

Spermatozoal transcriptome profiling for bull sperm motility: a potential tool to evaluate semen quality

Nathalie Bissonnette^{1,2}, Jean-Philippe Lévesque-Sergerie^{1,2}, Catherine Thibault¹ and Guylain Boissonneault²

¹Dairy and Swine Research and Development Center, Agriculture and Agri-Food Canada, 2000 College Road, Sherbrooke, Quebec, Canada J1M 1Z3 and ²Université de Sherbrooke, Sherbrooke, Quebec, Canada J1H 5N4

Correspondence should be addressed to N Bissonnette at Dairy and Swine Research and Development Center, Agriculture and Agri-Food Canada; Email: nathalie.bissonnette@agr.gc.ca

Abstract

Regarding bull fertility, establishing an association between *in vitro* findings and field fertility requires a multi-parametric approach that measures the integrity of various structures and dynamic functions, such as motion characteristics, among others. The heterogeneous RNA pattern of spermatozoa could be used in genomic analysis for evaluating both spermatogenesis and fertility potential of semen, mainly because of the static status of the transcriptome of this particular differentiated cell. In a previous study, we determined that some spermatozoal transcripts identified by PCR-based cDNA subtraction are associated with non-return rate, a field fertility index. In the present study, the microarray technology was used in conjunction with differential RNA transcript extraction. We have shown that among these genes, some transcripts are also associated with the motility status of a population of sperm cells fractionated from the same ejaculate. We highlighted a systematic data analysis and validation scheme important for the identification of significant transcripts in this context. With such an approach, we found that transcripts encoding a serine/threonine testis-specific protein kinase (*TSSK6*) and a metalloproteinase non coding RNA (*ADAM5P*) are associated with high-motility status ($P < 0.001$), also confirmed by quantitative PCR ($P = 0.0075$). This association was found only when transcripts were extracted using the hot-TRIzol protocol, whereas the cold-TRIzol RNA extract comprised mitochondrial transcripts. These results demonstrate that some transcripts previously identified in association with field fertility are also found associated with *in vitro* motility provided that a stringent RNA extraction protocol is used.

Reproduction (2009) 138 65–80

Introduction

The primary goal of semen evaluation is to assess ejaculate quality, but it also has the ultimate objective of checking the fertility potential of the sire. To confirm a bull's fertility, breeders begin with a physical assessment of the animal (e.g. testicular volume) as well as a summary evaluation of the semen (e.g. volume and concentration of the ejaculate, sperm morphology). During production, these semen evaluation methods generally use additional ejaculate criteria such as the percentage of live spermatozoa in the ejaculate and the morphology and motility pattern of individual spermatozoa. This common practice is a way to make a diagnostic estimate of both testicular and epididymal function as well as to predict acceptable production yield. However, this approach can detect only major cases of infertility.

Semen quality analysis constitutes a powerful tool to evaluate the fertility potential of males in several species. Visual parameters may not be sufficient for a thorough evaluation of the fertility potential of semen. A number of

in vitro semen analysis methods have been proposed. Many laboratory techniques have been shown to adequately confirm the fertility potential of a semen sample (Rodriguez-Martinez *et al.* 1997, Saacke *et al.* 1998, Larsson & Rodriguez-Martinez 2000, Rodriguez-Martinez 2000, 2003, 2006, Saacke *et al.* 2000, Rodriguez-Martinez & Barth 2007). Laboratory-evaluated parameters of sperm quality include assessment of the integrity of various structures such as genomic DNA (gDNA), acrosome, cell membrane, and mitochondria, in addition to dynamic parameters such as the hyper-activated motility (Mortimer 1990, Davis *et al.* 1992) as well as sperm–oocyte interaction (Madrid-Bury *et al.* 2005, Rodriguez-Martinez 2006). The above suggests that it may be possible to associate specific semen quality criteria with fertility. However, despite the fact that a number of techniques allowing more precise screening in human males are currently available, the high number of patients diagnosed with idiopathic infertility clearly shows that we have only limited knowledge of the processes related to fertility. It appears that certain contradictory results can be attributed not only to the

heterogeneous nature of ejaculate but also to the fact that fertility cannot be defined by a single parameter. An example is provided by the study of the link between fertility and motility. Indeed, it is generally accepted that the motility of spermatozoa is a determining factor in normal male fertility because of its essential role to transport the genetic materials to the site of fertilization. Although the correlation with fertility has not yet been properly defined, most of the statistically significant associations between sperm motility and fertility have ranged between 0.15 and reaching 0.83 in other studies (Amann & Hammerstedt 1993, Kjaestad *et al.* 1993, Bailey *et al.* 1994, Stalhammar *et al.* 1994, Januskauskas *et al.* 2003). Interestingly, the highest correlation ($r^2=0.83$) was measured when the motility assessments were combined with parameters of sperm function (patterns of sperm motion, DNA integrity, and viability; Januskauskas *et al.* 2001). Despite the efforts that have been made, the accuracy of those techniques basically depends on the statistical probabilities arising out of the number of observations made and on how representative is the analysis of the bull fertility (Rodriguez-Martinez 2006).

Recent developments in the field of comparative genomics and global gene expression analysis have provided new molecular detection tools that will allow a number of parameters to be effectively integrated into the evaluation of the fertilizing potential of semen. Thus, the heterogeneous RNA content of a spermatozoon could be used as a pattern for the genomic analysis of semen quality, in terms of both spermatogenesis and fertility potential. The various transcripts that are present in spermatozoa and affect different metabolic pathways or functions raise many questions about the complexity of spermatogenesis and the significance of the transcripts that are generated. We are currently witnessing the emergence of a myriad of diverse transcripts identified using different transcript profiling platforms (Mao *et al.* 2004, Dadoune *et al.* 2005, Zhang *et al.* 2006, Zhao *et al.* 2006, Gilbert *et al.* 2007, Lalancette *et al.* 2008b). In an earlier study, we compiled a list of genes that are differentially represented in the semen within sons of a bull family presenting different fertility status: high non-return rate (NRR) versus a low rate as an index of field fertility (Koops *et al.* 1995, Lalancette *et al.* 2008b). In the present study, we used the microarray approach to determine candidate genes within this subset that may be associated with spermatozoa of greater motility.

For the cattle breeding industry, subfertility is of great concern because it remains a non-predictable measure. Indeed, the selection of sires for species reproduction requires an *in vivo* evaluation of fertility that is carried out over a period that can run from several months to several years. During that period, the semen has already been treated and stored with a view to eventual marketing. The process involves the handling of thousands of ejaculates, the cryopreservation of tens

of thousands of doses, and breeding attempts that all have significant costs. The advent of a molecular tool that would make it possible to distinguish high-fertility bulls from those with lower fertility is still awaited. For the cattle industry, fertility is evaluated mainly by calculating the NRR. The bulls that are retained generally have an NRR that varies from 60 to 80%. A number of methods are used to try to correlate the fertility potential of a sire measured *in vitro* with the NRR in order to reduce the costs associated with subfertility. In a previous study, we assessed whether bulls with a higher NRR had a different sperm transcript profile. The presence of given spermatozoal transcripts can indeed correlate with NRR (Lalancette *et al.* 2008b). In the present study, we used that subset or collection of transcripts to assess whether some of them can be correlated with sperm motility. Although commercial microarrays are useful and have shown the variety of functions that are potentially associated with certain sperm transcripts (Gilbert *et al.* 2007), other transcripts are not represented on these commercial arrays (Lalancette *et al.* 2008b). As a result, we chose to use the cDNA collection on DNA microarrays and demonstrate that other genes, including the genes identified for their differing abundance based on bull NRRs, might be important to include in a spermatozoal transcriptome analysis.

Results

Although spermatozoal RNA is considered to be stored remnants of past spermatogenesis events, we hypothesized that the RNA signature could still reflect the fertility potential of the ejaculate, and motility was used to establish this potential link. To test this hypothesis by global transcript profiling, microarray is one of the best approaches. However, microarray requires substantial amounts of RNA ($\sim 10 \mu\text{g}$) and a standard procedure to maximize yield and quality should this approach become routinely used for fertility assessment. Therefore, we first tested different semen preparation and extraction procedures using commercial extraction kit and published protocols, and compared both yield and quality.

Sperm recovery and viability

In order to isolate motile spermatozoa, two separating methods, swim-up and Percoll gradients, were compared. The swim-up method was more delicate as it preserved the membrane integrity and spermatozoa presented greater motion activity (data not shown) as previously confirmed (Mehmood *et al.* 2008). However, percent recovery of motile sperm was much higher with Percoll separation, an observation also reported in other studies (Kjaestad *et al.* 1993, Mehmood *et al.* 2008). Furthermore, the Percoll gradient allows grading the motility according to the density fraction. We thus

examined spermatozoa stemming from three-phase discontinuous 40–70–90% Percoll gradient prepared as described in Materials and Methods. After centrifugation, contents of the three different layers were examined (visual examination under microscope) for morphology, subjective motility, and viability using a nigrosin–eosin staining technique. As observed previously (Mengual *et al.* 2003), the sperm fractions at the ‘70–90% interphase’ and the bottom 90% fraction presented normal morphology and no difference in viability (data not shown). In accordance with previous work (Mengual *et al.* 2003), a significant difference in viability and evident difference of subjective motility could be detected between these fractions and the first ‘40–70% interphase’ (data not shown). Centrifugation through gradient density increased homogeneity of the preparation as reported (Morrell *et al.* 2008). The viability assays were performed in parallel with a control sample for which spermatozoa were treated with Triton X-100 0.3% for 1 min. This treatment induced a complete loss of viability and motility as observed by visual examination under a microscope of spermatozoa stained with nigrosin–eosin (Materials and Methods). The access of these Triton-treated cells to higher density fractions were impaired by this treatment and cells, thus, laid into the first ‘40–70% interphase’ fraction (data not shown). The stringency of the Percoll gradient was also validated by fluorescent cell sorting method as described previously (Lalancette *et al.* 2008b). Briefly, isolated leukocytes (buffy coat) were added to semen prior to centrifugation and no somatic cell could reach denser fraction (>40%) but kept in the ‘40–70%’ interphase, as evaluated by fluorescent cell sorting using CD2, CD4, and CD8 antibodies as described earlier (Lalancette *et al.* 2008b). The absence of contamination by somatic and germ cells in collected fractions was always confirmed by visual examination of the semen fractions. Furthermore, the absence of contamination was also validated by PCR. The RNA samples extracted from all fractions were reverse transcribed and cDNA tested using specific primers listed in Supplementary Table 1, which can be viewed online at www.reproduction-online.org/supplemental/. An example of a PCR assessment, using these immune and testicular markers cells respectively *PTPRC* and *CD8B* (immune), and *CDH1* and *KIT* for germ cells (Materials and Methods), is presented in Fig. 1A. Both ‘70–90% interphase’ and the ‘bottom 90%’ fractions were always exempt from contamination, containing comparable cell viability, and were therefore assigned as the low- and high-motile fractions respectively.

