Animal (2013), **7:3**, pp 446–454 © The Animal Consortium 2012
doi:10.1017/S1751731112001620

Identification of some unknown transcripts from SSH cDNA library of buffalo follicular oocytes

S. K. Rajput[1], P. Kumar[1a], B. Roy[1], A. Verma[1], H. P. Pandey[2], D. Singh[1], S. De[1] and T. K. Datta[1†]

[1]*National Dairy Research Institute, Animal Biotechnology Centre, Karnal 132001, India;* [2]*Department of Biochemistry, Banaras Hindu University, 221005 Varanasi, India*

(Received 24 June 2011; Accepted 6 July 2012; First published online 10 August 2012)

A buffalo oocyte-specific subtracted cDNA library was constructed to identify exclusively or preferentially oocyte-expressed genes. The library represented an enriched population of transcripts obtained from oocytes of diverse ovarian follicular origin and at different stages of in vitro *maturation. A total of 1173 high-quality sequences of oocyte-specific genes were clustered into 645 unique sequences, out of which 65.76% were represented as singlets and 34.26% as contig expressed sequence tags (ESTs; clusters). Analysis of sequences revealed that 498 of these sequences were identified as a known sequence in mammalian species including buffalo, 103 as uncharacterized ESTs and 44 unknown sequences including 1 novel EST, so far not reported in any species. Gene ontology annotation classified these sequences into functional categories of cellular events and biological processes associated with oocyte competence. Expression status of the isolated unknown ESTs confirmed that many of these are expressed in oocytes exclusively and in others preferentially, some in excess of 80-fold greater in comparison with a variety of somatic tissues. The isolated novel EST was detected to be expressed exclusively in oocytes and testicular cells only. To our knowledge, this is the first report giving a detailed transcriptome account of oocyte-expressed genes in buffalo. This study will provide important information on the physiological control of oocyte development, as well as many questions yet to be addressed on the reproductive process of buffalo.*

Keywords: buffalo, oocyte, cDNA library, EST

Implications

A comprehensive description of all genes expressed in mammalian oocytes important for affecting 'oocyte competence' is still pending. The present work reports identification of some unknown transcripts identified in a buffalo oocyte-specific subtracted cDNA library, which were found to be expressed either exclusively or preferentially in oocytes up to several folds higher than other tissues examined. The high abundance pattern of these transcripts selectively in oocytes suggests that some of these could have a link with oocyte development and thereby have potential as markers of fertility. Data generated in this study will aid future investigation into poor reproductive rate and compromised success rate of assisted reproductive technologies in buffalo.

[a] Present address: Department of Hematology, BMC, Lund University, Lund 22184, Sweden.
[†] E-mail: tirthadatta@gmail.com

Introduction

Maturation of mammalian oocytes presents a protracted developmental process whereby an undifferentiated female germ cell prepares itself for undertaking the process of fertilization and guides the newly formed zygote through early developmental stages until the embryo's own genome becomes eligible to undertake its own transcription (Yamashita *et al.*, 2000; Zuccotti *et al.*, 2011). Transcriptionally, the period of oocyte maturation displays a typical cycle of mRNA synthesis and degradation starting with a very high turnover rate, which goes silent toward the end (Tomek *et al.*, 2002). A very fundamental yet practically relevant question that still remains unresolved is 'what makes an oocyte competent'? The ability to select the most competent oocytes out of a pool of oocytes will contribute to the success rate of many assisted reproduction technologies (ARTs) procedures awaiting large-scale implementation for genetic improvement of livestock, as well as for identifying new drug targets for infertility management (Armstrong, 2001).

