A small set of extra-embryonic genes defines a new landmark for bovine embryo staging

Séverine A Degrelle, Kim-Anh Lê Cao¹,², Yvan Heyman, Robin E Everts³, Evelyne Campion, Christophe Richard⁶, Céline Ducroix-Crépy⁵, X Cindy Tian⁶, Harris A Lewin³,⁷, Jean-Paul Renard, Christèle Robert-Granie¹ and Isabelle Hue

INRA-ENVA, UMR 1198 Biologie du Développement et Reproduction, Domaine de Vilvert, F-78350 Jouy en Josas, France, ¹INRA, UR631 Station d’Amélioration Génétique des Animaux, F-31326 Castanet, France, ²Institut de Mathématiques, Université de Toulouse et CNRS (UMR 5219), F-31062 Toulouse, France, ³Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA, ⁴INRA, UE331 Unité Commune d’Expérimentation Animale de Bressonvilliers, F-91630 Leudeville, France, ⁵Centre de Ressources Biologiques pour la Génomique des Animaux d’Elevages et d’Intérêt Economique, F-78350 Jouy-en-Josas, France, ⁶Department of Animal Science and Center for Regenerative Biology, University of Connecticut, Storrs, Connecticut 06269, USA and ⁷Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Correspondence should be addressed to I Hue; Email: isabelle.hue@jouy.inra.fr

R E Everts is now at SEQUENOM, Inc., 3595 John Hopkins Court, San Diego, California 92121, USA

Abstract

Axis specification in mouse is determined by a sequence of reciprocal interactions between embryonic and extra-embryonic tissues so that a few extra-embryonic genes appear as ‘patterning’ the embryo. Considering these interactions as essential, but lacking in most mammals the genetically driven approaches used in mouse and the corresponding patterning mutants, we examined whether a molecular signature originating from extra-embryonic tissues could relate to the developmental stage of the embryo proper and predict it. To this end, we have profiled bovine extra-embryonic tissues at peri-implantation stages, when gastrulation and early neurulation occur, and analysed the subsequent expression profiles through the use of predictive methods as previously reported for tumour classification. A set of six genes (CALM1, CPA3, CITED1, DLD, HNRNPDL, and TGFB3), half of which had not been previously associated with any extra-embryonic feature, appeared significantly discriminative and mainly dependent on embryonic tissues for its faithful expression. The predictive value of this set of genes for gastrulation and early neurulation stages, as assessed on naive samples, was remarkably high (93%). In silico connected to the bovine orthologues of the mouse patterning genes, this gene set is proposed as a new trait for embryo staging. As such, this will allow saving the bovine embryo proper for molecular or cellular studies. To us, it offers as well new perspectives for developmental phenotyping and modelling of embryonic/extra-embryonic co-differentiation.

Reproduction (2011) 141 79–89

Introduction

Developmental biologists have a specific interest in spatio-temporally regulated processes. These often appear conserved even though their timing can dramatically vary between species. To rely on conserved morphological traits rather than on time schedules, the use of developmental staging has provided worldwide standards for common classifications. This was prior to the advent of high-throughput genetics or transcriptomics but is nowadays increasingly important for species-centric as well as cross-species ontology (Hughes et al. 2008, Mungall et al. 2010), automated search of biological databases and literature mining.

Gastrulation is a highly dynamic but conserved process which enables two major events of early vertebrate development: i) the establishment of morphologically visible antero-posterior (or head–tail) and dorso-ventral (or mouth–spinal cord) axes, which outline the future body plan, and ii) the specification and early patterning of the germ layers, especially the mesoderm. In birds and mammals, the primitive streak signals the onset of gastrulation, identifies the posterior end of the embryo and morphologically defines the antero-posterior axis. Later, neural plate and head fold relate to the head process at the anterior end of the embryo. In the developing trunk, notochord and somite formation initiate neurulation. These conserved traits have thus quickly completed the morphological description of embryonic stages, initially defined as days post fertilisation. The morphology of the extra-embryonic tissues...
has also contributed to the refinement of these stages in the mouse (Downs & Davies 1993). These tissues, which provide maternal nutrients to the embryo, were indeed well characterised as a part of the placenta (Cross 2005). Moreover, gastrulation mutants recently revealed that axis specification in the embryo proper is a sequence of reciprocal interactions between embryonic and extra-embryonic tissues, so that genes involved in these interactions (Pfister et al. 2007, Arnold & Robertson 2009) could now contribute to a morphogenetic staging (Tam & Loebel 2007). Among them a few signals emanate from the extra-embryonic ectoderm (Exe) to pattern the anterior visceral endoderm and the epiblast while a few others emanate from the epiblast to restrict the expression of genes in the Exe. However, there are hardly any reports in which gene expression of mouse extra-embryonic tissues from E5.5 to E7.5 was studied (Ko et al. 1998, Hemberger et al. 2001, Frankenberg et al. 2007).

