Animal (2013), **7:3**, pp 378–385 © The Animal Consortium 2012
doi:10.1017/S1751731112001747

animal

Variability-specific differential gene expression across reproductive stages in sows

J. Casellas[1,2†], M. Martínez-Giner[1], R. N. Pena[1], I. Balcells[2], A. Fernández-Rodríguez[3], N. Ibáñez-Escriche[1] and J. L. Noguera[1]

[1]*Genètica i Millora Animal, IRTA-Lleida, 25198 Lleida, Spain;* [2]*Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain;* [3]*Departamento de Reproducción Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, 28080 Madrid, Spain*

(Received 14 February 2012; Accepted 20 June 2012; First published online 11 October 2012)

Differential gene expression analyses typically focus on departures across mathematical expectations (i.e. mean) from two or more groups of microarrays, without considering alternative patterns of departure. Nevertheless, recent studies in humans and great apes have suggested that differential gene expression could also be characterized in terms of heterogeneous dispersion patterns. This must be viewed as a very interesting genetic phenomenon clearly linked to the regulation mechanisms of gene transcription. Unfortunately, we completely lack information about the incidence and relevance of dispersion-specific differential gene expression in livestock species, although a specific Bayes factor (BF) for testing this kind of differential gene expression (i.e. within-probe heteroskedasticity) has been recently developed. Within this context, our main objective was to characterize the incidence of dispersion-specific differential gene expression in pigs and, if possible, providing the first evidence of this phenomenon in a livestock species. We evaluated dispersion-specific differential gene expression on ovary, uterus and hypophysis samples from 22 F_2 Iberian $\times$ Meishan sows, where a total of 15 252 probes were interrogated. For each tissue, heteroskedasticity of probe-specific residual variances was evaluated by three pairwise comparisons involving three physiological stages, that is, heat, 15 days of pregnancy and 45 days of pregnancy. Between 2.9% and 37.4% of the analyzed probes provided statistical evidence of within-tissue across-physiological stages dispersion-specific differential gene expression (BF $>$1), and between 0.1% and 3.0% of them reported decisive evidence (BF $>$100). It is important to highlight that $<$8% of the heteroskedastic probes were also linked to differential gene expression in terms of departures among the probe-specific mathematical expectation of each physiological stage. This discarded the disturbance of scale effects in a high percentage of probes and suggested that probe-specific heteroskedasticity must be viewed as an independent phenomenon within the context of differential gene expression. As a whole, our results report a remarkable incidence of dispersion-specific differential gene expression across the whole genome of the pig, establishing a very interesting starting point for further studies focused on deciphering the genetic mechanisms underlying heteroskedasticity.

Keywords: Bayes factor, gene expression, heteroskedasticity, microarray, sow reproduction

Implications

Dispersion-specific differential gene expression across physiological stages has been evaluated on ovary, uterus and hypophysis from 22 F_2 Iberian $\times$ Meishan sows. The heteroskedastic model evidenced relevant departures of the residual variance in 2.9% to 37.4% of the analyzed probes, highlighting a remarkable incidence of this phenomenon in pigs. Moreover, scale-effect disturbances were discarded in most of the probes with dispersion-specific differential gene expression. These results provided the first evidence of heteroskedasticity for gene expression in a livestock species, suggesting the possibility of dispersion-related regulatory genetic mechanisms involved in sow reproduction.

Introduction

The current availability of low-cost massive platforms for profiling gene expression levels in domestic species has opened new research possibilities with a plethora of potential contributions to our livestock industries. The analysis of these genomic data tries to identify differential gene expression (Wolfinger *et al.*, 2001) associated with some phenotype of interest, for example muscle growth (Reecy *et al.*, 2006) or stress tolerance (Collier *et al.*, 2006), although all studies in domestic species typically focus on differences linked to

†E-mail: Joaquim.Casellas@uab.cat

the mathematical expectation (i.e. mean) of two or more groups of microarrays, without considering alternative patterns of departure.

In humans, Bachtiary *et al*. (2006) and Somel *et al*. (2006) have provided evidence of differential gene expression in terms of dispersion parameters, although dispersion-specific methodologies for microarray experiments were not developed until the Bayes factor (BF) approach adapted by Casellas and Varona (2008). In that methodological paper, the authors revealed a remarkable incidence of genes with heterogeneous residual variances in great ape fibroblasts, and discarded the influence of scale effects in the greatest part of them (Casellas and Varona, 2008). This novel approach allows for a straightforward screening of dispersion-related differential gene expression phenomena and provides raw material for the identification of regulatory genetic mechanisms involved in the variability of gene transcription.

Unfortunately, we lack of any preliminary information about the incidence and relevance of this kind of differential gene expression in the swine species. Within this context, our principal objective was to analyze gene expression data from uterus, ovary and hypophysis of 22 F_2 Iberian $\times$ Meishan sows, characterizing the incidence of heterogeneous residual variances when comparing three different reproductive stages (i.e. heat, 15 days of gestation and 45 days of gestation).

Material and methods

This research was carried out on frozen tissues and live animals were not required. Nevertheless, all animal protocols regarding the original F_2 Iberian $\times$ Meishan sows were approved by the *Institut de Recerca i Tecnologia Agroalimentàries* (IRTA; http://www.irta.cat) Animal Care and Use committee within the context of research project AGL2000-1229-C03.