Evaluation of different extraction protocols

Four extraction protocols were evaluated based on RNA yield and purity, i.e. absence of gDNA: the RNAeasy (Qiagen), the RNA isolation kit (Gentra System), the

TRIZol (Invitrogen), and a hot-TRIZol extraction protocol as previously described (Ostermeier *et al.* 2005, Lalancette *et al.* 2008b). Both RNAeasy and the Gentra System RNA isolation kit required a spin column, a membrane-binding system. We could not get a detectable amount of RNA although the manufacturer’s recommendations were carefully followed. When the amount of spermatozoa was increased, the protocol failed because the columns got clogged or the RNA samples were extensively contaminated with gDNA (Fig. 1B). With either the cold- or hot-TRIZol protocol, a larger amount of bull spermatozoa could be handled and therefore a greater quantity of spermatozoal RNA was extracted. With the hot-TRIZol protocol, a higher RNA yield was obtained and could not be attributed to gDNA contamination. Indeed, all spermatozoal RNA samples were tested by PCR and no amplification were detected (Fig. 1C, lanes 1–6) whereas the reverse transcribed RNA (cDNA) was amplified (Fig. 1C, lanes 7–8). Because microarray studies require micrograms of RNA, it becomes evident that one ejaculate from one bull further fractionated using the discontinuous Percoll gradient centrifugation would not suffice. Indeed, <50 ng RNA was extracted from the bottom fractions of gradients pooled from one ejaculate, using one extraction protocol (i.e. not split; data not shown). The four bull semen samples were pooled and it was thus necessary to amplify RNA due to the low yield in spermatozoa. Many studies have evaluated the possible RNA amplification bias resulting of an exponential or linear amplification. The selection of the proper amplification method must consider both the RNA source and the yield (Patel *et al.* 2005, Zhu *et al.* 2006, Jones *et al.* 2007). We previously showed that an RNA profile was maintained within the range of 200–800 ng of starting spermatozoal RNA using linear amplification (SMART, Materials and Methods) following *in vitro* transcription (Lévesque-Sergerie *et al.* 2006).

This stringent extraction protocol resulted in a better RNA yield but required some assessments of RNA integrity. We used the Chain R gene that was detected as abundant cDNA copies in a previous study (Lalancette *et al.* 2008b); it corresponds to the large subunit of the mammalian mitochondrial ribosome. The impact of the extraction protocol on the integrity of the transcript was evaluated by comparing cDNA copies quantified by real-time RT-PCR (qRT-PCR), targeting both ends of the transcript in order to measure integrity. The respective 5’ and 3’ ends of the ribosomal Chain R gene (1568 nt) are respectively similar to the L_13_B10 and L_05_B02 gene IDs (Supplementary Table 2, which can be viewed online at www.reproduction-online.org/supplemental/), which are part of cDNA libraries generated in previous work (Lalancette *et al.* 2008b). This transcript integrity measured by qRT-PCR was performed on two different RNA preparations obtained from either the cold- or the hot-TRIZol protocol. Assays were performed in triplicate

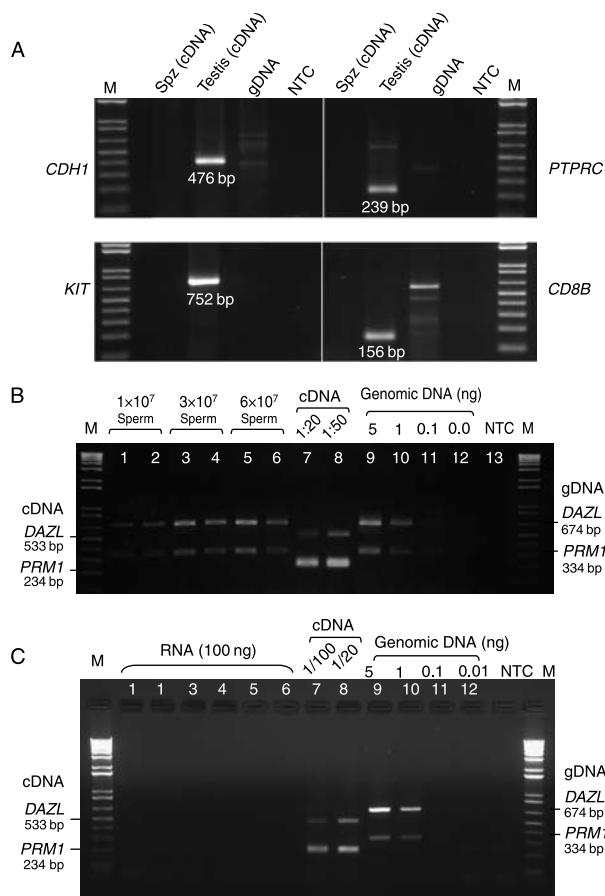


Figure 1 Verification of contamination in the spermatozoal RNA preparations. Somatic cellular, testicular, and genomic contaminations were verified by RT-PCR. (A) Absence of somatic cell contaminant was tested by PCR using an equivalent of 100 ng spermatozoal cDNA for the respective markers *CDH1*, *PTPRC*, *KIT*, and *CD8B* along with testis cDNA and gDNA samples. (B) Verification of contaminating genomic DNA of spermatozoal RNA samples extracted using the Genra System kit. The extraction was performed in duplicate 1×10^7 , 3×10^7 , and 6×10^7 spermatozoa and RNA samples were checked by PCR using *DAZL* and *PRM1* markers along with testis cDNA (dilution 1:20 and 1:50 of 2 μ g equivalent) and a different amount of gDNA, as indicated, for comparison purposes. (C) Purity of hot-TRIZOL RNA extract. The absence of gDNA was confirmed using a multiplex-PCR performed on RNA extracts. Equivalent of 100 ng spermatozoal RNA was tested by multiplex PCR using *DAZL* and *PRM1* markers (lanes 7–8) along with a different amount of gDNA for comparison (lanes 9–12). Ten microliters of the PCR were run on agarose gel electrophoresis. Each target gene was amplified using the appropriate primers (Materials and Methods; Supplementary Table 1). NTC, no template control. M, molecular ladder (1 kb Plus DNA ladder).

on three different fresh semen ejaculates after Percoll preparation (see Materials and Methods). The two qRT-PCR detection assays correspond to amplicons of 180 and 120 bp (Supplementary Table 3, which can be viewed online at www.reproduction-online.org/supplemental/) respectively, which are 846 nt away from each other as confirmed by sequencing (data not shown). qRT-PCR results are presented in Table 1. Whereas the same number of cDNA copies was measured in the

cold-TRIZOL fractions when both ends of the transcripts were targeted, a significant difference was measured in samples prepared with hot-TRIZOL ($P=0.044$). This protocol has a deleterious effect on the 5' end of the Chain R gene transcript.

Microarray experimental design and data analysis

Global transcript profiles have been detected by microarray using the cDNA chip approach. Complete description of arrays is presented in Materials and Methods section and summarized in Table 2. Figure 2A schematizes the semen treatment whereas Fig. 2B schematizes the microarray construction. The whole experimental design is represented in Fig. 2C. Spotting quality was ascertained by TOTO3 staining (data not shown).

We investigated an adequate reference sample according to the 'reference' design approach for microarray. The testicular sample was chosen first based on RNA material accessibility since hundreds of micrograms of total RNA can be extracted from 1 g tissue. Using testis as reference, only gene IDs that present true intensities in both channels (testis and spermatozoa) were kept in the subsequent analysis. This approach would have the consequence of eliminating a substantial amount of gene IDs (Supplementary Figure 1B, which can be viewed online at www.reproduction-online.org/supplemental/) because the abundance of transcripts in the respective tissue presented important relative disparities and gene IDs were thus eliminated using the 'stringent' option of the MIDAS software (The Institute for Genomic Research (TIGR) suite; Materials and Methods), a more statistically rigorous treatment. We decided to perform the hybridizations by comparing the different spermatozoal RNA preparations with one another.

The experimental design was defined using only spermatozoal RNA as a source of probes and for comparing the following treatments: two different Percoll fractions (70–90% interphase and bottom 90%-last fraction) and two different extraction protocols (cold- and hot-TRIZOL), as described in the Materials and Methods section. The design is schematized in Fig. 2C. Briefly, the comparisons made by microarray included three biological samples, corresponding to three ejaculates harvested respectively in weeks 1–3. Although there is rarely a male that produces semen with similar populations (e.g. homogenous ejaculate) at all times, Percoll gradients should attenuate this variability as reported (Mengual *et al.* 2003, Morrell *et al.* 2008). Moreover, by comparing three biological replicates (ejaculates of weeks 1–3), this analysis design took into account the variability between ejaculates and, therefore, significant difference rather relied on the technique being used (gradient fraction and extraction protocol). Furthermore, these comparisons were supplemented by

Table 1 Comparison of the effect of two different extraction protocols on the RNA integrity by real-time PCR quantification.

Clone	Putative identity ^b	GenBank or local database	Query start	Query end	Hit start	Hit end	Score	E value	QPCR: number of gene copies ^a (S.E.M.)		
									Cold-TRIZol	Heated-TRIZol	P value*
L_13_B10	Chain R, large subunit of the mammalian mitochondrial ribosome	Local	1	180	309	488	357	1E-95	521.4	2469.5	0.002
	2FTC_R GI:99032306	NT									
L_07_B02	Chain R, large subunit of the mammalian mitochondrial ribosome	Local	8	404	1120	1516	779	0	422.2	18 990.6	0.003
	2FTC_R GI:99032306	NT									
								P value*	0.623	0.044	

*Statistical analysis was performed using SAS. $P < 0.05$ was considered highly significant.

^aNumber of gene copies detected in aliquots of reverse transcribed samples that are equivalent to 10 ng amplified mRNA. The number of molecules was calculated based on an average quantity measured by spectrophotometry (ABS 260/280) converted in number of molecules. The respective amplicon length generated during real-time PCR is presented in Supplementary Table 3. Copy number was also normalized for *EGFP*, see results Table 3, in order to eliminate the difference in RT efficiency. ^bPutative identity is given regarding the best score according to the results of similarity search again NCBI databases (Materials and Methods). Next columns reports blast results of our sequences for this GI reference sequence.

three technical replicates, corresponding to three hybridization slides per biological replicate, which included dye swap, totaling nine slides per comparison. The hybridization results (gene expression base-2 log ratio) and statistics (*t*-test) for all spot elements are graphically presented in Fig. 3A–C) for comparisons 1–3.