Conversely, there are few descriptions of such interactions in other mammals due to the lack of bona fide ES cells (Munoz et al. 2008, Talbot & Blomberg 2008, Cao et al. 2009, Hall et al. 2009, Pant & Keefer 2009) to generate knockouts and use them in genetically driven approaches. Even though gene knockouts using somatic nuclear transfer have been reported, these are long lasting and far from routine procedures (three reports so far in goats (Yu et al. 2009), pigs (Fujimura et al. 2008) and cows (Sendai et al. 2006)). Nonetheless, alternative approaches can be used, including molecular profiling studies. Indeed, considering that reciprocal interactions between embryonic and extra-embryonic tissues are essential to all mammals, embryonic and/or extra-embryonic genes should be identifiable in other species as well. In this study, we have searched for genes from extra-embryonic tissues that would mirror the developmental stages of embryonic tissues and predict them.

To do so, we chose cattle as a model organism for the following reasons: i) gastrulation occurs well ahead of implantation so that conceptuses are easily recovered, ii) bovine extra-embryonic tissues elongate while gastrulation proceeds and only appose to the uterine epithelium at the time of neurulation (Maddox-Hytte et al. 2003) and iii) data on gene expression in bovine elongating tissues are available (Degrelle et al. 2005); however, prediction of embryonic stages through extra-embryonic signatures has not been considered so far. We conducted a gene expression study using elongating tissues from early gastrula to early bovine neurula, used statistical methods dedicated to prediction (Nuyten & van de Vijver 2008) and identified a small set of genes that accurately predicted the embryonic stages from early gastrulation to early neurulation in cows. We propose this predictive set as a new standard to replace morphological staging and thus save the embryo proper in the view of other studies and discuss its significance with respect to embryonic/extra-embryonic co-differentiations.

**Results**

**Hundreds of gene expressions are discriminative among extra-embryonic tissues**

To search for genes from extra-embryonic tissues, which are able to mirror and predict the developmental stage of the corresponding embryonic tissues, we first collected conceptuses representing the main stages encompassing gastrulation and early neurulation in cows. These stages were evaluated on the embryo proper according to morphological traits and Brachyury expression patterns (Hue et al. 2001) and named according to the chick classification system, also used for rabbit or pig embryos (Blomberg et al. 2008, Hassoun et al. 2009). We defined a pre-streak embryo as stage 2, an early streak embryo as stage 3, a late-streak embryo as stage 4 and a head fold embryo as stage 5 with a (+) when pairs of somites were formed (Fig. 1). While selecting 5 conceptuses per stage, we ended up with 20 conceptuses. The corresponding extra-embryonic tissues were submitted individually to gene expression profiling (Fig. 2). The array data set was scrutinised for discriminative expressions by combining four statistical tools: a filter method (ANOVA) and three predictive methods (random forests (RFs), optimal feature weighting algorithm (ofwCART and ofwSVM)). We identified hundreds of discriminative sequences and focused our interest on the first sequences selected by each method, according to their adjusted P values. As evidenced by a first screen through the Ingenuity Pathway Analysis (IPA) software, the main functions and pathways at work in these extra-embryonic tissues were as previously described for elongating conceptuses

![Figure 1](https://example.com/figure1.png)

**Figure 1** Embryonic classifications at gastrulating and early neurulating stages: simplified description of the main stages and classifications usually used. The chick classification is the HH classification (HH for Hamburger and Hamilton). Here, the bovine 5/5+ stage corresponds to HH10–HH14. Referring to this study, the morphology of the embryos as seen under the microscope and the Brachyury expression patterns (WISH) are presented. WISH, whole-mount in situ hybridisation. Scale bars = 500 μm.