Recent published literature suggests that discovery of novel transcripts in oocytes is still an open research agenda (Klatsky *et al*., 2010). Continued identification of new oocyte-specific genes and their functional annotation is providing additional insight into the molecular mechanism involved in folliculogenesis, oogenesis, fertilization and early embryogenesis (Bettegowda *et al*., 2007; Wells *et al*., 2008; Luo *et al*., 2010). Several techniques such as differential-display reverse transcription-PCR, microarray hybridization, serial analysis of gene expression and cDNA library construction have been used to reinforce the inventory of oocyte-expressed genes. Each of these techniques, however, suffers from the disadvantage of biased detection of more abundant transcripts and often misses rarely or lowly expressed transcripts. Suppressive subtractive hybridization (SSH) provides an attractive solution to identify rare transcripts, which may encode regulatory proteins and have crucial cellular role (Zeng and Schultz, 2003). The advantage of this method stems from its ability to normalize the mRNA population so that abundant mRNAs are reduced, whereas rare transcripts are enriched (Robert *et al*., 2000).

Buffalo (*Bubalus bubalis*) is an economically important dairy animal species in many Asian, African and Mediterranean countries because of its immense potential in agriculture, milk and meat production (Nanda and Nakao, 2003). Poor reproductive efficiency of buffalo, however, is a major hindrance toward implementing planned breeding improvement programs in this species. Late maturity, poor estrus expression, prolonged calving intervals, seasonal reproductive patterns and low conception rates directly affect the propagation of good germplasm in buffalo (Nandi *et al*., 2002; Drost, 2007). Although several ARTs have been demonstrated in this species, the overall success rates of these techniques remain low, thus constraining their widespread adoption (Singh *et al*., 2009). Delineating the exact cellular mechanism of oocyte function and identification of new genes in oocytes of this species would not only help to understand the buffalo oocyte biology better, but would also provide markers to study the oocyte competence phenomena in any species. The present study was undertaken to identify developmentally important genes expressed in buffalo oocyte during the maturation process using a complementary approach of SSH and customized cDNA macroarray analysis.

Material and methods

Buffalo oocytes and somatic tissues

Buffalo ovaries were collected from the abattoir and cumulus oocyte complexes (COCs) were aspirated from all visible follicles irrespective of the follicular diameter. COCs were classified as A, B and C grades based on cumulus cell and ooplasm characteristics as described by Kumar *et al*. (2011). About 200 COCs under each grade were cultured at 38.5°C under 5% CO_2 in TCM-199 HEPES-modified (Sigma, St. Louise, MO, USA) medium supplemented with 10% fetal bovine serum, 0.01% sodium pyruvate, 0.005% glutamine, 0.005% streptomycin, along with 5 μg/ml follicle-stimulating hormone(FSH-p) and 10 μg/ml leutinizing hormone. Batches of COCs were retrieved at 0, 8, 16 and 24 h of *in vitro* maturation culture and denuded completely by vortexing for 3 to 5 min. After washing, all oocytes were pooled (2400 oocyte) and stored at −80°C in 500 μl Trizol (Invitrogen, Carlsbad, CA, USA). Isolated cumulus cells were collected separately after washing three times with 1× phosphate-buffered saline and preserved in Trizol at −80°C. Buffalo somatic tissues (heart, kidney, intestine, liver, lung, muscle and spleen) were also collected at the abattoir in RNA later (Ambion, Drive Foster, CA, USA) and stored at −80°C till RNA isolation.

RNA extraction

Total RNA was isolated from somatic tissues (lung, kidney, liver, heart, intestine, spleen and cumulus cells) and the oocyte pool using Trizol protocol as per the manufacturer's instructions. The concentration and purity of the isolated RNA was determined using a NanoDrop ND-1000 spectrophotometer and integrity of the total RNA was assessed using a Bio-Analyzer 2100 (Agilent, New Delhi, India). RNA samples with RNA integrity number values of 7 and above were processed further for library preparation. Equal amounts of total RNA from each somatic tissue were pooled and adjusted to a final concentration of 0.5 mg/ml (used as 'driver' in the SSH protocol). RNA isolated from oocytes was precipitated using 1 μg glycogen to obtain a final concentration of 0.5 mg/ml and used as a 'tester' RNA. Total RNA (2 μg) isolated from oocytes and somatic tissues were incubated with DNase-I (Fermentas, Burlington, ON, Canada) to remove contaminating genomic DNA and purified by organic extraction (Chomczynski and Sacchi, 1987). After DNase-I treatment, the genomic DNA contamination was negated by buffalo genomic DNA-specific primer amplification with the 10 ng of RNA from each extracted samples.