The Volcano plot shown summarizes the microarray analysis. The graph depicts both the difference (expressed in base-2 log ratio of transcript abundance – intensity) between the two samples to be compared (*x*-axis), and the significance level associated to this value (*y*-axis). This significance level was attributed

Table 2 Description of controls, genes, and clones printed on the customary microarray.

Identity of data point	GenBank accession no.	Length (bp ± S.E.M.)	Frequency (Nbr spots)
Control genes ^a			
<i>GAPDH</i>	NM_001034034	299	1 (109)
<i>GNPDA1</i>	NM_001080287	287	1 (218)
<i>NLVCF</i>	XM_593707	307	1 (318)
<i>PRM1</i>	NM_174156	176	1 (529)
<i>PRM2</i>	NM_174157	282	1 (136)
<i>DAZL</i>	NM_001081725	316	1 (128)
<i>TNP2</i>	NM_174200.1	324	1 (96)
Reporter genes			
<i>EGFP</i> (pEGFP-N1)	U55762	290	1 (359)
<i>LUC</i> (pGL2)	X65326	318	1 (517)
<i>Arabidopsis thaliana</i>			
<i>PRN</i> ^b	NM_115784.2	319	1 (394)
<i>ATBI-1</i> ^c	NM_124083	314	1 (287)
<i>At4g02220</i> ^d	AY056352	300	1 (143)
<i>PDF1.2</i> ^e	AY133787	289/274	2 (585)
<i>K22G18.2</i> ^f	AY062824	310	1 (718)
Libraries ^g			
NRR-SSHf	–	490 ± 209	1776 (3552)
NRR-SSHL	–	435 ± 165	1248 (2496)
NRR-ASSHF	–	323 ± 92	171 (342)
Fertile bulls (F)	–	332 ± 124	96 (192)
Subfertile bulls (L)	–	326 ± 138	96 (192)
Total			11 311

^aThe respective similarity search results obtained from the different public databases are presented in details in Supplementary Table 2. ^b*PRN*: clone 177B6 – calmodulin binding. ^c*ATBI-1*: clone 186K17 – *Arabidopsis* Bax inhibitor 1. ^d*At4g02220*: clone #U09503 – putative zinc finger protein. ^e*PDF1.2*: clone 37f10 (289 pb) and u11912 (274 pb) – putative antifungal protein (two different fragments with overlapping sequences). ^f*K22G18.2*: clone u17022 – copin-like protein. ^gComplete description of the cDNA libraries is presented in Materials and Methods section. Libraries NRR-SSHf, NRR-SSHL, and NRR-ASSHF were normalized whereas F and L were not.

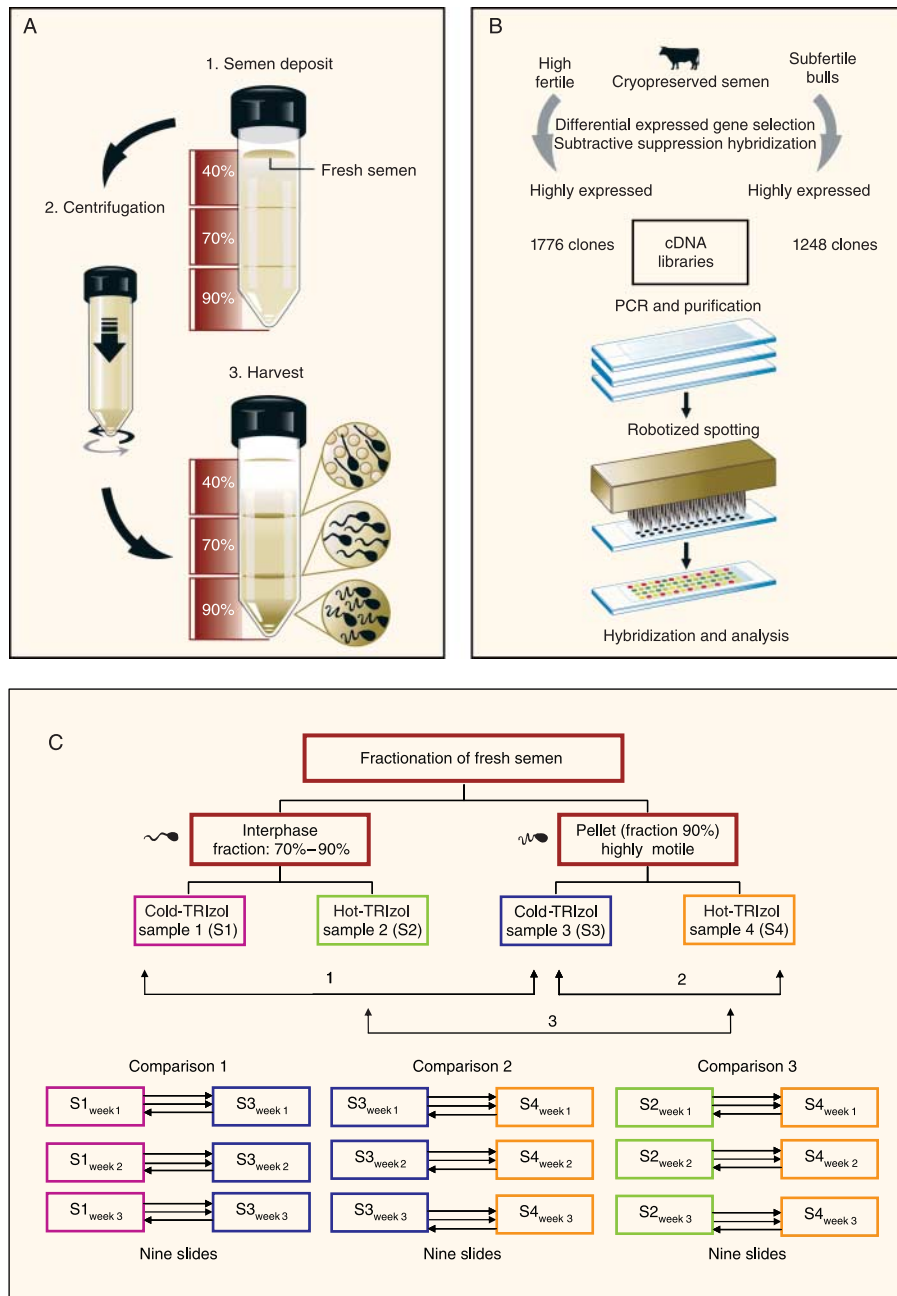


Figure 2 Experimental layout. (A) Fresh semen was fractionated by centrifugation on discontinuous Percoll gradient containing three phases: 40, 70, and 90%. (B) The cDNA chip comprised two cDNA libraries constructed using suppressive subtractive hybridization (SSH), which has been successfully applied to spermatozoal transcript profiling. (C) Microarray analyses were performed with one of the three RNA samples prepared on different days (biological replicates) whereas comparisons were performed on different fractions harvested from the same gradient (i.e. same ejaculate). All microarrays were processed at the same time. Arrows represent dye swap (i.e. beginning of the arrow represents cy3 and arrow cy5; reverse arrow represents the dye swap).

to each gene 'ratio', tested using the *t*-test, the null hypothesis indicating no difference between the two samples. This value is reported on the *y*-axis. Statistically interesting genes should present more than a twofold difference between both samples. In other words, genes harboring a 'gene expression mean value' superior to 1 or inferior to -1 are selected. The Volcano graphs are presented along with an inset graph that represents the same hybridization results but analyzed using the 'generous' option and not filtered for the negative controls threshold value ($2 \times \text{s.d.}$, see Materials and Methods). All data were normalized for LOWESS to

render the gene expression level from the two different dyes comparable (Quackenbush 2002, Yang *et al.* 2002, Morrison & Hoyle 2003).

In the first comparison (Fig. 3A), the '70–90% interphase' fraction of the Percoll gradient was compared with the 'bottom 90%' fraction and no significant gene ID was found when RNA was extracted using the cold-TRIZol method ($P > 0.001$, *y*-axis of the Volcano plot). In other words, using this cold-TRIZol protocol, no difference in the transcript profiles could be detected between low- and high-motile spermatozoa. However, the extraction protocol influenced the RNA