Tissue collection

A total of 61 tissue samples from ovary, uterus and hypophysis of 22 F_2 Iberian $\times$ Meishan sows were recovered at the commercial slaughterhouse (ESFOSA, Vic, Spain). More specifically, both the whole hypophysis and one ovary were collected when possible, whereas a sample from the apical uterus was obtained close to the oviduct. Samples were immediately frozen in liquid nitrogen and stored at $-80°C$ until laboratory processing. All sows were in their fifth gestation (i.e. they had delivered four times), although they were slaughtered at different reproductive stages: heat (HEAT; $n = 8$), 15 days of pregnancy (15d; $n = 7$) and 45 days of pregnancy (45d; $n = 7$). Note that pregnancies were confirmed before slaughtering by ultrasound. Moreover, sows contributing to this experiment were selected depending on their reproductive performance in order to generate two groups with extreme prolificacy histories measured as the total number of piglets born per litter, that is, low average litter size (*L*; 5.78 $\pm$ 2.53 piglets per litter) and high average litter size (*H*; 11.48 $\pm$ 1.47 piglets per litter). A summary of the number of tissue samples per reproductive stage and performance is shown in Table 1.

Table 1 *Number of microarrays for each tissue and physiological stage combination*

	Physiological stage		
Tissue	HEAT	15d	45d
Uterus	4[1]/4[2]	4/3	3/3
Ovary	4/4	4/3	3/4
Hypophysis	4/4	2/2	3/3

[1]Microarrays from lowly prolific sows.
[2]Microarrays from highly prolific sows.

RNA extraction

RNA was extracted from frozen tissue with 1 ml of TRIzol Reagent (Sigma-Aldrich Inc., St Louis, MO, USA), homogenized with a motor-driven homogenizer and incubated for 5 min at room temperature. 200 μl of chloroform was added to the supernatant and incubated in the same conditions. After incubation, the sample was centrifuged at 4°C for 10 min, the aqueous phase was mixed with 200 μl of ethanol, removed to a filter column (RiboPure™, Ambion Inc., Austin, TX, USA) and filtered by centrifugation (12 000 **g** for 30 s at room temperature). The filter was washed twice with Wash Solution (RiboPure™, Ambion Inc.) and finally, RNA was eluted in 100 μl of Elution Buffer (RiboPure™, Ambion Inc.). The quantity of RNA was determined using a NanoDrop ND-100 (NanoDrop Technologies Inc., Wilmington, DE, USA) and all samples were adjusted between 0.7 and 1 μg/μl. The quality of the samples was determined by using a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) with the RNA 6000 Nano Kit (Agilent Technologies).

Microarray processing

Each sample was independently hybridized on a GeneChip Porcine Genomic array (Affymetrix, Santa Clara, CA, USA) using the One-Cycle Target Labeling protocol described by the microarray manufacturers. Hybridization and scanning were performed in the *Unidad de Genómica* of *Vall d'Hebron* hospital (Barcelona, Spain).

Normalization and analyses of microarray data quality

Quality control of the microarrays was performed with Affy and SimpleAffy packages of the Bioconductor (Ihaka and Gentleman, 1996; Irizarry *et al*., 2003). All the microarrays fulfilled the quality requirements, thus all of them were used in the analysis. Data normalization was performed with the Robust Multichip Average algorithm (Irizarry *et al*., 2003) using the *RMAExpress* package (http://rmaexpress.bmbolstad.com/). After filtering, 15 535 probes (each probe is a fragment of complementary nucleic acid covering genomic or inter-genomic annotated regions) were used for gene expression analyses. Probe annotation was first done using the annotation file supplied by Affymetrix (http://www.affymetrix.com/index.affx).

Heteroskedastic linear mixed model

For each pairwise comparison between two physiological stages Ω_1 and Ω_2 of a given tissue, data from *n* replicates of non-competitive hybridization microarrays were available. Gene expression data were analyzed under the following hierarchical mixed model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \mathbf{Z}_1\mathbf{p}_1 + \mathbf{Z}_2\mathbf{p}_2 + \mathbf{Z}_3\mathbf{p}_3 + \mathbf{Z}_4\mathbf{p}_4 + \mathbf{e}$$

where $\mathbf{y}$ was the $(nm)\times 1$ column vector of intensity scores sorted by microarray, within treatment within probe, and $\mathbf{e}$ was the $(nm)\times 1$ column vector of residuals. Preliminary analyses on this data set suggested significant differential gene expression on the basis of sow prolificacy. Within this context, sows were classified as lowly (*L*) or highly (*H*) prolific (see Table 1), and the model accounted for the interaction between sow prolificacy and physiological stage with four levels, $L\times\Omega_1$ ($\mathbf{p}_1$; dimension $n_1\times 1$) and $H\times\Omega_1$ ($\mathbf{p}_2$; dimension $n_2\times 1$), $L\times\Omega_2$ ($\mathbf{p}_3$; dimension $n_3\times 1$) and $H\times\Omega_2$ ($\mathbf{p}_4$; dimension $n_4\times 1$), where $n = n_1 + n_2 + n_3 + n_4$. Note that $\mathbf{p}_1$ to $\mathbf{p}_4$ accounted for the effect of each probe, whereas $\boldsymbol{\mu}$ was a systematic effect characterizing the overall mean of each microarray, and $\mathbf{X}$, $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{Z}_3$ and $\mathbf{Z}_4$ were appropriate incidence matrices.

Following Casellas and Varona (2008), vector $\mathbf{e}$ was assumed to be normally distributed,

$$\mathbf{e} \sim \text{MVN}(\mathbf{0}, \mathbf{R})$$

with $\mathbf{R}$ being the matrix of residual (co)variances. Assuming null residual (co)variances (Kendziorski *et al.*, 2003; Newton *et al.*, 2004; Casellas *et al.*, 2008), heteroskedasticity between physiological stages was analyzed by stating

$$\mathbf{R} = \bigoplus_{i=1}^{m} \begin{bmatrix} \mathbf{I}_1\sigma^2_{e(i)}\pi_i & \mathbf{0} \\ \mathbf{0}' & \mathbf{I}_2\sigma^2_{e(i)}(1-\pi_i) \end{bmatrix}$$

where $\mathbf{I}_1$ was a $(n_1 + n_2)\times(n_1 + n_2)$ identity matrix, $\mathbf{I}_2$ was a $(n_3 + n_4)\times(n_3 + n_4)$ identity matrix, $\mathbf{0}$ was a $(n_1 + n_2)\times(n_3 + n_4)$ matrix of zeros, $\sigma^2_{e(i)}$ was the sum of residual variance for the *i*th probe in both physiological stages and π_i was the probe-specific heteroskedasticity parameter. As shown by Casellas and Varona (2008), $\pi_i = 0.5$ accounted for equal residual variances between treatments and $\pi_i \neq 0.5$ suggested within-probe (between treatments) heteroskedasticity. Given the reduced number of replicates within each physiological stage, a common residual variance was assumed for both *L* and *H* sows.