Construction of cDNA library

Equal quantity (1 μg) of the total RNA isolated from the oocytes (tester RNA) and somatic tissues (driver RNA) were processed simultaneously for cDNA synthesis using the Super SMART PCR cDNA synthesis kit (Clontech, Mountain View, CA, USA) following manufacturer's instructions. Synthesized cDNA was amplified by long-distance PCR (LD-PCR) to enrich full-length cDNA population. It was followed by RsaI digestion of tester and driver double-stranded (ds)-cDNA to prepare blunt-end cDNA for SSH experiments. In the next step, SSH was performed to subtract the oocyte cDNA (tester) with excess of cDNA from somatic tissues (driver) using the PCR-Select cDNA subtraction Kit (Clontech) as per the manufacturer's instructions. After SSH, the resultant primary and nested PCR products were amplified for 27 cycles. Diluted PCR products (10×) were further used for secondary amplification for 10 cycles to obtain the subtracted cDNA, representative of oocyte transcripts only. The subtraction efficiency was analyzed by PCR amplification of the subtracted and non-subtracted cDNAs using the constitutively expressed gene *GAPDH* using buffalo-specific primers (forward 5′CCAGAATATCACCCCTGCTT3′ and reverse

5′CCCAGCATCGAAGGTAGAAG3′). The PCR amplification was performed for 18, 23, 28 and 33 number of cycles with thermal conditions of 95°C for 2 min followed by '*n*' number of cycles at 95°C for 15 s, 56°C for 30 s, 72°C for 30 s and final incubation at 72°C for 5 min. Five microliters of PCR product was collected at 18, 23, 28 and 33 cycles and analyzed on 2% agarose gel (Figure 1). The secondary PCR products obtained after the SSH experiment was purified using Qiagen column, cloned into pGEMT easy vector (Promega, Madison, WI, USA) and transformed into XL-Blue *Escherichia coli* competent cells (Invitrogen) to establish an oocyte-specific subtracted cDNA library. Transformed colonies were randomly picked up and screened for the recombinant clones by colony PCR using a vector-specific M13 primer. PCR was performed in 50 μl at 95°C for 3 min followed by 30 cycles of 95°C for 30 s, 50°C for 15 s and 72°C for 1 min 30 s, and finally incubated at 72°C for 10 min. Cloning efficiency was analyzed by running 10 μl of PCR product on 1.5% agarose gel electrophoresis. Clones with a minimum inset size of 200 bp were processed for sequencing.

Sequence analysis of the cloned expressed sequence tags (ESTs)

Putative identity of the sequences was determined using DNA (megablast, blast-n, t-blastx) and protein (blastx) sequence search tools against several databases (Nucleotide, ESTs, RefSeq RNA, PDB, Swiss-Prot, Genome) for different mammalian species (human, bovine, ovine, murine) available at National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) using *e*-value 10^{-2} to 10^{-5}. On the basis of the results, sequences were classified as known, uncharacterized, unknown and novel sequences. Known sequences were defined as cases where a high degree of the homology was found with already identified genes available in mammalian databases. Sequences that did not match with known genes but possessed homology with EST databases were grouped as uncharacterized, and those that did not show any significant similarity with nucleotide and the EST database but showed some similarity with unannotated genomic DNA in bovine (considered homologous species to buffalo) were classified as unknown. Novel ESTs were designated as the cases where no match could be established using different genomic and EST databases available. Sequences included in known, uncharacterized and unknown categories were further analyzed against *Bos taurus* genome database available at NCBI to search the 5′untranslated region (UTR), 3′UTR, poly-A tail and other structural elements. Potential open reading frames (ORFs) in the identified sequences were searched for all the unknown and novel sequences using Geneious Pro (Drummond *et al.*, 2007) and sequences having more than 200 bp ORF length were considered as putative coding RNA. Analyzed sequences were submitted to the EST database at NCBI GenBank.