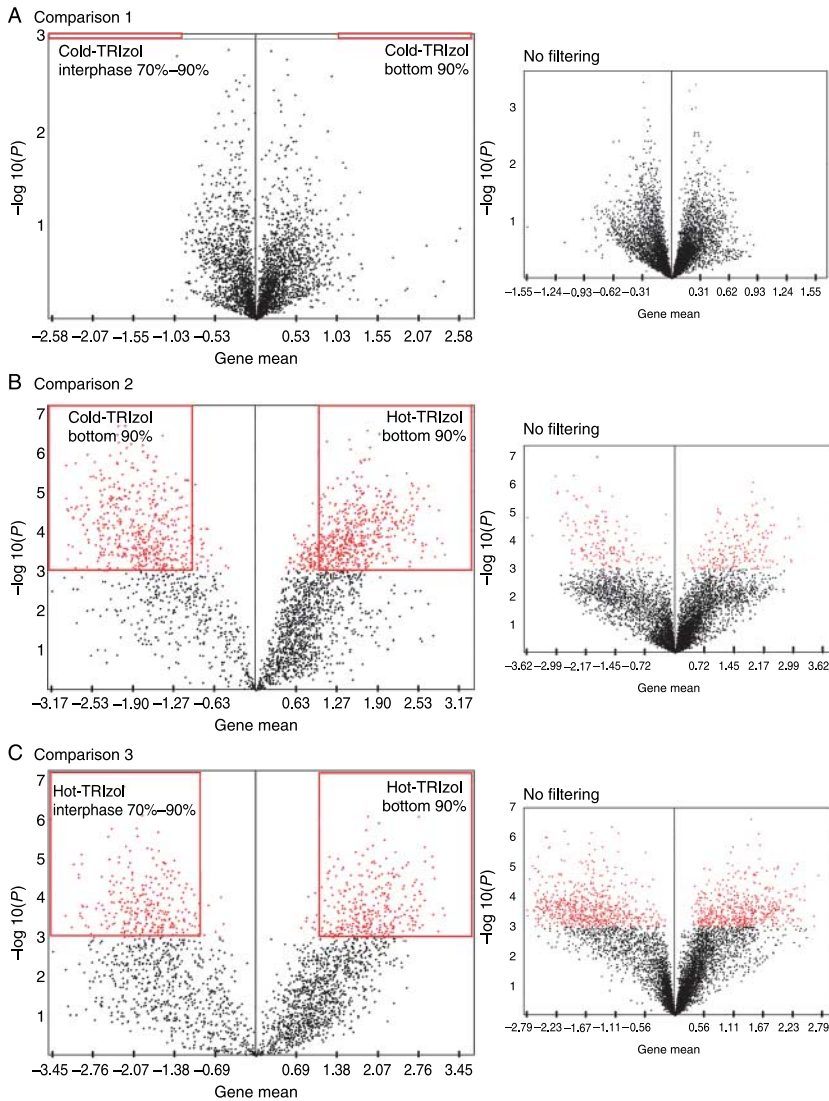


Figure 3 Volcano plot of the hybridization results for the three comparison assays of different spermatozoal RNA preparations. The hybridization results present the three block assays (A–C), which refer to comparisons 1–3 respectively, that followed a design schematized in Fig. 1C. In each graphical representation (Volcano plot) is plotted the intensity regarding a specific gene given by the two samples to be compared in a dual-hybridization assay. This intensity is presented as a ratio of the respective intensity. Each point corresponds to a transcript abundance ratio (expressed as a mean value of base-2 log ratio, x-axis) obtained from the combined hybridization of the compared spermatozoal samples on the microarray. Each gene expression ratio is plotted on the x-axis, while the corresponding statistical P value is presented on the y-axis. Each gene ‘ratio’ has been statistically tested by a t -test, the null hypothesis indicating no difference between the two samples. This t -test was applied on nine hybridization results (slides) and a statistical value was obtained for each gene and reported on the y-axis. Significant genes are marked as red dots and framed into the red box of the Volcano graph indicating a P value < 0.001 , which corresponds to $-\log_{10}(P \text{ value}) > 3$. The Volcano graphs presented (inset) correspond to the respective hybridization results but analyzed using less stringent parameter. By this ‘generous’ option of the ‘Channel Tolerance’ parameter, a gene value was not discarded because the intensity was too low for one or the other spermatozoal sample (called ‘absent’).

profile depicted. In the second comparison, when both cold- and hot-TRIZol protocols were compared on the same ‘bottom 90%’ fraction, split into two parts, several RNAs were differentially expressed (Fig. 3B). Whereas a broader gene diversity was found among the significant genes in the hot-TRIZol fraction (\log_2 intensity > 1 , x-axis of the Volcano plot), the significant gene ID (i.e. significant spot elements of the microarray) associated to the cold-TRIZol fraction (\log_2 intensity < -1 , x-axis of the Volcano plot; Fig. 3B) was essentially mitochondrial ribosomal gene transcripts (data not shown).

While the main goal of this work was to determine whereas spermatozoa presenting different motility status also differed in their RNA profile, such assumption was rejected if the cold-TRIZol protocol was used (comparison 1, Fig. 3A). However, if the hot-TRIZol protocol was applied (third comparison, Fig. 3C), the microarray analysis revealed that sperm preparation with different motility status presented different global transcript profiles.

To confirm these results, real-time qRT-PCR assays were performed to monitor the absolute cDNA copies, i.e. the RNA contained in spermatozoal fraction for a selection of gene ID (i.e. significant spot elements of the microarray).

Validation of significant transcripts

In order to be classified as an interesting gene for the qRT-PCR study, both spot elements – each clone (gene) was spotted in duplicate on each slide – had to be significant in the microarray analysis ($P < 0.001$) across the nine hybridized slides. In other words, to be selected, the respective statistical result of both spots for a given transcript had to be significant. The selected gene IDs were sequenced (both complementary strands, data not shown) and a consensus sequence was assembled using MegAlign software allowing primers design for quantitative PCR (qPCR). We noticed that some gene IDs harbored repetitive sequences (e.g. SINE elements) and were eliminated.

Table 3 Validation by real-time PCR for clones selected by microarray analysis.

Clone	Putative identity ^b	Microarray		Quantitative real-time PCR		
		Gene mean ratio ^c (s.e.m.)	Number of gene copies ^a in RNA extract (inf. born (CI); sup. born (CI))	Interphase fraction 70–90%	Bottom Percoll fraction 90%	Amplification fold ^d
L_02_C06	Unknown (LOC617010 similar to ADAM5P (<i>Bos taurus</i>))	Significant 2.44 (0.30) $P=0.0002/\text{Adj.}=0.037$		Significant 196 (58; 660) $P=0.00075$	6863 (2547; 18 491)	1.48
RUR_C08	T5SK6 (LOC539707 similar to serine/threonine protein kinase (<i>Bos taurus</i>))	1.52 (0.36) $P<0.001/\text{Adj.}=0.014$		140 (90; 201) $P=0.00075$	1487 (1078; 2051)	1.78
L_04_G12	Unknown similar to the secreted protein ALLW1950 (<i>Homo sapiens</i>)	ND		ND	521 (380)	–
PRM1	Protamine 1 (<i>PRM1</i>)	–		72 (24; 216) $P=0.00075$	20 238 (8277; 49 483)	13.73
PRM2	Protamine 2 (<i>PRM2</i>)	–		11 (3; 36) $P=0.0151$	268 (98; 736)	1.26
F_18_F05	Unknown (chromosome 27 (<i>Bos taurus</i>))	Not significant (gene mean) 0.99 (0.21) $P=0.0033/\text{Adj.}=0.03$		Significant 2063 (1883; 3599) $P=0.00074$	107 347 (74 073; 155 569)	1.52
F_16_B08	Unknown (<i>Bos taurus</i> hypothetical protein LOC100125878 mRNA)	Not significant/significant one spot on two Spot 1: 1.89 (0.17) $P=0.0002/\text{Adj.}=0.0036$; Spot 2: 1.6 (0.25) $P=0.0079/\text{Adj.}=0.17$		Not significant 129 563 (38 664; 434 162) $P=0.5274$	337 326 (125 675; 905 420)	0.08
F_18_H08	Unknown (<i>Bos taurus</i> similar to limb expression 1 (MGC142689))	Not significant (both spots) 1.46 (1.33) $P=0.0174/\text{Adj.}=0.388$		Not significant 115 176 (43 388; 305 742) $P=0.9538$	201 795 (90 934; 447 815)	0.04
RT-PCR efficiency						
<i>EGFP</i> ^e	Enhanced green fluorescent protein	NA		208 872 (164 836; 264 671) $P=0.0003$	117 310 (90 253; 152 479)	–

NA, not applicable; ND, not detected.

^aNumber of gene copies detected in aliquots of 1 µg reverse transcribed amplified mRNA. The number of molecules was calculated from a calibration curve established using serial dilutions of the gene amplicons (Materials and Methods). Results in parentheses are confidence intervals. Copy number was also normalized for *EGFP* in order to eliminate the difference in RT efficiency. ^bPutative identity is given regarding the best score according to the results of similarity search against NCBI databases (Materials and Methods). The acronyms refer to the HUGO gene nomenclature when available, but refer to the *Bos taurus* ortholog, unless stated. ^cMicroarray results are reported for each gene, which comprise the mean of the ratio of the base 2 log-transformed intensity of hybridized spots (e.g. ratio of the compared RNA extracts). ^dThe relative mRNA expression levels are given by the arithmetic formula $2^{-\Delta\Delta C_t}$ (Livak & Schmittgen 2001) where the C_t of the target gene (for both samples compared) are normalized with the housekeeping gene value detected in the housekeeping gene value detected in the respective sample (ΔC_t). The difference between the two samples to be compared is calculated with the normalized C_t , i.e. ΔC_t and $\Delta\Delta C_t$ are then converted into amplification fold using the $2^{-\Delta\Delta C_t}$ formula since each $\Delta\Delta C_t$ value corresponds to a twofold PCR amplification. ^e*EGFP* detection assay is used to standardize the RT efficiency of the respective assays. An amount of 2.6×10^6 mRNA copies of the *EGFP* transcript is spiked in all RT as described (Levesque-Sergerie *et al.* 2007).

We investigated two candidates by real-time qRT-PCR that presented significant differential expression as revealed by the microarray results ($P < 0.001$): gene IDs RUR_C08 and L_02_C06 (Table 3). The real-time RT-PCR results confirmed the differential transcript abundance: gene ID L_02_C06 and gene ID RUR_C08 were respectively 35 and 11 times more abundant in the 'bottom 90%' fraction than in the '70–90% interphase' fraction. Sequence similarity searches against the RefSeq_RNA database of NCBI indicated that gene ID RUR_C08 is similar to a serine/threonine testis-specific protein kinase (*TSSK6*; Supplementary Table 2). The sequence of gene ID L_02_C06 is similar to a metalloproteinase as predicted by automated computational analysis (annotated genomic sequence NW_001494418) and also presents high similarity to a testis cDNA sequence of *Canis familiaris*. The L_02_C06 sequence is similar to a multidomain protein member containing a metalloproteinase-like domain, a disintegrin-like domain as well as a cysteine-rich domain (the MDC protein family), which proteins are expressed abundantly in testis (Smith *et al.* 2002, White 2003). The similarity search results obtained from the different public databases are presented in detail in Supplementary Table 2.

In parallel, we tested whether the microarray analysis and the arbitrary selected twofold gene threshold were adequate; this involved determining to what extent the microarray analysis limited false negative results. To be considered significant based on microarray analysis and to be selected for further analysis, the following three parameters needed to be met: 1) base-2 log ratio (> 1 or lower than -1), 2) duplicate spot elements across the nine slides, and 3) *t*-test statistics for overall- α (critical *P* value < 0.001). We chose one gene ID that failed in one of three categories: F_18_F05 (base-2 log ratio = 0.99), F_16_B08 (only one significant out of two: clone spotted in duplicate), and F_18_H08 (both gene IDs were not significant). The putative identity of these genes corresponds to a hypothetical protein (Supplementary Table 2). The real-time qRT-PCR results for these gene IDs are shown in Table 3. The gene ID F_18_F05 is representative of a false-negative result since this transcript, as measured by real-time qRT-PCR, was more abundant in the 'bottom 90%' Percoll compared with the '70–90% interphase' fraction ($P = 0.0003$; Table 3). This gene ID was rejected based on twofold selection (log₂ ratio of 0.99; Table 3). The following gene IDs were not significant in either microarray and real-time qRT-PCR: F_16_B08 was only significant for one of the two spot elements and therefore did not qualify, which was also confirmed by real-time qRT-PCR ($P = 0.8788$); F_18_H08 was not significant for both spot elements in microarray analysis or in real-time qRT-PCR ($P = 0.7715$; Table 3), which validated the negative result. Microarray analysis was thus considered rigorous.