This model was analyzed under a standard Bayesian development with the following joint posterior probability:

$$\begin{aligned} &p\left(\boldsymbol{\mu}, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4, \boldsymbol{\sigma}, \boldsymbol{\pi}, \sigma^2_{p1}, \sigma^2_{p2}, \sigma^2_{p3}, \sigma^2_{p4} \middle| \mathbf{y}\right) \\ &\propto p\left(\mathbf{y}\middle|\mathbf{a}, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4, \mathbf{R}\right)p(\boldsymbol{\mu}) \times p\left(\mathbf{p}_1\middle|\sigma^2_{p1}\right)p\left(\sigma^2_{p1}\right) \\ &p\left(\mathbf{p}_2\middle|\sigma^2_{p2}\right)p\left(\sigma^2_{p2}\right)p\left(\mathbf{p}_3\middle|\sigma^2_{p3}\right)p\left(\sigma^2_{p3}\right)p\left(\mathbf{p}_4\middle|\sigma^2_{p4}\right) \\ &p\left(\sigma^2_{p4}\right) \times p(\boldsymbol{\sigma})p(\boldsymbol{\pi}) \end{aligned}$$

where $\boldsymbol{\sigma}' = \left[\sigma^2_{e(1)}\ \sigma^2_{e(2)}\ \ldots\ \sigma^2_{e(m)}\right]$, $\boldsymbol{\pi}' = \left[\pi_1\ \pi_2\ \ldots\ \pi_m\right]$ and the Bayesian likelihood was stated as multivariate normal,

$$\begin{aligned} &p\left(\mathbf{y}\middle|\boldsymbol{\mu}, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4, \mathbf{R}\right) \sim \text{MVN} \\ &\left(\mathbf{X}\boldsymbol{\mu} + \mathbf{Z}_1\mathbf{p}_1 + \mathbf{Z}_2\mathbf{p}_2 + \mathbf{Z}_3\mathbf{p}_3 + \mathbf{Z}_4\mathbf{p}_4, \mathbf{R}\right) \end{aligned}$$

The *a priori* distribution for the probe effects was assumed:

$$p\left(\mathbf{p}_j\middle|\sigma^2_{pj}\right) \sim \text{MVN}\left(\mathbf{0}, \mathbf{I}_{n_j}\sigma^2_{pj}\right), \qquad j \in [1, 2, 3, 4]$$

where $\mathbf{I}_{n_j}$ was a $n_j\times n_j$ identity matrix and σ^2_{pj} was the variance component for $\mathbf{p}_j$. Scaled χ^{-2} priors with hyperparameters $S^2 = 0$ and $v = -2$ were assumed for all variance components σ^2_{pj}, and flat priors were also assumed for $\boldsymbol{\mu}$. Given that we lacked *a priori* information about probe-specific dispersion patterns in pigs, the prior distribution for $\boldsymbol{\pi}$ was defined as a mixture of distributions involving a uniform distribution between 0 and 1, and null probability otherwise (Casellas and Varona, 2008):

$$p(\boldsymbol{\pi}) \sim \prod_{k=1}^{m} 1 \qquad \text{if } \pi_k \in [0, 1] \text{ and 0 otherwise}$$

Note that this prior distribution is the key point for further testing of within-probe heteroskedasticity and covers all possible values taken for π_k with equal probability, following Verdinelli and Wasserman (1995), García-Cortés *et al.* (2001) and Varona *et al.* (2001).

BF for testing within-probe heteroskedasticity

The BF is the basic tool for comparing models in the Bayesian framework (Kass and Raftery, 1995). This factor provides the ratio of the posterior probability of both competing models without requiring the definition of the null and alternative hypotheses. Differences in residual variance between physiological stages were checked probe-by-probe by applying the BF proposed by Casellas and Varona (2008). This approach straightforwardly compares $\pi_i = 0.5$ (within-probe homogeneous residual variance for the *i*th probe, within-probe heterogeneous residual variances for the remaining probes; Model HO_i) against $\pi_i \neq 0.5$ (within-probe heterogeneous residual variances for all probes; Model HE). This BF tests probe-by-probe dispersion patterns, although it does not inform about the best analytical model for the joint inference of all probes (Casellas and Varona, 2008). The BF between Model HE and Model HO_i ($\text{BF}_{\text{HE/HO}_i}$) was calculated from the Markov chain Monte Carlo sampler output of Model HE, by averaging the full conditional densities of each cycle at $\pi_i = 0.5$ using the Rao–Blackwell argument (Wang *et al.*, 1994). Note that $\text{BF}_{\text{HE/HO}_i} > 1$ shows that Model HE is more suitable than Model HO_i, revealing a statistically relevant degree of heteroskedasticity in probe *i*. On the other hand, within-probe homogeneous residual variances are corroborated under $\text{BF}_{\text{HE/HO}_i} < 1$. Following

García-Cortés *et al*. (2001) and Varona *et al*. (2001), the posterior density $p(\pi_i = 0.5|\mathbf{y})$ suffices to obtain BF_{HE/HO_i},

$$BF_{HE/HO_i} = \frac{p(\pi_i = 0.5)}{p(\pi_i = 0.5|\mathbf{y})} = \frac{1}{p(\pi_i = 0.5|\mathbf{y})}$$

because $p(\pi_i = 0.5)$ was previously stated by $p(\boldsymbol{\pi})$ at $\pi = 0.5$.