Nonradioactive labeling of cDNA

Both oocyte- and tissue-specific nonradioactive Digoxigenin (DIG)-labeled probes were synthesized using the super smart cDNA synthesis kit (Clontech) with minor modifications. In brief, 1 μg each of oocyte and somatic tissue RNA were used as templates to prepare full-length cDNA. The purified cDNAs were used as templates to generate the DIG-labeled ds-cDNA probe from oocytes and somatic tissues using LD-PCR amplification. Total 100 μl PCR reaction mixture was prepared containing 120 ng of ss-cDNA, 10 μl of 10× PCR buffer (Clontech), 10 μl of 10× PCR DIG probe synthesis mix (Roche, Mississauga, ON, Canada), 10 μl of 10× deoxyribonucleotide triphosphate (dNTP) stock solution (Roche), 2 μl of 12 μM 5′ PCR primer-IIA (Clontech) and 2 μl of 50× advantage cDNA pol mix (Clontech). Identical PCR reactions were carried out with cDNA from both oocytes and somatic tissues. The PCR mixture was subjected to a total of 30 cycles using thermal cycles of 95°C for 1 min, followed by 10 cycles at 95°C for 20 s, 65°C for 30 s, 68°C for 3 min and remaining 11 to 30 cycles were performed at 95°C for 20 s, 65°C for 30 s, 68°C for 3 min plus additional 40 s for each successive cycle and final incubation at 68°C for 7 min. Labeled ds-cDNA were purified using size exclusion column chromatography (Clontech) to remove the unincorporated DIG-labeled dNTPs.

cDNA macroarray hybridization

A custom cDNA macroarray was prepared for all the identified and nonredundant clones. For this, the inserts were amplified using M13 primer (as described above) and purified using QIAquick PCR purification kit (Qiagen, Valencia, CA, USA). Each purified PCR products (10 ng) were blotted in duplicate onto Hybond N+ membrane (GE healthcare, Gurgaon, Haryana, India) using vacuum slot blot as per the manufacturer's instructions. The blotted DNA was cross-linked to the membrane using UV linking under 70 000 μJ/cm^2 followed by baking at 80°C for 2 h. Housekeeping genes, *GAPDH* and *ACTB*, were also blotted on the same membrane as positive controls. Custom cDNA macroarrays were hybridized with DIG-labeled cDNA probes generated from either oocytes or somatic tissue RNA. Prehybridization (to reduce the background) and hybridization of the labeled probe were performed using the standardized protocol (Sambrook *et al.*, 1989) and were immunodetected using the DIG labeling detection system (Roche) according to the manufacturer's instructions. Images obtained after hybridization of the probe generated from oocyte and somatic tissues were recorded on X-ray film and data were compared between oocytes and somatic tissue by quantification of the signal using Image-J software.

Reverse transcription (RT)-PCR

RT-PCR was performed to validate the results obtained from macroarray hybridization-based differential screening. In total, 24 clones with more than 10-fold overexpression in oocytes as compared with somatic tissues were selected to design sequence-specific primers (Table 1). For cDNA synthesis, 500 ng of total RNA each from oocytes and pooled somatic tissues were used to synthesize single-strand cDNA using SMART scribe reverse transcriptase (Clontech). After purification, 10 ng of oocyte cDNA and 100 ng of somatic tissues cDNA were used as templates in RT-PCR analysis of

Table 1 *Primers used for RT-PCR analysis of 24 genes selected for confirming their oocyte-specific expression*