Several gene IDs were not retained and they were denoted as 'absent' because either the normalized intensity value of one of the two spots was not $2 \times \text{s.d.}$ above the surrounding background intensity, or because the spot elements did not satisfy the filtering threshold requirement set at two times the mean value of negative controls (see Materials and Methods section for data analysis). This was the case for gene ID L_04_G12, which was found to be associated with NRR (Lalancette *et al.* 2008b), a field fertility index, and which was only detected in the 'bottom 90%' Percoll fraction extracted with hot-TRIzol (Table 3 and data not shown). As well, both protamine *PRM1* and *PRM2* genes were denoted as 'absent' in the microarray analysis although they could be detected by real-time qRT-PCR assay (Table 3). Both *PRM1* and *PRM2* genes were found to be differentially represented, being 280 and 26 times more present in the 'bottom 90%' versus the '70–90% interphase' fractions respectively (Table 3). These results were explained by the very few cDNA copies detected by qRT-PCR in the '70–90% interphase' fraction for both *PRM1* and *PRM2* transcripts (mean copy number < 100 per qPCR). Indeed, a closer look at the microarray results (integrated intensity) confirmed that intensities were slightly above background but not above the acceptable value ($> 2 \times \text{s.d.}$), thus these gene IDs were eliminated in the downstream analysis. The microarray detection is robust and allows for a considerable throughput analysis; however, it is not as sensitive as qPCR.

Discussion

The long-term objective of this work is to identify transcripts whose quantitative expression predicts motility potential of a bull ejaculate combined with markers that also hold promise to be linked with sperm fertility. In previous works, we have depicted the transcriptome by microarray using commercial human microarrays. This survey confirmed that spermatozoa harbor a complex mixture of messengers implicated in a wide array of cell functions (Gilbert *et al.* 2007). In a more recent study, we have created a cDNA library made of transcripts differentially represented between highly fertile and subfertile bulls (Lalancette *et al.* 2008b). Because some of these genes are not represented in commercial arrays, we constructed custom microarrays using these cDNA libraries. Among the 180 fertile sons of one bull sire, extreme bulls were selected to create this inventory of transcripts that present difference in abundance between the two groups: five high fertile and five subfertile bulls. Notwithstanding that this custom array was not a complete representation of the spermatozoal transcriptome, we identified novel transcripts whose abundance was also associated with motility. It confirms that commercial array needs improvement in order to depict a closer representation of spermatozoal transcriptome.

For decades, efforts have been put in the determination of the best parameters to be used to predict the fertility potential of spermatozoa. Recently, the community of reproductive biologists has reached the conclusion that fertility is a multi-factorial process. It thus requires a multi-parametric assessment, which includes the evaluation of morphology, motility, acrosome reaction, and genome integrity, to reach a predictable level. In other words, prediction can be achieved when several assessments are performed (multi-parametric evaluation and repetition – including large sperm counts). Furthermore, the approach might consider that ejaculate is a heterogeneous mixture of haploid cell population harboring genetic diversity. This heterogeneity is now a recognized fact and we also observed that this heterogeneity was conveyed in the transcript abundance of a specific gene among different sperm cells. Indeed, we observed that the level of a specific transcript vary among a heterogeneous spermatozoal cell population (see Supplementary Figure 2, which can be viewed online at www.reproduction-online.org/supplemental/). There is, undeniably, a difference in terms of transcript level among spermatozoa. If the transcript profile is a reflection of past events (e.g. spermatogenesis) and sperm dynamics (e.g. motility), this transcript profile assessment might be considered in itself as a multi-parametric tool. Furthermore, spermatozoal transcript profiling of semen sample could be considered as a non-invasive approach avoiding biopsies. Thus, this perspective would have interesting clinical applications for the identification of the idiopathic cause of infertility or subfertility and to grade semen quality. To test this hypothesis, motility was the first parameter evaluated since large quantity of semen could be processed by sedimentation. To discriminate motility status, we prepared semen using the discontinuous Percoll gradient centrifugation since this processing increased homogeneity of the sperm preparation based not only on motility, but also on morphology and DNA integrity (Tomlinson *et al.* 2001, Mengual *et al.* 2003). Precisely, denser fraction of the Percoll gradient not only increases the total motility status of the different fractions but also discriminates spermatozoa following motion characteristics like linearity, progressive, and grade motility (e.g. grade A \rightarrow 25 μ m/s, World Health Organization; Mengual *et al.* 2003, Kliesch & Cooper 2008). The specific objective of this study was to verify whether reliable biomarkers could be identified among the gene transcripts isolated from bovine spermatozoa as previously reported (Lalancette *et al.* 2008b). The global transcript profiling by microarray produces an overview of past events, mainly because the transcriptome of this particular differentiated cell is under a static status. It is tempting to speculate that the genetic content, not only the DNA integrity, could be used as a hallmark of a successful spermatogenesis process and depict the potential fertility of semen sample. However, as

microarray technology presents some limitations, a careful approach must be undertaken as outlined in this manuscript. Most importantly, a closer look at the representation of transcripts present in this particular cell is needed. As a first limitation, microarrays can provide information only about the genes that are included on the array resulting in a limited overview of the spermatozoal transcriptome. For example, we searched similarity using the BLAST algorithm of NCBI among our cDNA sequences against the commercial bovine genome array downloaded on our server. Several spermatozoal transcripts are said to be absent (data not shown) and among those, both *ADAM5P* (L_02_C06) and *TSSK6* (RUR_C08) gene sequences are not represented. Therefore, it confirms the necessity of using DNA chips that reflect the diversity of spermatozoal transcripts.

The second limitation concerns gene abundance. Several spermatozoal genes were found by microarray; however, some significant genes were also missed, including L_04_G12, a gene associated with bull fertility (Lalancette *et al.* 2008b), which was not found significant by microarray. The L_04_G12 transcript was present only in the highly motile fraction, as validated by qRT-PCR, and found absent both in the lower motile extract and in the RNA samples extracted by cold-TRIZol. In other words, two main reasons explain why this gene ID was not retained in the microarray analysis: the extraction protocol and/or the limited absolute abundance of a transcript. Indeed, the presence of the transcript above a certain detection level is mandatory. As detected by the highly sensitive qRT-PCR molecular assay, both *PRM1* and *PRM2* were nearly at the limit of detection (<100 copies per qRT-PCR assay) in the low-motile fraction. Indeed, the microarray analysis excludes all the weak abundant transcripts unless the expression is above a significant threshold, which is calculated using an algorithm that takes into account the intensity of the surrounding background area. In other words, the so-called 'on/off' genes qualified as low abundant transcripts are excluded during the analysis.

Since spermatozoa are specialized cells with a low RNA content, it is of prime importance to recover as much sample as possible. As micrograms of cDNA are required in probing microarrays, efficient RNA extraction and amplification methods are required not only to generate sufficient amount of hybridization target for completing a dual-dye microarray design, but also to depict a relevant transcript profile. There is no doubt that different extraction protocols (mild and stringent) yield different transcript profiles, a phenomenon that was previously observed (Miller & Ostermeier 2006b). As seen for the first comparison (Fig. 3), no difference in transcript profile between high- and low-motile sperm was observed using the cold-TRIZol protocol (Fig. 3A). The depicted analysis remains partial since the cDNA library used to create the custom microarrays contains

only genes that present a difference between bulls that differ in their NRR, a field fertility index. Nevertheless, it appears that the cold-TRIZOL protocol essentially provided a superficial overview of the transcriptome, mainly consisting of mitochondrial ribosomal gene transcripts. Interestingly, a broader gene diversity was found when the hot-TRIZOL was applied to compare high- and low-motile sperm (Fig. 3C) although RNA integrity is affected (Table 2). This is the case of L_02_C06 and RUR_C08 transcripts only detected when hot-TRIZOL protocol was applied. Our custom microarray incorporated genes of three normalized cDNA libraries that were created using RNAs extracted according to the hot-TRIZOL protocol. We trusted that RNA degradation bias is also integrated in the array, i.e. transcripts resistant to degradation are also represented on the microarray. As mentioned above, microarray results from comparison 2 clearly depict the restricted transcripts diversity of sample extracted with the cold-TRIZOL protocol, which were essentially mitochondrial ribosomal transcripts, confirming the relevance of using the hot-TRIZOL method to extract transcripts that otherwise would not be detected.

Since these L_02_C06 and RUR_C08 transcripts were highly associated to motility status (35- and 11-fold in the high- versus low-motile fraction respectively), we investigated whether this novel 'unknown' transcripts (L_02_C06) could be detected by *in situ* hybridization in spermatozoa. Preliminary *in situ* hybridization of the L_02_C06 probe revealed that the spermatozoon contains this specific transcript principally in two regions of the cell: a major signal co-localized with the nuclei and a second signal located near the centriole (Supplementary Figure 2). Similar pattern that emerged from the centriole region and the gDNA if the spermatozoon was decondensed with dithiothreosulfate (DTT) has been reported elsewhere (Miller & Ostermeier 2006a). The L_02_C06 signal appeared like a shadow, co-localized with spreading DNA, which induced an enlargement (swelling) of the head as observed in time course DTT treatment (data not shown). Because the *Bos taurus ADAM5P* gene sequence is located at a unique loci on the haploid genome, this increased signal cannot emerge from the RNA/DNA hybridization as confirmed by the absence of signal obtained with the probe sense (Supplementary Figure 2). On the higher structural level, it has been proposed that chromatin in mammalian sperm is organized into loop domains attached at their bases to a nuclear matrix (Ward & Coffey 1991, Yaron *et al.* 1998, Ward & Ward 2004). Looped organization of sperm DNA is disputed in other works (Sanchez-Vazquez *et al.* 1998) but the presence and localization of specific RNA within or close to the sperm nuclear matrix have recently been demonstrated (Lalancette *et al.* 2008a). It is thus logical to speculate that a more drastic hot-TRIZOL protocol was required to show up transcripts enclosed into more complex and less accessible chromatin organization.