Markov chain Monte Carlo sampling

A total of nine independent analyses were performed, accounting for each within-tissue pairwise comparison between physiological stages. For each analysis, autocorrelated samples from the relevant marginal posterior distribution of all unknowns in the model were obtained by Gibbs sampling (Gelfand and Smith, 1990), with the exception of π_i, which required a Metropolis–Hastings step (Hastings, 1970). A single chain with 500 000 elements was launched for each pairwise comparison, and the first 50 000 samples were discarded as burn-in. Sampling convergence was checked by visual inspection according to Raftery and Lewis (1992).

Correction for multiple testing

Following Kass and Raftery (1995), the posterior odds between Model HE and Model HO_i (PO_{HE/HO_i}) can be calculated as follows:

$$PO_{HE/HO_i} = BF_{HE/HO_i} \times \frac{p_{HE}}{p_{HO_i}}$$

where p_{HE} was the *a priori* probability of Model HE and p_{HO_i} was the *a priori* probability of Model HO_i. PO_{HE/HO_i} could be viewed as a weighted BF, which accounted for more realistic *a priori* probabilities for both models under multiple testing. In the standard development of the BF described above, we assumed that the ratio between p_{HE} and p_{HO_i} (prior odds) was 1 and the *a priori* probabilities for Model HE and Model HO_i were both 0.5 at each probe. This assumption provided an *a priori* expected number of probes with heteroskedastic dispersion pattern of 7676.5 in each analysis. To provide a more realistic and conservative framework, and according to dispersion-specific gene expression results reported by Casellas and Varona (2008), we *a priori* assumed that only one out of 100 probes could have heterogeneous residual variances, applying a multiplicative correction factor of 1/99 to all BF. All BF provided in this study were corrected by multiple testing.

Results and discussion

As highlighted by Wolfinger *et al*. (2001), inference in microarray gene expression analyses is typically focused on

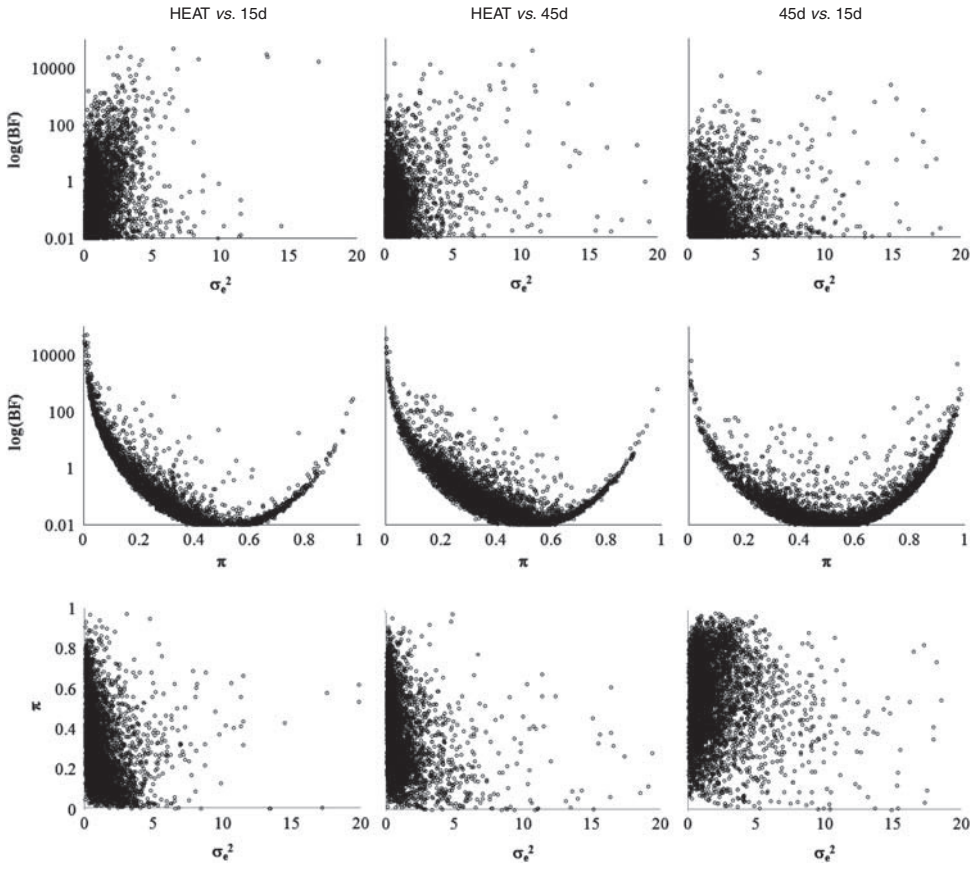

Figure 1 Plots of the pairwise distribution of $\log(BF_{HE/HO_i})$, σ_e^2 and π parameters in the uterus.