---

*Hensen's node (chick)*
and TGFB3 clear in the box plots as in the array data. Conversely, expression at stage 2 as compared with older stages. showed, in Fig. 4A and B, a discriminatively low expression at all stages and were therefore not validated extra-embryonic tissues by qualitative RT-PCR (Fig. 4A).

To check the expression of each sequence on bovine morphological staging (Fig. 1). As clearly evidenced in such PCR analyses to mirror the qualitative traits of the gene effects were significant. This allowed us to rely on embryos per stage (Fig. 4B) and confirmed that stage and to assess the biological variability among similarly staged embryos, we repeated these analyses on five embryos per stage (Fig. 4B) and confirmed that stage and gene effects were significant. This allowed us to rely on such PCR analyses to mirror the qualitative traits of the morphological staging (Fig. 1). As clearly evidenced in Fig. 4A, though less clear on Fig. 4B, most profiles but not all were similar to the profiling data. For example, CAPZA2, BF039259 and CN434538 displayed a similar expression at all stages and were therefore not validated as discriminative (not shown). DLD and CALM1 behaved differently in the box plots and the array data but showed, in Fig. 4A and B, a discriminatively low expression at stage 2 as compared with older stages. Conversely, HNRNPD showed a discriminatively high expression at stage 2, in Fig. 4A and B, though not as clear in the box plots as in the array data. CPA3, CITED1 and TGFB3 behaved, in Fig. 4A and B, as expected from the array data and helped discriminating stage 5 from others (CPA3, CITED1) or stage 4 from others (TGFB3). In the view of an easy staging procedure by qualitative PCR, we retained six sequences out of nine as discriminative among stages and studied their localisation within elongating extra-embryonic tissues, at one of the stages where they were highly expressed (Fig. 4C). Subtle distinctions appeared between the genes expressed in all trophoblast cells (CALM1, CITED1 and DLD) and those expressed mainly in the differentiated trophoblast cells, namely the bovine binucleated cells (CPA3).

In a search for extra-embryonic genes that were able to mirror and predict the developmental stage of the corresponding embryonic tissues, we achieved the first step but had to assess whether the expression profile of these six discriminative genes could predict the stage of an embryo proper when assessed by qualitative RT-PCR on unclassified extra-embryonic tissues.

A small set of these genes predicts embryonic stages and depends on the embryonic tissues

On the basis of 15 embryos (collected independently of the first ones within the same developmental window and thus referred as ‘test’ sample as opposed to the 20 first embryos defined as ‘training’ sample; see Fig. 2) and using the RF method, the 6 discriminative genes we identified predicted proper stages in 93% of the cases (Table 2). All predictions were confirmed using microscopy and referring to the morphological traits described in Fig. 1 but were additionally challenged by whole-mount in situ hybridisation with a well established marker of these stages, the Brachyury gene (Fig. 1). All appeared properly predicted except one (B175). Predicting accurately the stages of 14 embryos out of 15, we propose this extra-embryonic gene set as a new qualitative trait of bovine embryonic development at gastrulation and early neurulation stages and advise its use referring to the scheme in Fig. 5A.

To assess the biological significance of this gene set, we generated bovine extra-embryonic vesicles by ablation of the embryonic tissues on in vivo developed conceptuses at stage 3 (detailed procedure in section Materials and Methods). This evidenced that

![Figure 2 Experimental design.](image)

(www.reproduction-online.org

Table 1 Gene networks and pathways corresponding to the two top functions identified by Ingenuity Pathway Analysis (IPA) on bovine extra-embryonic tissues, after the use of each prediction method.

<table>
<thead>
<tr>
<th>IPA analyses (version 6.5)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F test</td>
<td>RF</td>
<td>ofwCART</td>
<td>ofwSVM</td>
</tr>
<tr>
<td>Top functions</td>
<td>Cancer (32 genes)</td>
<td>Cellular cycle (38 genes)</td>
<td>Cellular development (22 genes)</td>
<td>Cell cycle, cancer (33 genes)</td>
</tr>
<tr>
<td></td>
<td>Cellular growth and proliferation (26 genes)</td>
<td>Cellular movement (21 genes)</td>
<td>Cancer, cell death (22 genes)</td>
<td>Embryonic development (22 genes)</td>
</tr>
<tr>
<td>Top pathways</td>
<td>Oxidative stress</td>
<td>G1/S transition of the cell cycle</td>
<td>TGFβ signalling</td>
<td>Gene regulation by peroxisome proliferators via PPARA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PPATA/RXR activation</td>
<td>G1/S transition of the cell cycle</td>
</tr>
</tbody>
</table>