GenBank Accession no.	Forward primer sequences (5′ to 3′)	Reverse primer sequences (5′ to 3′)
HO0000106[1a]	TGCCTTAAACTTGCCCAAAG	CATGGAGGAGTGGAAATTGG
GW813181[2a]	CATCTCTCCATCTCCCCTTG	GAACTCTGACTGCTGACAATG
HO0000099[2a]	GCAAGGTCATATAACTAATGAGTG	GACATATCCCAGCATAAAGTTTAG
HO0000102[2a]	GAACACAAGCAGAAAGTGAAC	GCCTTTGACCTTGGACATG
HO0004899[2a]	CACGATTCCAGAATCCAGATTC	AACAGAAGTATTCAACTGTGACC
HO0000101[2a]	AAGAACCCCACGTTCCATG	TGAGCGTAAATGCAACTGTG
HO0004900[2b]	GCAAACCCCAGAGCTAATTC	ATCTACCTTCTTAGAGAGGTGTC
GW813175[2b]	TTGTCCACTCTCTTTTATTCACC	TCTTTAATGTTCACCACTGTATCC
GW813176[2b]	TTCTGGTCAAGGCTCTCAAG	CGTCAGAGGGATGCATGG
GW813177[2b]	GTAAACGACCGCCCACAG	CTCCTCTGTAAACCCTTCCTAC
HO0004902[2b]	AGAGGCAAAGAACTGAAACAC	AAGTCTTCTTCCTACTCCTGTC
HO0000097[2b]	CCTGAAAATCGACACCTGTAG	GCTACTTGTGCTCACTCTATTC
GW840938[3a]	CGCTACTACCGATTGGATGG	ACCTACGGAAACCTTGTTACG
GW840948[3a]	AGAGGCTGCTGTTTCCTG	GTTCGGTTGATGTAGTTCTCC
GW840947[3a]	CCAAAACCCGAGGGTGAG	GTGAGGCTTCTGTCGTTTG
HO0000354[3a]	TCTTCGTCTTGCGCTTCC	AGCCTCTTCATCCTCTTTCC
HO0000355[3a]	GACATAAGAAGGGCCACATG	GATAGGCACCTTGTCTTGTTC
HO0000361[3a]	GGATGGAGCAGTGGTCAAG	ATAGTAGACACAGTCACAGAGC
GW840919[3a]	GGCTTCCTGTTCCTTCTTTACTC	ACCTGACCTCCTCCATCGG
HO0000328[3a]	CTCTCTGATGGAGCTGAAGG	TGAACCAGATGACCTGAGTG
GW840956[3a]	TCCTGGGCTTTACATCAGC	TGAACTGGACTTCAGAGGAG
GW840936[3a]	TCATCAATAGCCCACCACTG	GATCAGACCCAGAAAGTTCTTG
HO0000349[3a]	TCAGAAGGACACAGGCACACAC	CTGACAACTTGAGAGCGTGAAGC
GW840946[3a]	CTAAGTGTGATCCCCAAAACC	ATAGCCGTGGTCTTCTTCTC

RT-PCR = reverse transcription-PCR; EST = expressed sequence tags; ORF = open reading frame.
[a]ESTs having more than 200 bp ORF; [b]ESTs without ORF.
[1]Novel EST.
[2]Unknown EST.
[3]Uncharacterized EST.

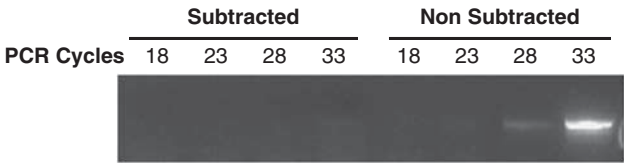

Figure 1 Amplification results of *GAPDH* from subtracted and non-subtracted cDNA at respective PCR cycle numbers.

selected transcripts. The expression status of the identified novel gene (BOE283N) was studied independently in different somatic (cumulus, heart, lung, kidney, spleen, intestine) and reproductive tissues (oocyte, testis, COC) using equal quantities of cDNA (100 ng). PCR reaction was performed on each cDNA sample under the following conditions: 95°C for 2 min; 30 cycles at 95°C for 30 s, 58°C for 15 s and 72°C for 20 s; and then 72°C for 5 min.