Table 2 *Distribution of the number of probes (and percentage) for each analysis according to Jeffreys' (1984) scale of evidence for Bayes factors*

	Bayes factor of Model HE against Model HO_i (BF_{HE/HO_i})					
	$BF_{HE/HO_i} < 1$	$1 \leqslant BF_{HE/HO_i} \leqslant 3.16$	$3.16 \leqslant BF_{HE/HO_i} \leqslant 10$	$10 \leqslant BF_{HE/HO_i} \leqslant 31.62$	$3.162 \leqslant BF_{HE/HO_i} \leqslant 100$	$BF_{HE/HO_i} \geqslant 100$
Uterus						
HEAT *v.* 15d	11 915 (77.6)	1200 (7.8)	845 (5.5)	578 (3.8)	354 (2.3)	461 (3.0)
HEAT *v.* 45d	13 058 (85.0)	1013 (6.6)	561 (3.7)	325 (2.1)	163 (1.1)	233 (1.5)
15d *v.* 45d	14 014 (91.2)	590 (3.8)	345 (2.2)	199 (1.3)	121 (0.8)	84 (0.5)
Ovary						
HEAT *v.* 15d	14 714 (95.8)	329 (2.1)	141 (0.9)	83 (0.5)	46 (0.3)	40 (0.3)
HEAT *v.* 45d	14 644 (95.3)	378 (2.5)	172 (1.1)	83 (0.5)	44 (0.3)	32 (0.2)
15d *v.* 45d	14 917 (97.1)	235 (1.5)	128 (0.8)	38 (0.2)	22 (0.1)	13 (0.1)
Hypophysis						
HEAT *v.* 15d	9613 (62.6)	3060 (19.9)	1631 (10.6)	643 (4.2)	212 (1.4)	194 (1.3)
HEAT *v.* 45d	14 767 (96.1)	352 (2.3)	147 (1.0)	54 (0.4)	24 (0.2)	9 (0.1)
15d *v.* 45d	14 590 (95.0)	482 (3.1)	158 (1.0)	69 (0.4)	36 (0.2)	18 (0.1)

Table 3 *Distribution of the number (and percentage) of probes with mean-related differential expression on the dispersion-related categories defined by the Bayes factor of Model HE against Model* HO_i *(*BF_{HE/HO_i}*)*

	Bayes factor of Model HE against Model HO_i (BF_{HE/HO_i})					
	$BF_{HE/HO_i} < 1$	$1 \leqslant BF_{HE/HO_i} \leqslant 3.16$	$3.16 \leqslant BF_{HE/HO_i} \leqslant 10$	$10 \leqslant BF_{HE/HO_i} \leqslant 31.62$	$3.162 \leqslant BF_{HE/HO_i} \leqslant 100$	$BF_{HE/HO_i} \geqslant 100$
Uterus						
HEAT *v.* 15d	262[1]/268[2] (2.1/2.2)	18/21 (1.5/1.7)	6/4 (0.7/0.4)	8/10 (1.3/1.7)	2/6 (0.5/1.6)	5/2 (1.0/0.4)
HEAT *v.* 45d	240/381 (1.8/2.9)	17/26 (1.6/2.5)	4/11 (0.7/1.9)	4/5 (1.2/1.5)	1/5 (0.6/3.0)	1/5 (0.4/2.1)
15d *v.* 45d	168/146 (1.1/1.0)	6/4 (1.0/0.6)	5/5 (1.4/1.4)	1/0 (0.5/0)	0/0 (0/0)	1/0 (1.1/0)
Ovary						
HEAT *v.* 15d	421/498 (2.8/3.3)	13/20 (3.9/6.0)	6/9 (4.2/6.3)	2/5 (2.4/6.0)	2/1 (4.3/2.1)	2/3 (5/7.5)
HEAT *v.* 45d	386/544 (2.6/3.7)	18/19 (4.7/5.0)	4/6 (2.3/3.4)	3/2 (3.6/2.4)	3/3 (6.8/6.8)	1/0 (3.1/0)
15d *v.* 45d	140/109 (0.9/0.7)	1/4 (0.4/1.7)	2/0 (1.5/0)	1/3 (2.6/7.8)	0/0 (0/0)	0/0 (0/0)
Hypophysis						
HEAT *v.* 15d	16/20 (0.1/0.2)	10/11 (0.3/0.3)	2/2 (0.1/0.1)	0/1 (0/0.1)	1/1 (0.4/0.4)	0/0 (0/0)
HEAT *v.* 45d	68/59 (0.4/0.3)	1/1 (0.2/0.2)	0/2 (0/1.3)	0/0 (0/0)	0/0 (0/0)	0/0 (0/0)
15d *v.* 45d	10/14 (0.0/0.0)	0/0 (0/0)	0/0 (0/0)	0/0 (0/0)	0/0 (0/0)	0/0 (0/0)

[1]Comparison between lowly prolific sows.
[2]Comparison between highly prolific sows.

gene-specific differences between mathematical expectations of two (or more) groups of biological conditions. However, discrepancies in gene expression could also be characterized by other statistics of interest like dispersion parameters (Bachtiary *et al.*, 2006; Somel *et al.*, 2006). Despite the recent release of massive microarray platforms for evaluating gene expression levels, dispersion-related differential gene expression patterns were poorly investigated in mammals and we lack prospective results in domestic species of livestock. The recent development of a specific BF approach to check for within-probe between-groups heterogeneous residual variances in microarray studies (Casellas and Varona, 2008) opened a new research framework for gene expression analyses, allowing for genome-wide screenings of dispersion-related differential gene expression with minimum computational requirements. We adopted this methodology to characterize the incidence of dispersion-related differential gene expression in sows, the first screening of this phenomenon in a livestock species. Nevertheless, this must be viewed as one of the multiple statistical approaches that can be used to unravel dispersion-specific differential gene expression in livestock. Indeed, this model can be easily adapted to account for tail-area probability approaches (see Blasco (2005) for additional details) or even equivalence testing (Morey and Rouder, 2011) among others.