---

Downloaded from Bioscientifica.com at 02/27/2019 11:28:35PM via free access
prediction of our gene set once applied to these vesicles did not fit into stage 4 or stage 3 (Fig. 5B), though these stages were accurately predicted on the test sample (Table 2). Moreover, our controls made us confident that the experimental steps required for this experiment did not prevent a day-15 embryonic disc to reach stage 4 (see section Materials and Methods). Despite an ongoing elongation of the extra-embryonic tissues on their own, this misclassification confirmed that this extra-embryonic gene set i) mirrors the developmental stage of the embryo proper and ii) mainly depends on it for its faithful expression within the extra-embryonic tissues.

Discussion

A small set of extra-embryonic genes predicts embryonic stages

We initially wondered whether the tissues that will contribute to the placenta could be used to predict gastrulating and early neurulating stages in the embryo proper. As an answer, we identified in bovine extra-embryonic tissues a set of 6 genes that i) accurately predicted the embryonic stage of 14 embryos in a sample of 15 and ii) was mainly dependent on the embryonic tissues for its faithful expression. So far, molecular traits of embryonic differentiation in livestock species referred to the expression patterns of embryonic genes such as Goosecoid (van de Pavert et al. 2001), Brachyury (Hue et al. 2001), Eomesodermin (Guillomot et al. 2004), POU5F1 (Oct4; Degrelle et al. 2005) or NODAL (Hassoun et al. 2009). To our knowledge, the prediction of gastrulating and early neurulating stages based on the expression pattern of a few extra-embryonic genes such as CPA3, CALM1 or HNRNPDL is new.

This gene set brings an innovative concept

In the absence of the genetically driven approaches easily used in the mouse, the use of a molecular profiling study to identify extra-embryonic genes that would mirror developmental stages of embryonic tissues and predict them was successful. Nevertheless, two limitations should be underlined: first of all, the use of a bovine array of limited size, which did not include the full spectrum of genes to be expressed in these tissues and stages (what RNA sequencing would now do) and second the incomplete annotation of the bovine genome that leaves sequences unidentified and genomic locations unsolved in either Btau4.2 or UMD3.1 (http://www.ncbi.nlm.nih.gov/sites/genesis).

As frequently used in tumour classification, the use of statistical methods dedicated to prediction helped us to i) identify large number of discriminative sequences within the bovine extra-embryonic data set and ii) validate a small subset of genes that accurately predicted the embryonic stages from early gastrulation to early neurulation. Discriminative and predictive genes seem to be odd concepts due to the prevalence of differential gene expressions within developmental profiling studies. In this study, we understand them as follows: discriminative refers to genes that are able to distinguish different consecutive stages (likely the few that are involved in the emergence and determination of a given stage), while predictive corresponds to the necessary and sufficient discriminative genes, the pattern of which identifies these stages on a naïve sample. Though not applied to genes but to morphological traits, discriminative and predictive are criteria applied to all staging procedures.
Most of the misclassifications that occurred on the test sample were subtle, predicting stage 2 instead of stage 3. Such misclassifications were described as well with morphological traits of gastrulation. The pre-streak stage of the mouse embryo, the equivalent of our stage 2, was indeed reported by Downs & Davies (1993) as difficult to identify unambiguously with the use of morphological traits only, these leading sometimes to 11% of wrong assessment. It was therefore no surprise that the stages we did not easily discriminate corresponded to late stage 2/early stage 3. On the other hand, these two stages were also the ones that displayed high variability for the expression of three genes of the predictive set: DLD, HNRNPDL and CALM1. Predicting close stages via in situ hybridisation patterns also proved difficult on Drosophila embryos (Ye et al. 2006), evidencing that the reduction of developmental dynamics into timely series of events is difficult. Enlarging our training set in these delicate stages, refining them through sub-stages before array profiling, mining our data sets to extract new predictors for these close stages, adding more genes to the array we used or updating this whole study through the use of RNA sequencing will undoubtedly decrease pre-streak misclassification.