Interestingly, the L_02_C06 transcript has not yet been reported (no EST deposited by 19 November 2008). The only identity being predicted by bioinformatics is similar to the *B. taurus ADAM5P* gene (Matched accession no: NR_001448.1). Both L_02_C06 and NR_001448.1 sequences are located solely on the respective genomic emplacement. In addition to being expressed in testis, we found this *ADAM5P* transcript expressed in bovine liver and heart tissues but found it to be undetectable in muscle, lung, ovary, spleen, or kidney (data not shown). The presence of this non-coding spermatozoal bovine *ADAM5P* transcript a pseudogene that would encode an enzyme harboring a metalloendoproteinase activity, is intriguing. The next step would be to understand the biological meaning of this transcript in spermatozoa, which encodes a metalloproteinase that catalyzes the hydrolysis of nonterminal peptide linkages in oligopeptides or polypeptides. If the corresponding protein is demonstrated to be present, these proteases could proteolytically modify cell surface and extracellular matrix in order to alter cell behavior. Indeed, the protein encoded by this putative bovine transcript (similar to accession no: NR_001448.1) contains a conserved domain known as ADAM (membrane-anchored protein with a disintegrin and a metalloprotease domain). This cysteine protease is detected in early *Xenopus* embryos (Smith *et al.* 2002). When active, this protein retains a variety of functions by cleaving and remodeling components of the extracellular matrix, including potentially releasing growth factors, cytokines, cell adhesion molecules, and receptors (White 2003). These activities might be important for fertilization or embryo survival since it has been demonstrated that spermatozoal RNA can be transferred into the embryo following fecundation (Ostermeier *et al.* 2004).

Before it is possible to have a reliable representation of spermatozoal transcriptome, some technical considerations are required as outlined in this paper. It has been revealed that differences in extraction stringency could lead to different transcript profiles. A reliable RNA extraction could open up possibilities to use an extensive transcriptome analysis to study the fertility potential of semen sample and to potentially extrapolate the male fertility. The number of diverse applications of DNA microarray technology has permitted the simultaneous study of thousands of genes as a means of identifying differentially expressed genes. Thus, the microarray analysis finds applications as long as the genes of interest are present on the DNA chip. In this study, among the spermatozoal transcripts associated with bull NRR (Lalancette *et al.* 2008b), a field fertility index, some transcripts were found associated with motility. Some of them were not present on commercial bovine DNACHIP. Thus, further consideration should be put in getting a better knowledge of the transcriptome content of the spermatozoa, first, in order to propose commercial microarray that comprises this diversity when studying

the spermatozoal transcriptome. The next question to be raised concerns not only the role of the transcript found in spermatozoal cells but rather if using other parameters, strong correlations could be established through a microarray analysis. This paper therefore establishes the sperm RNA transcriptome a potential multi-parametric tool for a predictive evaluation of fertility allowing the identification of markers specific for the different parameters. In this study, motility was investigated as one such key parameter.

Materials and Methods

Semen collection and Percoll gradient

Fresh semen ejaculates from four different bulls were sampled weekly at the Centre d'Insémination Artificielle du Québec (CIAQ, Sainte-Madeleine, Quebec, Canada), for a period of three consecutive weeks and constitute the three biological replicates. Bulls were handled according to the Guide for the Care and Use of Agricultural Animals by the Institutional Animal Care and Use Committee of Agriculture and Agri-Food Canada's Dairy and Swine Research and Development Center (document #160, Oct 2002). Semen was collected with an artificial vagina and all ejaculates were examined individually by routine laboratory measurements – (ejaculate volume, sperm concentration, and motility rate) to ensure that samples present similar characteristics. Progressive motility and concentration were assessed before and after sperm separation. The sperm concentration was $>20 \times 10^6/\text{ml}$, with motility $>65\%$ and containing $<1\%$ immature germ cells and no lymphocyte following Coomassie Blue G-250 and nigrosine-sin staining method. The latter staining procedure allows to readily visualize the paddle-shaped heads of the sperm morphology whereas acrosome integrity was evaluated by Coomassie Blue G-250 staining method to assess membrane integrity as described elsewhere (Ramirez *et al.* 1992, Mehmood *et al.* 2008). Visual motility was evaluated under $400\times$ magnification with a phase contrast microscope. Sperm concentration was estimated using a hemocytometer. The four bull semen samples were pooled and a maximum of 2 ml aliquot was overlaid on each three-layer (40–70–90%) discontinuous Percoll gradient (Amersham) in 15 ml conical plastic test tubes (Fig. 2A). The Percoll layers were prepared by diluting Percoll solution in HEPES buffered Tyrode's lactate (TL) sperm medium (100 mM NaCl, 25 mM NaHCO_3 , 10 mM HEPES, 0.3 mM NaH_2PO_4 , 3.11 mM KCl, 0.035 mM sodium lactate, 2 mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ and 4 mM $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$; pH 7.3; Parrish *et al.* 1995). The spermatozoa and gradient were then centrifuged at 700 g for 45 min at room temperature. After centrifugation, the 0.5 ml fractions were carefully collected: the 70–90% Percoll interphase fraction (low motile) and the 90% Percoll bottom fraction (high motile; Fig. 2A). They were collected from each gradient and washed twice in five volumes of TL-SPERM solution by centrifugation at 250 g for 5 min to remove the Percoll. Finally, the pellets were resuspended in 150 μl TL-SPERM solution and kept at -80°C until RNA extraction.

RNA extraction

During preliminary experimentations, four different protocols were tested: the RNeasy extraction kit (Qiagen), the Genra System (Inter Medico, Markham, Ontario, Canada), the TRIzol (Invitrogen) as reported (Lalancette *et al.* 2008b), and a 0.3% Triton X-100 (Sigma–Aldrich) treatment, as reported (Ostermeier *et al.* 2005). Both RNeasy and Genra System RNA extraction protocols were performed as recommended by their respective manufacturers. Precisely, the hot-TRIzol protocol consisted of a lysis of 30 min incubation at 60°C with vortexing every 10 min. Following steps were similar to the cold-TRIzol protocol, which was performed as recommended by the company. After treatment with RNase-free rDNAse I (Ambion), samples were precipitated and pellets were resuspended in Nuclease-free water (Ambion) containing 1 U/ μl SUPERase In (Ambion). RNA was quantified spectrophotometrically at 260 nm with a NanoDrop ND-1000 spectrophotometer. Following each extraction and DNase treatment, a multiplex PCR designed for the bovine deleted in azoospermia-like (*DAZL*; accession no. NM_001081725) and protamine 1 (*PRM1*; accession no. NM_174156.2) gene, both spanning an intron, was performed on RNA to check for contamination by gDNA. Primer sequences used are listed in Supplementary Table 1. Each reaction contained 500 nM *DAZL* primers, 200 nM *PRM1* primers, and Multiplex Master Mix (Qiagen) to $1\times$ concentration. Thermal cycling conditions included an initial denaturation step at 95°C for 15 min followed by 35 cycles at 94°C for 30 s, annealing at 55°C for 90 s, extension at 72°C for 1 min, and a final extension step at 72°C for 10 min. Each assay included dilutions of bovine gDNA, bovine testicular cDNA, and water as controls. PCR products were then analyzed by DNA electrophoresis on a 1.5% agarose gel.

Evaluation of sperm cell preparation by RT-PCR

To validate the absence of a possible contamination of spermatozoa preparation by somatic cell, an RT-PCR approach was used. After extraction, 150 ng RNA was amplified using the SMART mRNA Amplification Kit (Clontech) following the manufacturer's protocol. Then, 250 ng amplified mRNA were reverse-transcribed with SuperScript II (Invitrogen) using oligo dT_{12–18} (Invitrogen). PCRs were then performed in a 50 μl volume. Oligonucleotides (Operon Biotechnologies, Huntsville, AL, USA) were designed on different exons when information was available. Primers sequences for the *B. taurus* E-cadherin (*CDH1*; accession no. NM_001002763), v-kit oncogene homolog (*KIT*; accession no. XM_612028), and the cytokines *PTPRC* (accession no. XM_599431) and *CD8B* (accession no. NM_001105344) genes are listed in Supplementary Table 1. Thermal cycling conditions were, except for *CD8B* gene, as follows: an initial denaturation step at 94°C for 3 min followed by 35 cycles at 94°C for 30 s, annealing at 55°C for 45 s, extension at 72°C for 2 min, and a final extension step at 72°C for 2 min. The PCR conditions for the *CD8B* gene were: annealing at 58°C for 30 s, extension at 72°C for 30 s, and a final extension step of 10 min at 72°C . PCR products were analyzed on a 2% agarose gel electrophoresis stained with ethidium bromide.

mRNA amplification

The SMART mRNA Amplification kit (Clontech) was used to produce sense mRNA for cold- and hot-TRIZOL samples from both 70–90% interphase fraction and 90% bottom fraction. This kit uses the template switch mechanism to generate a double-stranded cDNA necessary for *in vitro* transcription without a PCR step generally associated with a conventional SMART technique. To produce sufficient amount of mRNAs, we used a two-round approach. The first round started with 150 ng RNA replicate while about 500 ng mRNA from the first round were used to repeat the mRNA amplification procedure to achieve sufficient amount to produce labeled probes for microarray analysis. To prepare mRNA from testis, total RNA was extracted using TRIZOL (Invitrogen), following the protocol recommended by the company, and 500 ng was used to produce amplified mRNA, as described above, by also performing a second round of amplification.

Construction of the microarrays

The construction of the cDNA chip is schematized in Fig. 2B. The microarray comprised PCR amplicons of three normalized cDNA libraries constructed using the suppressive subtractive hybridization (SSH) technique and a list of reference genes (Table 2). Two of those cDNA libraries were previously described (Lalancette *et al.* 2008b). In short, the first library contains genes highly expressed in high-fertile bull spermatozoa (F), status determined by their NRR (a fertility index). The resulting library was called NRR-SSH and was normalized with subfertile bull transcripts. The second library also reports highly expressed genes but in subfertile bull spermatozoa (L), generating the NRR_SSHL library, normalized with high-fertile bull transcripts using the SSH technique. These libraries were constructed with several doses of the respective bull, and these bulls were selected because they presented extreme NRR among 180 sons of one bull sire family. The third cDNA library was constructed using the SSH technique applied to genetically unrelated bulls (A), which are known to be highly fertile (F), resulting in the NRR-ASSH and was normalized with unrelated subfertile bull transcripts. These libraries represented 1776, 1248, and 171 EST for the respective NRR-SSH, NRR_SSHL, and NRR-ASSH libraries. A set of 96 EST prepared from a pool of transcripts (cDNA) from highly fertile and subfertile bulls, used in the cited manuscript (Lalancette *et al.* 2008b), were included on the chip. Controls were also selected (Supplementary Table 4, which can be viewed online at www.reproduction-online.org/supplemental/). Among them (Table 2), six *Arabidopsis thaliana* genes and bovine genes were selected were amplified using appropriate primers (Supplementary Table 6, which can be viewed online at www.reproduction-online.org/supplemental/). All of the PCR products were transferred to 384 well plates and purified using MultiScreen PCR Cleanup Plates (Millipore, Mississauga, Ontario, Canada). Prior to spotting, PCR products were resuspended in a 3×SSC and spotted in duplicates on GAPS II slides (Corning, New York City, NY, USA) using the BioRobotics MicroGrid Compact (Genomics Solution, Ann Arbor, MI, USA). Slides were then cross-linked with u.v. according to the

manufacturers' instructions. In order to check the spot integrity, microarray slides were incubated with TOTO3 dye (Molecular Probes, distributed by Invitrogen) in PBS for 10 min, washed with PBS, and dried prior to visualization.