Results

Validation of the SSH library

An average of 4.2 μg of total RNA was recovered from a pool of 2400 oocytes (1.75 ng/oocyte). Total RNA (1 μg) isolated from oocyte and pool of somatic tissues was subjected to ds-cDNA synthesis and amplification following LD-PCR. SSH was performed with good-quality amplified cDNA and subtraction efficiency was evaluated by detecting the housekeeping gene (*GAPDH*) in both subtracted and non-subtracted cDNA pools. The PCR product of *GAPDH* appeared to be detectable after amplification of 33 PCR cycles with subtracted cDNA compared with 28 cycles without subtraction. This indicated a reduction of *GAPDH* transcript by more than 30-fold after subtraction (Figure 1). After cloning and transformation, more than 2000 clones from the subtracted library were amplified by PCR using M13 primers against the vectors. This procedure revealed that 1875 clones from the bovine library were positive insert-containing clones. The insert size of the clones ranged from 0.1 to 2.0 kb, and the average fragment insert size was ~500 bp, which is in agreement with that statistically predicted by *Rsa*I digestion.

Sequence analysis of SSH cDNA clones

Table 2 presents a summary of sequence results from our cDNA library. A total of 1173 high-quality sequences were obtained, which were found to represent a maximum of 645 unique sequences based on clustering analysis results. Approximately 85% of the genes were present once or twice in the library, confirming effective normalization of the subtraction procedure used (Figure 2). The level of redundancy in the library was found to be ~45%. Further, putative identity of the 645 unique sequences were determined using BLAST search against NCBI databases and classified them in known,

Table 2 *Summary of the sequences generated from oocyte-specific cDNA library*

Sequence categories	Number of clones
Total clones selected from the library	1588
Number of high-quality sequences obtained	1173
Number of unique and non-redundant sequences	645
Sequences that matched with buffalo GO annotation	2
Sequences matched with GO annotation in homologous species but uncharacterized in buffalo	496
Sequences matched with EST database but function not known: uncharacterized	103
Sequences matched with the unannotated genomic DNA of the *Bos taurus:* unknown	43
Sequence could not find a match with any database: novel	1

GO = gene ontology; EST = expressed sequence tags.

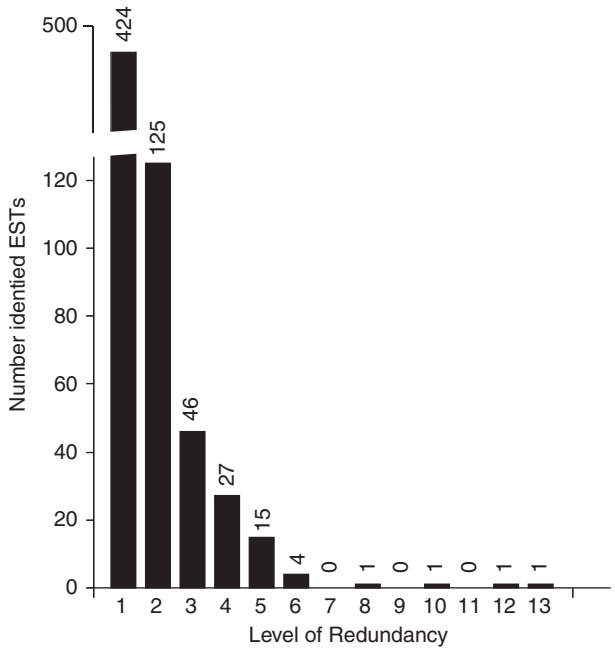

Figure 2 Frequency distribution of assembled expressed sequence tags (EST) sequences obtained in cDNA library representing the normalization effect of suppressive subtractive hybridization procedure and level of redundancy in cDNA library. The *X*-axis represents the number of times the same sequence is represented in the library.

uncharacterized, unknown and novel categories. BLAST analysis revealed that 77.21% (498) of the unique sequences with significant similarity with known genes, 15.97% (103) matched with functionally unannotated ESTs of bovine, ovine and murine origin, and hence were classified as uncharacterized sequences. About 6.5% (43) of sequences did not match with any nucleotide or EST databases but showed similarities with unannotated sequences from the bovine genome and thus were classified as unknown sequences. One sequence was found to be novel, as it did not match with any reported sequence available in GenBank.

