (65%).

Cette opération réitérée 20 fois permet une estimation sans biais des erreurs associées aux modèles. L'approche PLS suggère un minimum en test atteint pour N=3 prédicteurs. Les 20 modèles PLS et RF générés fournissent alors une liste de 67 prédicteurs potentiels à affiner.

2.2. Modèle optimal de prédiction du WBcuit

Les critères d'optimalité de 3 modèles linéaires sont comparés (Figure 1) :

-le modèle 1 (31 prédicteurs) minimise l'AIC

-le modèle 2 (14 prédicteurs) minimise le BIC

-le modèle 3 (7 prédicteurs) minimise l'erreur Test.

Puis, les performances du modèle 3 choisi (R^2 ajusté=0,19) sont comparées à celles de 2 modèles phénotypiques sélectionnés à 4 (modèle 4) et 2 (modèle 5) prédicteurs. Le modèle 4 est établi à partir des mesures de la couleur rouge (a^*), du pH 30 min, du potentiel glycolytique et du taux de muscle. Le modèle 5 est établi à partir des mesures de la couleur rouge (a^*) et du potentiel glycolytique. L'estimation des erreurs associées montre que le modèle 3 est le seul modèle « génotypique » (modèles 1, 2 et 3) à proposer une erreur test comparable à celles observées avec les modèles « phénotypiques » (modèles 4 et 5).

Il semble donc le plus intéressant en termes de prédiction. Un tel modèle pourrait par exemple inclure la mesure des transcrits des gènes SEPT7, ST3GAL6, RARB, NRG4, DISP1 B3GNT5 et DCXR.

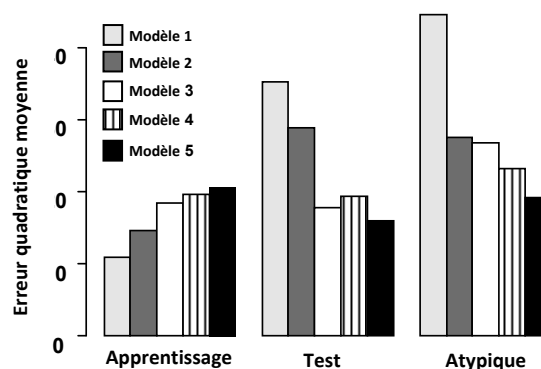


Figure 1 - Caractéristiques des modèles proposés

CONCLUSION

Après validation par qPCR, la combinaison à 7 prédicteurs pourrait constituer un outil de prédiction du WBcuit, mesure instrumentale de la tendreté de la viande.

Cet outil reste perfectible en robustesse mais serait pertinent pour les acteurs de la filière, car utilisable au stade du classement des carcasses à l'abattage

REMERCIEMENTS

Cette étude a été réalisée dans le cadre du projet GENMASCO, financé par l'Agence Nationale pour la Recherche dans le Programme National de Recherches en Alimentation, édition 2006.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Breiman L., 2001. Random Forests. *Machine Learning* 45, 5-32.
- Laville E., Sayd T., Terlouw C., Chambon C., Damon M., Larzul C., Leroy P., Glénisson J., Chérel P., 2007. Comparison of sarcoplasmic proteomes between two groups of pig muscles selected for shear force of cooked meat. *J. Agric. Food Chem.* 55, 5834-5841.
- Lobjois V., Liaubet L., SanCristobal M., Glénisson J., Fève K., Rallières J., Le Roy P., Milan D., Chérel P., Hatey F., 2008. A muscle transcriptome analysis identifies positional candidate genes for a complex trait in pig. *Anim. Genet.* 39, 147-162.
- Maltin C., Balcerzak D., Tilley R., Delday M., 2003. Determinants of meat quality: tenderness. *Proc. Nutr. Soc.*, 62, 337-347.
- Wold S., Martens H., Wold H., 1983. The multivariate calibration problem in chemistry solved by the PLS method. In: *Proc. Conf. Matrix Pencils Lecture Notes in Mathematics* 973, 286-293.