This new staging procedure is robust and only misclassified at late stage 2/early stage 3

The use of an extra-embryonic gene set to stage gastrulating and early neurulating embryos, instead of morphological traits and embryonic markers as classically done, is certainly a new concept. To convince ourselves that using it as a new staging landmark was reasonable, we examined misclassified stages and embryos but also considered biological variability within the gene set patterns.

This staging gene set is connected to patterning genes

Briefly, this gene set was composed of two groups of genes: i) those known to be expressed in mammalian
extra-embryonic tissues or involved in placental, trophoblastic or extra-embryonic functions, such as CITED1 (Rodriguez et al. 2004), TGFβ3 (Gupta et al. 1998, Wyatt et al. 2007) and DLD (Johnson et al. 1997); and ii) those that had not been bound to any extra-embryonic feature (CPA3 and HNRNPDL) or only once (CALM1; GSE1414 in Ushizawa et al. (2004)). Among these, CPA3 is one of the 5 carboxypeptidases identified so far in humans, rats or mice, one of which is imprinted in humans (Singh et al. 2006). CALM1 may be not important by itself but through the Ca²⁺/calmodulin-dependent pathways and HNRNPDL, which belongs to protein complexes that bind pre-mRNA and regulate transcription, may be important in many other tissues.

In silico examination of the networks surrounding these genes suggested why they are useful for predicting embryonic stages, highlighting putative connections between them and the bovine orthologues of the mouse patterning genes. These connections went through extra-cellular signalling pathways (FGF, TGFβ, tumour necrosis factor α (TNF), platelet-derived growth factor (PDGF) and nuclear targets (CTNNB: catenin β, transcription factors)), involving up to five members of the predictive gene set (Fig. 6A). Limiting these connections to direct links did not abolish them but strengthened the FGF and TGF/NODAL signalling pathways, reduced the role of PDGF, left CTNNB as the main nuclear hub but brought HNF4a to light.

Though most of these bovine patterning orthologues were not present on the bovine array we used, we know that most are expressed in elongating extra-embryonic tissues. We reported it earlier for CDX2, ELF5, EOMES and ETS2 (Degrelle et al. 2005) and evidenced it for BMP4 and HNF4a in Fig. 6B. As suggested by its expression profile, BMP4 could have been discriminative among stages when present on the array. Conversely, FGFR2 and DAB2, which were on the array (Fig. 6C), did not appear discriminative in our study (DAB2) or did so with only one statistical method (FGFR2; F test), thus falling out from our selection (Fig. 3A). However, all these genes are expressed in bovine extra-embryonic tissues (including AFP; Maddox-Hyttel et al. 2003). This makes these speculative networks plausible though formal proofs await a jump in the ease of functional invalidations in vivo.

How to use it and why?

In the view of a molecular staging, array data were only a source of discriminative genes and potential predictive ones. We thus do not aim, as in most differential gene expression studies, at validating individual data coming out from the arrays because i) staging is not a quantitative variable and ii) the discriminative capacity of this new trait is not linked to any of the six genes alone but to their use as a set. As the assessment of a primitive streak’s length under a microscope is qualitative, the
assessement of this gene set expression pattern does not need to be quantitative. We indeed evidenced on the test sample that a qualitative PCR assessment is enough to draw a relative expression profile referring to the scheme in Fig. 5A. Owing to the biological variability encountered at late stage 2/early stage 3, a quick examination of the embryo proper will help the discriminative process (Fig. 1).

Interestingly enough, this new staging would allow saving the embryo proper for other studies, such as profiling studies, epigenomic studies or cell culture. Alternatively, based on the results we got with trophoblastic vesicles (TVs), we hypothesise that this molecular landmark could detect other developmental failures or describe atypical phenotypes in suboptimal developmental conditions. To us, the atypical staging that occurred in the TVs, predicting neither stage 4 nor stage 3, did not question the validity of the staging technique but questioned the assumption that embryonic tissues are dispensable for extra-embryonic tissues to elongate (reviewed in Hue et al. (2007)).