Taking the estimates from the analysis performed on uterus samples as an example (Figure 1), our results can be straightforwardly characterized by plotting the dispersion pattern of the BF itself and the two main variables involved in its parameterization, π_i and $\sigma^2_{e(i)}$. As expected, $\sigma^2_{e(i)}$ estimates accumulated to the left boundary of the positive parametric space (lower and upper rows, Figure 1), resulting in a shaped exponential pattern with minimum incidence of large variances (e.g. 1.4% (HEAT *v.* 15d) to 3.2% (45d *v.* 15d) of $\sigma^2_{e(i)}$ estimates were >5) and without showing a clear linkage with realized BF_{HE/HO_i} (Figure 1, upper row). Conversely, π_i estimates distributed across all the parametric

space (i.e. from 0 to 1) and showed a clear linkage with BF_{HE/HO_i}. As previously reported by Casellas and Varona (2008), BF_{HE/HO_i} resulted in a very characteristic U-shaped pattern when plotted against π_i (Figure 1, medium row), increasing BF_{HE/HO_i} with extreme values of π_i (i.e. π_i tending to 0 or 1) and decreasing with intermediate estimates of π_i. Both $\sigma^2_{e(i)}$ and π_i gave us a detailed idea about the overall dispersion pattern of gene expression within each tissue and reproductive stage (Figure 1, lower row). Focusing on uterus, HEAT *v.* 15d and HEAT *v.* 45d analyses provided a skewed distribution of both $\sigma^2_{e(i)}$ and π_i, with a higher incidence of small π_i estimates and suggesting an increase of the average $\sigma^2_{e(i)}$ with smaller π_i. Both results indicated a smaller residual variability in gene expression during heat when compared with further gestational progress (i.e. 15d and 45d), which suggested a higher need for accurate genetic regulation during this initial reproductive stage. The last comparison (45d *v.* 15d) tended to show a higher incidence of large π_i values (i.e. closer to 1), which was indicative of a smaller overall residual dispersion of gene expression at 15d than at 45d. As a whole, these plots showed a regular temporal trend across heat and gestation in sows, suggesting that the variability of the overall gene expression increased with time, at least from heat to 45d of gestation. Very similar plots and the same overall trends were observed in both ovary and hypophysis tissues (results not shown).

As shown in Table 2, most of the analyzed probes did not reveal relevant departures from the null hypothesis of homogeneous residual variance ($BF_{HE/HO_i} < 1$), ranging between 77.6% and 91.2% of the total in uterus, 95.8% and 97.1% in ovary and 62.6% and 96.1% in hypophysis. Nevertheless, the main result was not these non-heterogeneous probes but the 2.9% to 37.4% of probes providing statistical evidence of dispersion-related differential gene expression. Note that within-probe heterogeneous residual variances were previously reported when comparing great ape tissue samples (Casellas and Varona, 2008), although our results contribute the first evidence of within-probe heteroskedasticity in a domestic species of mammals. Following Jeffreys' (1984) scale of evidence for BFs, 1.5% to 19.9% of probes provided almost irrelevant evidence favoring Model HE ($1 \leqslant BF_{HE/HO_i} \leqslant 3.16$), whereas substantial (0.8% to 10.6%; $3.16 \leqslant BF_{HE/HO_i} \leqslant 10$), strong (0.2% to 4.2%; $10 \leqslant BF_{HE/HO_i} \leqslant 31.62$), very strong (0.1% to 2.3%; $31.62 \leqslant BF_{HE/HO_i} < 100$) and decisive (0.1% to 3.0%; $BF_{HE/HO_i} \geqslant 100$) results must be highlighted in our study. These estimates reported a remarkable incidence of dispersion-related differential gene expression in sows, which were revealed in all three biological tissues included in our study (i.e. ovary, uterus and hypophysis) and across all the reproductive stages (i.e. HEAT, 15d and 45d of pregnancy; Table 1).

One could hypothesize that this between-groups heterogeneity of the residual variance of gene expression could be linked to scale effects because of differential gene expressions in terms of mathematical expectations. Nevertheless, this hypothesis was discarded by the results shown in Table 3. Only 0% to 7.5% of the genes with larger than

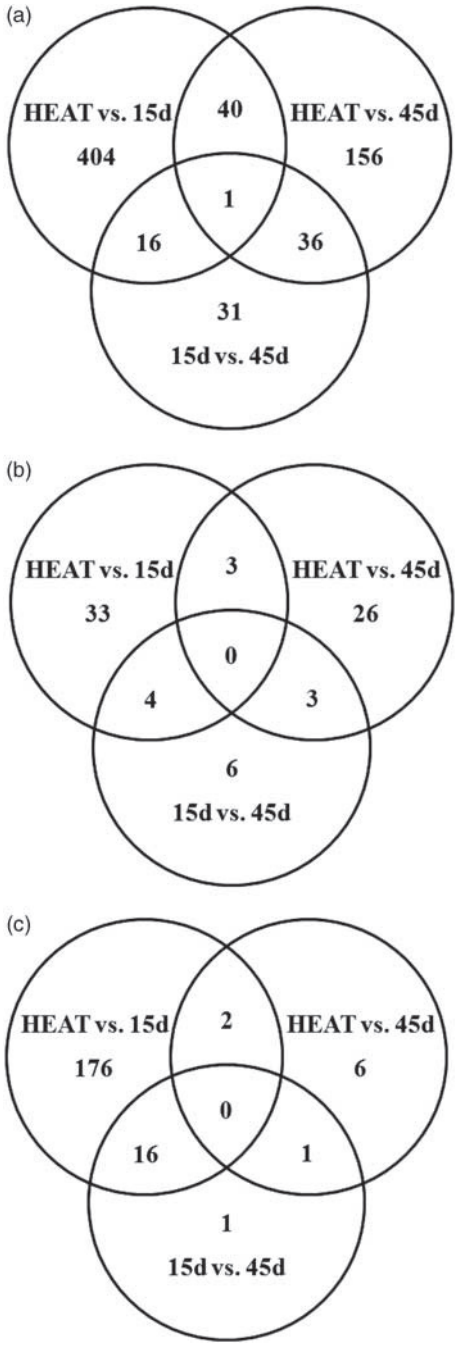

Figure 2 Venn diagrams accounting for the number of highly heteroskedastic probes ($BF_{HE/HO_i} > 100$) in uterus (a), ovary (b) and hypophysis (c).

one BF_{HE/HO_i} for heterogeneous residual variances also revealed relevant differential gene expressions in terms of mathematical expectations at a posterior probability of 95% (see Casellas *et al.* (2008) for additional details on the calculation of this posterior probability). These results agreed with Casellas and Varona (2008) and, although they cannot be extrapolated to all microarray experiments, they suggest that heterogeneous residual patterns could be a biological phenomenon of special interest in further analyses of gene expression data in the swine species, independent from scale-effect disturbances.