Indirect labeling of RNA and hybridization design

The experimental design of the hybridization replicates is schematized in Fig. 2C. One microgram of sense-amplified mRNA of each RNA sample was reverse-transcribed by SuperScript II (Invitrogen) in the presence of amino allyl dUTP (Sigma–Aldrich). cDNA was then labeled with either Alexa Fluor 555 or 647 Reactive Dye (Invitrogen) to perform dye-swap hybridizations. Each labeling reaction included a spiking control mix containing 2 ng of each *A. thaliana* transcripts produced *in vitro* with T7 RNA polymerase (Epicentre Biotechnology, Madison, WI, USA). The PDF1.2 transcripts were produced from the u11912 and 37f10 gene IDs and the K22G18.2 transcript from the u17022 gene ID (Table 2) using appropriate primers (Supplementary Table 5). These spiking controls acted as positioning references for the grid applied on the image, which delimited the spots for data acquisition (Supplementary Figure 3, which can be viewed online at www.reproduction-online.org/supplemental/). In the microarray hybridization design, each fraction (70–90% interphase or the pellet 90% samples) treated with the respective protocol (hot- and cold-TRIZOL) was directly compared and included biological (i.e. experiment reproduced three times) and technical (hybridizations including dye-swap) replicates. Each comparison set totalized nine hybridizations (Fig. 2C). Before hybridization of the labeled probes, customarily printed arrays were washed twice in a 0.1% SDS solution at room temperature for 10 min. Hybridization was made overnight at 37 °C in DIG Easy Hyb (Roche) containing 300 ng/μl Salmon Sperm DNA (Ambion) and 300 ng/μl yeast tRNA. Slides were then washed twice for 15 min in 1×SSC, 0.1% SDS with gentle shaking at 50 °C, and once in 0.1×SSC, 0.2% SDS for 15 min at 50 °C with gentle shaking.

Image collection and data analysis

Slides were scanned with a ScanArray Express scanner (Perkin Elmer, Woodbridge, Ontario, Canada). Laser power and photomultiplier tube settings were adjusted to obtain the highest intensity without image saturation as recommended by the manufacturer. Scan Array Express Software v3.0 was used to quantify the spots for each array. Overall, a total of 18 data points were collected for each gene ID, with each of two points on the array across the nine slides used per comparison. The normalization modules provided by the open-source TM4 software suite of TIGR was used to normalize and filter ('low-intensity cutoff') the data. The 'surrounding' background-corrected signal intensity of each spot was normalized using the locally weighted linear regression (LOWESS; Yang *et al.* 2002, Morrison & Hoyle 2003). This can be linked with filters of the base 2 log-transformed using the MIDAS software module (TIGR; Quackenbush 2002). Each intensity value reflected the transcript abundance in the hybridized samples, which also include non-specific noise. In addition to the

background subtraction, filtering was applied using a threshold value determined by the s.d. of the intensity of each samples (cy3 and cy5 labeled) of negative controls spotted on the array, if any were detected. Next, the MIDAS software module allows to select the 'stringent' option of the 'Channel Tolerance' parameter (applied on cy3 and cy5 channels). Indeed, when stringent option was chosen, both channels passed only if the normalized intensity value was above the 'two-s.d. levels above background' and the 'two-s.d. levels above negative controls'. In other words, if one of the signal intensities of the spotted gene is too low, this gene information was removed from the analysis (called 'absent'). By opposite, if the 'generous option' was selected, a gene value was not discarded based on weak intensity from one or the other channel (hybridized spermatozoal sample). The stringent selection (both channels passed) was selected in this study, unless specified (inset of Fig. 3). The differences on gene expression levels in each time point were analyzed with *t*-test using the MeV application (TIGR). All *P* values were also corrected for the false discovery rate and are presented as an adjusted *P* value. Criteria on detection of differentially expressed gene were twofold or greater change in expression level with $P < 0.001$ and adjusted $P < 0.05$. Statistical significance was performed using the MULTTEST module of Statistical Analysis System, Release 9.1, 2002 (SAS Institute Inc., Cary, NC, USA). A selection of significant genes were further validated by real-time PCR order to further confirm the relative expression level (above section).

Sequence identity

Sequencing reactions were performed with the BigDye V3.1 chemistry and analyzed with a 3100-Avant Genetic Analyser (Applied Biosystems, Foster City, CA, USA). Sequences were aligned for similarity search by BLAST (local server) against the NCBI-downloaded bovine 'reference sequences RNA' (RefSeq_RNA), EST, NT, 'reference protein sequences' (RefSeq_prot), NR, Swissprot, and Bovine Genome version 3.1 (BTA 3.1) datasets (latest download: 2008/03/06). Sequence and BLASTout information were imported in a relational database (MYSQL). The sequencing files (for both complementary strands) of a gene ID were aligned using the multiple sequence alignment methods (CLUSTALW) of the program MegAlign of the DNASTAR suite (Lasergene version 7.2; DNASTAR, Madison, WI, USA). Similar bovine sequences, if any existed, were also uploaded directly from NCBI to assemble the similar sequence into a single contig and used to generate the consensus sequence.

Reverse transcription

All amplified RNA procured by the hot-TRIZOL protocol from both fractions were reverse transcribed in triplicate using the SuperScript II (Invitrogen) with oligo dT₁₂₋₁₈ (Invitrogen) according to the manufacturer's protocol. In order to normalize the difference in RT efficiencies across samples, each RT reactions was spiked *EGFP* mRNA. An equivalent of 1 pg/μg of spermatozoal RNA was added in pooled RNA and each RT reaction made with 1 μg included an equal amount of *EGFP* transcripts in each 20 μl reaction as described (Levesque-Sergerie *et al.* 2007).

qRT-PCR data analysis

Primers used for qRT-PCR were designed on the consensus sequence of each gene ID using PrimerExpress 3.0 software (Applied Biosystems; Supplementary Table 3). qRT-PCR results were analyzed following two approaches: the absolute and the relative quantification methods (Livak & Schmittgen 2001). The relative transcript levels are given by the arithmetic formula $2^{-\Delta\Delta C_t}$ (Livak & Schmittgen 2001). The absolute quantification determines the input copy number of the transcript of interest by relating the PCR signal to a standard curve. For qPCR assay, a gene-specific standard curve was generated for each gene ID using a PCR amplicon quantified by spectrophotometry (ABS 260/280), and converted into number of molecules, and serially diluted to produce the five reference points of the calibration curve. The formula takes into account the respective amplicon length generated during real-time PCR (Supplementary Table 3) as described (Lalancette *et al.* 2008b). Only the absolute transcript copies were normalized using the exogenous *EGFP* transcripts as described (Levesque-Sergerie *et al.* 2007), whereas the relative method readily includes a normalization using *PPIA*, an endogenous 'housekeeping' gene. All points of the standard curves and each query samples were carried out with triplicates and run in the ABI PRISM 7500 Sequence Detection System with software version 1.4 (Applied Biosystems). All runs included 3–6 negative controls (without target DNA). *PRM1* gene and all gene IDs were analyzed with the Fast SYBR Green chemistry while *EGFP* and *PRM2* genes were evaluated with the TaqMan Fast Universal master mix (Applied Biosystems). The thermal cycling conditions, for the SYBR Green amplifications, were as follows: 50 °C for 2 min and 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C for 15 s, and annealing at 60 °C for 60 s. Melting curves analysis (dissociation assay) was performed by plotting the fluorescence intensity in a graphic model. The single melting temperature peak confirmed the presence of a specific amplicon. For the *EGFP* and *PRM2* genes, amplified by TaqMan probe process, 20 s at 95 °C, was followed by 40 cycles of a two-step cycle: 3 s at 95 °C and an annealing at 60 °C for 20 s.

Statistical analysis for qRT-PCR

qPCR data were analyzed using the appropriate threshold setup as recommended by Applied Biosystems. The slope of the calibration curve was calculated from the plot of base 10 log of the initial target copy number versus corresponding C_t allowing the determination of gene copies number in each sample. In the absence of housekeeping genes, each set of qPCR data was normalized using the corresponding *EGFP* sample value as described previously (Lalancette *et al.* 2008b). Significance of the number of gene copies (real-time qRT-PCR measurements) among groups was performed using a Statistical Analysis System (Release 9.1, 2002; SAS Institute). Data recorded as the number of copies were normalized using a log transformation. Statistics were performed using the nonparametric Wilcoxon Mann–Whitney test. All tests were performed at a confidence level of 95%. All values are

presented as means of number of gene copies (transformation inverse of the log value) with the confidence interval (estimate of the transformed inverse corresponding s.e.m. values).

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 Grants from the Dairy Cattle Genetics Research and Development (DairyGen) Council and Agriculture and Agri-Food Canada.

Acknowledgements

The authors thank L'Alliance Boviteq Inc. for providing semen samples and Valérie Émond and Éric Bouchard for performing the *in situ* fluorescence microscopy. Epifluorescence was performed at the Cell Imaging Facility in the Genetics Service of the Department of Pediatrics, Faculty of Medicine and Health Sciences, at the Université de Sherbrooke, Sherbrooke, Canada. The facility is funded by the Centre de recherche clinique Étienne-Le Bel of the Centre Hospitalier de l'Université de Sherbrooke and by grants from Canadian Foundation for Innovation.