**Figure 6** These predictive genes are *in silico* connected to the patterning genes. (A) Each top list of discriminative genes (*n* = 50) was loaded into the Ingenuity Pathway Analysis software (IPA version 6.5) with most of the bovine orthologues of the mouse patterning genes identified today (*n* = 55) and converted into gene networks. Here is the IPA analysis for the bovine genes from the RF selection. Five predictive genes out of six appear on the graph as putatively connected to patterning ones: CITED1, TGFβ3, CALM1, CPA3 and HNRNPD1. All connections were drawn: direct (solid lines) and indirect (dashed lines). The direction of the arrows indicates whether a gene works upstream or downstream of another. In red, the bovine genes; in green, the bovine patterning orthologues; in bold letters are the orthologues, the expression of which was confirmed in bovine extra-embryonic tissues by us (this study or Degrelle et al. 2005) or by other studies.

The function of each gene is represented as proposed by the IPA software for receptors, transcription factors and so on. For details, see the IPA web site (http://www.ingenuity.com/). (B and C) Experimental data confirming the expression of a few bovine patterning orthologues in the bovine extra-embryonic tissues: (B) expression pattern of HNF4a (hepatocyte nuclear factor 4, α) and BMP4 (bone morphogenetic protein 4) as measured by RT-PCR on pools of extra-embryonic tissues (*n* = 5) at each stage (st2, st3, st4 and st5); (C) expression pattern of DAB2 (disabled homologue 2) and FGFR2 (fibroblast growth factor receptor 2) in the array data set, as measured by their relative expression among stages. Each box plot corresponds to five embryos per stage and the developmental stages are st2, st3, st4 and st5.

**How universal is it?**

The last question to be answered for a new staging is the same for every classification system: how universal is it, here that is to say how suitable is it to other species, sheep, pig, rabbit or mouse, for example. This awaits further studies but may be first tested on each species, through a quick assessment by qualitative RT-PCR as we did on the test sample. On the other hand, one could search *de novo* for ovine, porcine or mouse predictive genes to appreciate the generic value of our strategy. Unfortunately, gene expression profiles reported so far on sheeps and pigs were conducted on whole elongating conceptuses (Cammas et al. 2005, Ross et al. 2009), which may confuse the analysis.

Last but not least, knowing that species-specific mechanisms do exist, anatomical equivalents are difficult to define and gene regulation is difficult to transpose (Degrelle et al. 2005, Rieland et al. 2008), it is hard to give an answer on the universal value of gene expression patterns in the absence of an evo-devo approach including out-groups, as illustrated recently (Coolen et al. 2009).
Materials and Methods

**Embryo collection**

Bovine conceptuses from days 14 to 25 post insemination (day of insemination as day 0) were either collected by uterus flushing (days 14–18) or surgically collected after slaughter due to the onset of implantation (days 19–25). For each conceptus, the embryonic disc was dissected out and stored separately from the extra-embryonic tissues. Tissues for in situ hybridisation were fixed whereas tissues for RNA extraction were snap-frozen. Animal use and care were performed in accordance with the International Guiding Principles for Biomedical Research involving animals.

**TV production**

Bovine conceptuses developed after artificial insemination up to day 15 were collected by uterus flushing. The embryonic tissues were surgically ablated while the corresponding extra-embryonic tissues were cut into pieces before being transferred to the uterus of recipient cows. TVs were then collected by uterus flushing at day 18, 3 days after the transfer. Conversely, to the first reports by Heyman et al. (1984) and Flechon et al. (1986), the whole process (recovery–ablation–transfer) was achieved within 2 h. Tissues for RNA extraction were snap-frozen. As control, a ‘false’ vesicle (still harbouring a disc) was transferred to the uterus of a recipient cow and recovered 3 days later. Assessed morphologically, this ‘false’ vesicle had reached stage 4. This control was repeated three times.

**RNA extraction and T7 linear amplification**

Total RNA from AI extra-embryonic tissues \( (n = 20) \) was isolated using Trizol (Invitrogen) and linear amplification was performed using MessageAmp aRNA kit (Ambion, Cortaboeuf, France), starting from 1 \( \mu g \) total RNA as in Degrelle et al. (2008). Two pools of five TVs each were similarly treated for RNA extraction and amplification.