Table 4 *Summary of probes with decisive evidence of heterogeneous ($BF_{HE/HO_i} > 100$) and/or homogeneous ($BF_{HE/HO_i} < 0.01$) residual variances in all three pairwise comparisons*

		HEAT *v.* 15d		HEAT *v.* 45d		15d *v.* 45d	
Gene	Tissue	π_i	BF_{HE/HO_i}	π_i	BF_{HE/HO_i}	π_i	BF_{HE/HO_i}
(HEAT = 15d) < 45d[1]							
ARPP21	Ut	0.47	9×10^{-3}	0.01	2558.7	0.02	360.0
APOA2	Ut	0.47	9×10^{-3}	0.02	1078.0	0.03	108.1
ASB11	Ut	0.52	8×10^{-3}	0.01	2375.4	0.01	749.1
BPI	Ut	0.54	1×10^{-2}	0.00	5262.7	0.01	1469.7
CPB2	Ut	0.51	9×10^{-3}	0.02	498.4	0.02	127.5
FDX1	Ut	0.54	9×10^{-3}	0.01	1338.2	0.02	491.1
C7orf45	Ut	0.55	9×10^{-3}	0.01	1413.6	0.01	742.9
SLC16A12	Ut	0.56	9×10^{-3}	0.02	659.8	0.01	438.1
PFKM	Ov	0.52	9×10^{-3}	0.99	855.3	0.98	361.9
C5orf23	Ut	0.53	9×10^{-3}	0.03	335.2	0.03	132.8
HEAT < (15d = 45d)[2]							
B3GAT2	Ut	0.03	210.2	0.04	132.2	0.51	9×10^{-3}
GALNT7	Ut	0.03	161.3	0.02	298.4	0.43	8×10^{-3}
GIT1	Ut	0.03	233.9	0.04	117.5	0.52	9×10^{-3}
MBNL1	Ut	0.02	698.0	0.02	304.3	0.50	7×10^{-3}
C5orf32	Ut	0.01	6068.0	0.02	316.4	0.56	8×10^{-3}
TMEM123	Ut	0.02	891.6	0.03	372.8	0.55	9×10^{-3}
TMEM45B	Ut	0.01	12 557.7	0.00	14 264.2	0.41	9×10^{-3}
HKDC1	Ov	0.02	1009.9	0.03	200.7	0.58	9×10^{-3}
SERPINB7	Ut	0.02	514.4	0.02	203.4	0.51	7×10^{-3}
SLC16A1	Ut	0.03	190.1	0.02	373.4	0.41	8×10^{-3}
SYT13	Ut	0.01	1555.7	0.03	346.9	0.58	1×10^{-2}
TPP2	Ut	0.03	204.1	0.03	207.1	0.48	8×10^{-3}
HEAT < (15d > 45d)[3]							
RPS3A	Ut	0.00	25 529.8	0.20	143.8	0.74	137.738

Ut = uterus; Ov = ovary.
Note that 118 (uterus) and 161 (ovary) probes with homogeneous residual variances ($BF_{HE/HO_i} < 0.01$) across all tissues have been omitted.
[1]Residual variances were homogeneous between HEAT and 15d, but were heterogeneous between these two groups and 45d.
[2]Residual variances were homogeneous between 15d and 45d, but were heterogeneous between these two groups and HEAT.
[3]Residual variances were heterogeneous across all groups, greater in 15d than in 45d, and greater in 45d than in HEAT.

Venn diagrams provided evidence of a higher incidence of heterogenically expressed genes (with $BF_{HE/HO_i} \geqslant 100$) in uterus than in ovary or hypophysis (Figure 2), although only one gene in uterus (*RPS3A*; Table 4) revealed decisive evidence of heteroskedasticity in all three pairwise comparisons between reproductive stages. This gene, which encodes for the 40S ribosomal protein S3a of the small ribosomal subunit and has been linked to apoptosis and cell differentiation phenomena (Naora and Naora, 1999), showed greater gene expression variability at 15d than at 45d and also at 45d compared with HEAT. Regarding the remaining genes with decisive evidence of heteroskedasticity or homoskedasticity in all three pairwise comparisons (Table 4), they were linked to multiple biological processes and metabolic pathways, suggesting that dispersion-specific differential gene expression could be widely distributed across the whole genome. Although the main target of this research was not focused on the identification of specific heteroskedastic genes but on checking for the suitability of this phenomenon, and its incidence in the whole genome of swine, probes shown in Table 4 represent a very promising starting point for identifying the genetic mechanisms involved in the regulation of gene transcription on a dispersion basis. Moreover, commonly used housekeeping genes such as *GAPDH*, *HMBS*, *RPL32*, *SDHA* and *YWHAZ* (Vandesompele *et al.*, 2002) were also included in the analysis and none of them showed statistically relevant dispersion-specific differential gene expression; this provided additional confidence about differentially expressed genes shown in Table 4.