References

- Amann RP & Hammerstedt RH 1993 *In vitro* evaluation of sperm quality: an opinion. *Journal of Andrology* **14** 397–406.
- Bailey JL, Robertson L & Buhr MM 1994 Relation among *in vivo* fertility, computer-analysed motility and Ca⁺⁺ influx in bovine spermatozoa. *Canadian Journal of Animal Science* **74** 53–58.
- Dadoue JP, Pawlak A, Alfonsi MF & Siffroi JP 2005 Identification of transcripts by macroarrays, RT-PCR and *in situ* hybridization in human ejaculate spermatozoa. *Molecular Human Reproduction* **11** 133–140.
- Davis RO, Niswander PW & Katz DF 1992 New measures of sperm motion. I. Adaptive smoothing and harmonic analysis. *Journal of Andrology* **13** 139–152.
- Gilbert I, Bissonnette N, Boissonneault G, Vallee M & Robert C 2007 A molecular analysis of the population of mRNA in bovine spermatozoa. *Reproduction* **133** 1073–1086.
- Januskauskas A, Johannisson A & Rodriguez-Martinez H 2001 Assessment of sperm quality through fluorometry and sperm chromatin structure assay in relation to field fertility of frozen–thawed semen from Swedish AI bulls. *Theriogenology* **55** 947–961.
- Januskauskas A, Johannisson A & Rodriguez-Martinez H 2003 Subtle membrane changes in cryopreserved bull semen in relation with sperm viability, chromatin structure, and field fertility. *Theriogenology* **60** 743–758.
- Jones GM, Song B, Cram DS & Trounson AO 2007 Optimization of a microarray based approach for deriving representative gene expression profiles from human oocytes. *Molecular Reproduction and Development* **74** 8–17.
- Kjaestad H, Ropstad E & Berg KA 1993 Evaluation of spermatological parameters used to predict the fertility of frozen bull semen. *Acta Veterinaria Scandinavica* **34** 299–303.
- Kliesch S & Cooper TG 2008 Semen analysis: spermogram according to WHO criteria. *Der Urologe. Ausgabe A* **47** 1548–1554.
- Koops WJ, Grossman M & Den Daas JH 1995 A model for reproductive efficiency of dairy bulls. *Journal of Dairy Science* **78** 921–928.
- Lalancette C, Miller D, Li Y & Krawetz SA 2008a Paternal contributions: new functional insights for spermatozoal RNA. *Journal of Cellular Biochemistry* **104** 1570–1579.
- Lalancette C, Thibault C, Bachand I, Caron N & Bissonnette N 2008b Transcriptome analysis of bull semen with extreme nonreturn rate: use of suppression-subtractive hybridization to identify functional markers for fertility. *Biology of Reproduction* **78** 618–635.
- Larsson B & Rodriguez-Martinez H 2000 Can we use *in vitro* fertilization tests to predict semen fertility? *Animal Reproduction Science* **60–61** 327–336.
- Lévesque-Sergerie JP, Lalancette C & Bissonnette N 2006 Amplification of RNA from bull spermatozoa: towards a better characterization of spermatozoal RNA and its use in reproductive success evaluation. *8th World Congress on Genetics Applied to Livestock Production*, Belo Horizonte, MG, Brazil. Proceedings 11–19.
- Levesque-Sergerie JP, Duquette M, Thibault C, Delbecchi L & Bissonnette N 2007 Detection limits of several commercial reverse transcriptase enzymes: impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR. *BMC Molecular Biology* **8** 93.
- Livak KJ & Schmittgen TD 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} method. *Methods* **25** 402–408.
- Madrid-Bury N, Perez-Gutierrez JF, Perez-Garnelo S, Moreira P, Pintado Sanjuanbenito B, Gutierrez-Adan A & de la Fuente Martinez J 2005 Relationship between non-return rate and chromatin condensation of deep frozen bull spermatozoa. *Theriogenology* **64** 232–241.
- Mao XM, Ma WL, Feng CQ, Zou YG & Zheng WL 2004 An initial examination of the spermatozoal gene expression profile. *Di Yi Jun Yi Da Xue Xue Bao* **24** 1033–1036.
- Mehmood A, Anwar M & Naqvi SM 2008 Motility, acrosome integrity, membrane integrity and oocyte cleavage rate of sperm separated by swim-up or Percoll gradient method from frozen–thawed buffalo semen. *Animal Reproduction Science* **111** 141–148.
- Mengual L, Ballesca JL, Ascaso C & Oliva R 2003 Marked differences in protamine content and P1/P2 ratios in sperm cells from percoll fractions between patients and controls. *Journal of Andrology* **24** 438–447.
- Miller D & Ostermeier GC 2006a Spermatozoal RNA: why is it there and what does it do? *Gynécologie, Obstétrique & Fertilité* **34** 840–846.
- Miller D & Ostermeier GC 2006b Towards a better understanding of RNA carriage by ejaculate spermatozoa. *Human Reproduction Update* **12** 757–767.
- Morrell JM, Johannisson A, Dalin AM & Rodriguez-Martinez H 2008 Morphology and chromatin integrity of stallion spermatozoa prepared by density gradient and single layer centrifugation through silica colloids. *Reproduction in Domestic Animals* **44** 512–517.
- Morrison N & Hoyle D 2003 Normalization: concepts and methods for normalizing microarray data. In *A Practical Approach to Microarray Data Analysis*, pp 76–90. Eds DP Berrar, W Dubitzky & M Granzow. Boston, MA: Kluwer Academic Publishers.
- Mortimer D 1990 Objective analysis of sperm motility and kinematics. In *The Handbook of the Laboratory Diagnosis and Treatment of Infertility*, pp 97–133. Eds BA Keel & BW Webster. Boca Raton, FL: CRC Press.
- Ostermeier GC, Miller D, Huntriss JD, Diamond MP & Krawetz SA 2004 Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature* **429** 154.
- Ostermeier GC, Goodrich RJ, Diamond MP, Dix DJ & Krawetz SA 2005 Toward using stable spermatozoal RNAs for prognostic assessment of male factor fertility. *Fertility and Sterility* **83** 1687–1694.
- Parrish JJ, Krogenaes A & Susko Parrish JL 1995 Effect of bovine sperm separation by either swim-up or percoll method on success of *in vitro* fertilization and early embryonic development. *Theriogenology* **44** 859–869.
- Patel OV, Suchyta SP, Sipkovsky SS, Yao J, Ireland JJ, Coussens PM & Smith GW 2005 Validation and application of a high fidelity mRNA linear amplification procedure for profiling gene expression. *Veterinary Immunology and Immunopathology* **105** 331–342.
- Quackenbush J 2002 Microarray data normalization and transformation. *Nature Genetics* **32** 496–501.
- Ramirez JP, Carreras A & Mendoza C 1992 Sperm plasma membrane integrity in fertile and infertile men. *Andrologia* **24** 141–144.

- Rodriguez-Martinez H** 2000 Evaluation of frozen semen: traditional and new approaches. In *Topics in Bull Fertility*, pp 1–6. Ed PJ Chenoweth. Ithaca, NY: International Veterinary Information Service.
- Rodriguez-Martinez H** 2003 Laboratory semen assessment and prediction of fertility: still utopia? *Reproduction in Domestic Animals* **38** 312–318.
- Rodriguez-Martinez H** 2006 Can we increase the estimated value of semen assessment? *Reproduction in Domestic Animals* **41** 2–10.
- Rodriguez-Martinez H & Barth AD** 2007 *In vitro* evaluation of sperm quality related to *in vivo* function and fertility. *Society of Reproduction and Fertility* **64** 39–54.
- Rodriguez-Martinez H, Larsson B, Zhang BR & Söderquist L** 1997 *In vitro* assessment of viability and fertilizing capacity of bull spermatozoa. *Journal of Reproduction and Development* **43** 1–11.
- Saacke RG, Dalton JC, Nadir S, Bame JH & Nebel RL** 1998 Spermatozoal characteristics important to sperm transport, fertilization and early embryonic development. In *Gametes: Development and Function*, pp 320–335. Eds A Lauria, E Gandolfi, G Enne & L Gianaroli. Milano, Italy: Serono Symposia.
- Saacke RG, Dalton JC, Nadir S, Nebel RL & Bame JH** 2000 Relationship of seminal traits and insemination time to fertilization rate and embryo quality. *Animal Reproduction Science* **60–61** 663–677.
- Sanchez-Vazquez ML, Reyes R, Ramirez G, Merchant-Larios H, Rosado A & Delgado NM** 1998 DNA unpacking in guinea pig sperm chromatin by heparin and reduced glutathione. *Archives of Andrology* **40** 15–28.
- Smith KM, Gaultier A, Cousin H, Alfandari D, White JM & DeSimone DW** 2002 The cysteine-rich domain regulates ADAM protease function *in vivo*. *Journal of Cell Biology* **159** 893–902.
- Stalhammar EM, Janson L & Philipsson J** 1994 The impact of sperm motility on non-return rate in preselected dairy bulls. *Reproduction, Nutrition, Development* **34** 37–45.
- Tomlinson MJ, Moffatt O, Manicardi GC, Bizzaro D, Afnan M & Sakkas D** 2001 Interrelationships between seminal parameters and sperm nuclear DNA damage before and after density gradient centrifugation: implications for assisted conception. *Human Reproduction* **16** 2160–2165.
- Ward WS & Coffey DS** 1991 DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. *Biology of Reproduction* **44** 569–574.
- Ward MA & Ward WS** 2004 A model for the function of sperm DNA degradation. *Reproduction, Fertility, and Development* **16** 547–554.
- White JM** 2003 ADAMs: modulators of cell–cell and cell–matrix interactions. *Current Opinion in Cell Biology* **15** 598–606.
- Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J & Speed TP** 2002 Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research* **30** e15.
- Yaron Y, Kramer JA, Gyi K, Ebrahim SA, Evans MI, Johnson MP & Krawetz SA** 1998 Centromere sequences localize to the nuclear halo of human spermatozoa. *International Journal of Andrology* **21** 13–18.
- Zhang JS, Liu Q, Li YM, Hall SH, French FS & Zhang YL** 2006 Genome-wide profiling of segmental-regulated transcriptomes in human epididymis using oligo microarray. *Molecular and Cellular Endocrinology* **250** 169–177.
- Zhao Y, Li Q, Yao C, Wang Z, Zhou Y, Wang Y, Liu L, Wang L & Qiao Z** 2006 Characterization and quantification of mRNA transcripts in ejaculated spermatozoa of fertile men by serial analysis of gene expression. *Human Reproduction* **21** 1583–1590.
- Zhu B, Xu F & Baba Y** 2006 An evaluation of linear RNA amplification in cDNA microarray gene expression analysis. *Molecular Genetics and Metabolism* **87** 71–79.

Received 3 December 2008

First decision 6 January 2009

Revised manuscript received 9 April 2009

Accepted 6 May 2009