**INRA bovine 10K array**

This array (GPL7417) was partly developed in the laboratory. It contains 7800 cDNA inserts from term placenta (Everts et al. 2005) and 2400 cDNA inserts from extra-embryonic tissues of bovine embryos (D14–D24) as well as young foetuses (D36 and D64). The corresponding libraries are indexed in Unigene as Lib. 3743, 17188, 15992. Finally, 10 214 unique cDNAs and D64). The corresponding libraries are indexed in Unigene bovine embryos (D14–D24) as well as young foetuses (D36 2005) and 2400 cDNA inserts from extra-embryonic tissues of 1000 EST (Dudoit & Fridlyand 2002) was carried out with an RF test at \( P \) value < 0.1. RFs (Breiman 2001): a stabilised version of RF (Bonnet et al. 2008) was applied by combining several RFs. RF outputs an important measurement of all variables based on the way each tree in the forest is constructed and how important is the role of the variables splitting each node of the tree to infer a good classification of the samples (mean decrease Gini measure). This measurement was used to select the most discriminative variables in our study. Optimal feature weighting algorithm (Gadat & Younes 2007) has been specifically developed for highly dimensional data-like arrays, to deal with more than two classes. The main principle in OFW algorithm consists of repeatedly selecting a small subset of variables (genes or EST) and evaluating their predictive ability to rightly classify the arrays into their respective class (here the development stages). This evaluation step is performed via a classification method such as CART (Breiman et al. 1984) or SVM (Vapnik 1999), which outputs a classification error rate based on the subset of variables and on a bootstrap sample of the original arrays. Thus, at iteration \( n \) of the algorithm, important weights \( p_i \) will be given to very discriminative genes, and weights close to zero will be given to irrelevant genes (noisy or uninformative for the classification task), \( i = 1 \ldots G \), where \( G \) is the total number of genes that are spotted on the array. Weights of all variables are then normalised so that \( \sum_i p_i = 1 \) and \( \forall i, p_i \geq 0 \). The next subset of genes to evaluate is then sampled with respect to this weight probability \( p_i \). The evaluation of every possible subset of variables is computationally infeasible and OFW uses stochastic approximations to overcome this problem. Note that at iteration 0, the probability weight is set to the uniform distribution (all variables can be potentially chosen at the first iteration). Both versions of OFW (ofwCART and ofwSVM (Lê Cao et al. 2007, 2009)) were applied to the data set using the R package ofw (Lê Cao & Chabrier 2008).

The data set including the TVs has been used to look for the relative expression of the six predictive genes (CALM1, CITED1, CPA3, DLD, HNRNPDL and TGFβ3) in TV as compared to whole conceptuses. This data set was mean centred and log transformed.

**Gene expression data analysis**

The data set (GSE13013) including the extra-embryonic tissues from whole conceptuses with 10 214 EST was mean centred and log transformed. A basic \( F \) test was used to seek whether at least the variance of one group is significantly different from the other variance groups, with a false discovery rate controlling procedure (Benjamini & Hochberg 1995). A pre-filtering of 2000 EST (Dudoit & Fridlyand 2002) was carried out with an \( F \) test at \( P \) value < 0.1. RFs (Breiman 2001): a stabilised version of RF (Bonnet et al. 2008) was applied by combining several RFs. RF outputs an important measurement of all variables based on the way each tree in the forest is constructed and how important is the role of the variables splitting each node of the tree to infer a good classification of the samples (mean decrease Gini measure). This measurement was used to select the most discriminative variables in our study. Optimal feature weighting algorithm (Gadat & Younes 2007) has been specifically developed for highly dimensional data-like arrays, to deal with more than two classes. The main principle in OFW algorithm consists of repeatedly selecting a small subset of variables (genes or EST) and evaluating their predictive ability to rightly classify the arrays into their respective class (here the development stages). This evaluation step is performed via a classification method such as CART (Breiman et al. 1984) or SVM (Vapnik 1999), which outputs a classification error rate based on the subset of variables and on a bootstrap sample of the original arrays. Thus, at iteration \( n \) of the algorithm, important weights \( p_i \) will be given to very discriminative genes, and weights close to zero will be given to irrelevant genes (noisy or uninformative for the classification task), \( i = 1 \ldots G \), where \( G \) is the total number of genes that are spotted on the array. Weights of all variables are then normalised so that \( \sum_i p_i = 1 \) and \( \forall i, p_i \geq 0 \). The next subset of genes to evaluate is then sampled with respect to this weight probability \( p_i \). The evaluation of every possible subset of variables is computationally infeasible and OFW uses stochastic approximations to overcome this problem. Note that at iteration 0, the probability weight is set to the uniform distribution (all variables can be potentially chosen at the first iteration). Both versions of OFW (ofwCART and ofwSVM (Lê Cao et al. 2007, 2009)) were applied to the data set using the R package ofw (Lê Cao & Chabrier 2008).