As recently demonstrated at a gene-specific level (Lo and Gottardo, 2007), an accurate modeling of residual dispersion patterns allows for a more realistic fit of gene expression data, reducing the rate of false positives when differential gene expression is characterized in terms of mathematical expectations or their differences (Kendziorski *et al.*, 2003; Newton *et al.*, 2004; Lo and Gottardo, 2007). Moreover, the analysis of heterogeneous residual dispersion patterns opens up promising research possibilities within the gene expression framework, where heterogeneity in residual variability could be viewed as an alternative and plausible characterization of differential expression patterns. The current study provided the first evidence of massive heteroskedasticity

for gene expression in the swine species, a phenomenon that must be accurately considered in further studies in order to provide a detailed characterization of the genetic architecture and regulation of domestic pigs. Moreover, and if confirmed in future analyses, the list of genes provided in Table 4 must be considered as an appealing starting point for a deeper understanding of the genetic pathways involved in sow reproduction, and may reveal dispersion-related regulatory mechanisms.

Acknowledgments

Financial support was provided by Spain's *Ministerio de Ciencia y Tecnología* (AGL2000-1229-C03), *Ministerio de Educación y Ciencia* (AGL2004-08368-C03) and *Ministerio de Ciencia e Innovación* (AGL2008-04818-C03). The authors are indebted to the staff of Nova Genètica (Solsona, Spain) for cooperating in the experimental protocol, in particular to F. Márquez, I. Riart, R. Malé, F. Rovira, M. González and E. Ramells, to M. Arqué (IRTA-Lleida, Lleida, Spain) for her active and friendly collaboration during data collection and to the staff of the *Escorxador Frigorific d'Osona SA* (ESFOSA) for facilitating tissue sample collection at slaughter. The authors gratefully acknowledge the contributions of the INRA (France) and the SIA El Dehesón del Encinar (Spain) for providing the purebred Meishan sows and Iberian boars, respectively. The research contract of J. Casellas was partially financed by Spain's *Ministerio de Ciencia e Innovación* (programs *Juan de la Cierva*, *José Castillejo* and *Ramon y Cajal*). M. Martínez-Giner was funded by a fellowship from the Spain's *Ministerio de Educación y Ciencia*.

References

Bachtiary B, Boutros PC, Pintile M, Bastianutto C, Li J-H, Schwock J, Zhang W, Penn LZ, Jurisica I, Fyles A and Liu F-F 2006. Gene expression profiling in cervical cancer: an exploration of intratumor heterogeneity. Human Cancer Research 12, 5632–5640.

Blasco A 2005. The use of Bayesian statistics in meat quality analyses: a review. Meat Science 69, 115–122.

Casellas J and Varona L 2008. Between-groups within-gene heterogeneity of residual variances in microarray gene expression data. BMC Genomics 9, 319.

Casellas J, Ibáñez-Escriche N, Martínez-Giner M and Varona L 2008. GEAMM v.1.4: a versatile program for mixed model analysis of gene expression data. Animal Genetics 39, 89–90.

Collier RJ, Stiening CM, Pollard BC, VanBaale MJ, Baumgard LH, Gentry PC and Coussens PM 2006. Use of gene expression microarrays for evaluating environmental stress tolerance at the cellular level in cattle. Journal of Animal Science 84, E1–E13.

García-Cortés LA, Cabrillo C, Moreno C and Varona L 2001. Hypothesis testing for the genetic background of quantitative traits. Genetics, Selection, Evolution 33, 3–16.

Gelfand AE and Smith AFM 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85, 398–409.

Hastings WK 1970. Monte-Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Ihaka R and Gentleman RR 1996. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics 5, 299–314.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4, 249–264.

Jeffreys H 1984. Theory of probability. Clarendon Press, Oxford, UK.

Kass RE and Raftery AE 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.

Kendziorski CM, Newton MA, Lan H and Gould MN 2003. On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. Statistics in Medicine 22, 3899–3914.

Lo K and Gottardo R 2007. Flexible empirical Bayes methods for differential gene expression. Bioinformatics 23, 328–335.

Morey RD and Rouder JN 2011. Bayes factor approaches for testing interval null hypotheses. Psychological Methods 16, 406–419.

Naora H and Naora H 1999. Involvement of ribosomal proteins in regulating cell growth and apoptosis: translational modulation or recruitment for extraribosomal activity? Immunology & Cell Biology 77, 197–205.

Newton MA, Noueiry A, Sarkar D and Ahlquist P 2004. Detecting differential gene expression with a semiparametric hierarchical mixture method. Biostatistics 5, 155–176.

Raftery AE and Lewis SM 1992. How many iterations in the Gibbs sampler? In Bayesian statistics IV (ed. JM Bernardo, JO Berger, AP Dawid, AFM Smith), pp. 763–774. Oxford University Press, Oxford, UK.

Reecy JM, Moody Spurlock D and Stahl CH 2006. Gene expression profiling: insights into skeletal muscle growth and development. Journal of Animal Science 84, E150.

Somel M, Khaitovich P, Bahn S, Pääbo S and Lachmann M 2006. Gene expression becomes heterogeneous with age. Current Biology 16, Article no. R360.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A and Speleman F 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology 3, Article no. 34.

Varona L, García-Cortés LA and Pérez-Enciso M 2001. Bayes factors for detection of quantitative trait loci. Genetics, Selection, Evolution 33, 133–152.

Verdinelli I and Wasserman L 1995. Computing Bayes factors using a generalization of the savage-dickey density ratio. Journal of the American Statistical Association 90, 614–618.

Wang CS, Rutledge JJ and Gianola D 1994. Bayesian-analysis of mixed linear-models via Gibbs sampling with an application to litter size in Iberian pigs. Genetics, Selection, Evolution 26, 91–115.

Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C and Paules RS 2001. Assessing gene significance from cDNA microarray expression data via mixed models. Journal of Computational Biology 8, 625–637.