The data set including the TVs has been used to look for the relative expression of the six predictive genes (CALM1, CITED1, CPA3, DLD, HNRNPDL and TGFβ3) in TV as compared to whole conceptuses. This data set was mean centred and log transformed.
RT-PCR

Total RNA (n = 15) was extracted from extra-embryonic tissues, using RNeasy Mini Kit (Qiagen) and first strand synthesis was carried out using oligo(dT) primers and Superscript II (Invitrogen) reverse transcriptase. cDNA was amplified with gene-specific primers and 10 μl of each PCR loaded onto 2% agarose gels. Gels were then visualised after scanning (FLA3000; Fuji, Courbevoie, France) and the signals were quantified with the Advanced Image Data Analyser software (AIDA, Fuji). PCR conditions are shown in Table 3.

RT-PCR data analysis

The embryos from the training sample were further used to validate the discriminative capacity of the selected genes. The experimental design was based on nine genes, four stages with five embryos per stage and two RT-PCRs per embryo. Based on the intensities of each amplified fragment, a simple linear model with heterogeneous variances was performed to test the effects of the gene, the developmental stage, the experiment (RT-PCR) and the interactions ‘experiment’ × ‘gene’ and ‘experiment’ × ‘stage’: \( y_{ijk} = \text{gene}_i + \text{stage}_j + \text{experiment}_k + (\text{experiment} \times \text{gene})_{ik} + (\text{experiment} \times \text{stage})_{jk} + e_{ijk} \) where \( e_{ijk} \sim N(0, \sigma^2) \) and \( y_{ijk} \) are the expressions of the EST \( i \) for the stage \( j \) and the experiment type \( k \).

For each gene, the RT-PCR data were normalised to remove the experiment \( \times \) gene and experimental effects. Only the stage and gene effects were kept for the analysis.

Biological network and pathway analysis

Each top list of discriminative genes (n = 50 per list) was loaded with the mouse patterning genes (n = 55 [Pfister et al. 2007, Tam & Loebel 2007, Arnold & Robertson 2009]) into the IPA software (version 6.5) and converted into gene networks, based on the human, mouse and rat database created by the IPA that integrates both functional annotations and bibliographic data of a list of selected genes.

In situ hybridisation

Extra-embryonic tissue sections (10 μM) were hybridised as described earlier [Degrelle et al. 2005] with DIG-labelled riboprobes (Roche Diagnostic). The hybridised sections were observed and photographed using a microscope, a digital camera and software from Olympus. The bovine cDNA fragments encoding CALM1, CITED1, CPA3, DLD, HRNPDL and TGFβ3 were directly amplified from the bacterial clones using specific couples of primers for pTT7Pac: (for sense probes T7/M13R: 5'-TGCTGTGACGGCCGCATTTGT-3'/5'-CACAGAATACGCCAGACCAGAGAATCGCAGTT-3'; for antisense probes: T3/M13F: 5'-GCCCTCGAGGCCGAAGATTTGT-3'/5'-GTAAAAC-3'); 100 ng of purified cDNA were then in vitro transcribed with the proper RNA polymerase.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This work was supported by awards from INRA (no. 20 058 to J-P Renard), the French Ministry of Research (no. 20085 to I Hue and J-P Renard), the EU (SABRE to J-P Renard), USDA CSREES National Research Initiative Grant 2002-35205-11548 and USDA ARS contracts AG 58-1265-2-018 and 58-1265-2-020. The support of the International Relationship Direction of INRA is warmly acknowledged. S A Degrelle was an MNERT fellow. K-A Lê Cao was a PhD student with a shared CNRS/INRA (PHASE-GA) fellowship.

Acknowledgements

The authors gratefully acknowledge D Le Bourhis (UNCEIA, France) for his invaluable help in the production of trophoblastic vesicles by embryonic excision, A Hernandez...
References


Received 6 April 2010
First decision 24 May 2010
Revised manuscript received 28 September 2010
Accepted 6 October 2010