On the basis of the information available from GenBank database, we observed that for the cDNA corresponding to known genes (498) in our subtracted library 69% of the cDNA contained UTRs at 5′ or 3′ end. A majority of these clones had a 3′ UTR region (51%) and a smaller number had a 5′ UTR region (18%). Confirmation of the absence of housekeeping genes (*GAPDH* and *ACTB*) and detection of several oocyte- and embryo-specific genes like zona pellucida genes (*zp1-4*); *gdf3* and *9*, *bmp15, jy1, btg4, npm2, nalp14, osp1* (Table 3) in subtracted oocyte cDNA pool authenticated the quality of our SSH library.

Functional annotation of ESTs

Functional annotation of identified known genes was carried out using the gene ontology (GO) program as per panther classification system (http://www.pantherdb.org). Identified genes were classified into three categories of cellular function, biological process and cellular component (Figure 3). It was observed that in cellular function category, the most abundant sequences were associated with binding (35.8%) and catalytic activity (29.3%). For biological process, the most frequent process identified was the cellular and metabolic process (46%) followed by cell communication (9.0%) and cell cycle (8.1%). The most frequent cellular components identified were intracellular (56.2%) in which a majority of the sequences were associated with cytoskeleton. Eighteen unknown and the novel ESTs in the present library were confirmed as having a potential ORFs with a minimum length of 200 bp (Drummond *et al.*, 2007; Table 4).

Screening of ESTs for oocyte specificity by cDNA macroarray

The results of differential screening of the ESTs in oocyte *v.* somatic tissues are presented in Figure 4. Most of the 645 spotted PCR products of the selected clones were observed to yield hybridization signals with the probe made from the pool of oocytes RNA (Figure 4a). Signals obtained from equivalent spots hybridized with probe made from other somatic tissue cocktail are presented in Figure 4b. In either case, the signal intensity of each spot was normalized against that of *GAPDH* and was compared between oocytes and somatic tissues. Differential signal intensity of a majority of spots hybridized with oocyte *v.* tissue probe suggested that about 21% of these are exclusively oocyte-expressed genes, whereas the remaining genes were expressed preferentially in oocytes. Only a few spots (16) were found positive for both the tissue and oocyte blots and hence could be designated as false positives. The group interpreted as exclusively oocyte-expressed genes represented already known oocyte-specific genes such as *GDF3, GDF9, BMP15, JY1, NPM2, NALP14, OSP1, zp2, zp3* and *zp4*. A summary of the preferentially expressed known ESTs are provided in Table 5.

RT-PCR validation of oocyte-specific ESTs

Subsequent to differential screening of identified ESTs by cDNA macroarray hybridization, the next level of validation was undertaken to confirm their oocyte-specific expression using a list of 24 representative genes, which revealed at

Table 3 *Identified oocyte-specific genes in our library that are involved in different stages of oocyte and preimplantation embryo development in bovine, human, ovine and murine*

Oogenesis	Folliculogenesis	Fertilization	Zygotic gene activation	Embryonic development
zp2, zp3, zp4, osp1,cdk, cyclinb1, polR2C, CPEB, DDX39 PIAS1, KPNA2, RBBP7	GDF9, GDF3, BMP15, PIAS1	Map2k1, ADAMTS12 zp2, zp3, zp4, PIAS1	oct4, SW1/SNF, PIAS1, DNMT3A, HMGN3	Dnmt1o, npm2, nalp7, dppa3, btg4, CUL1, jy1, dpy30, PAPOLG, CSTF3,

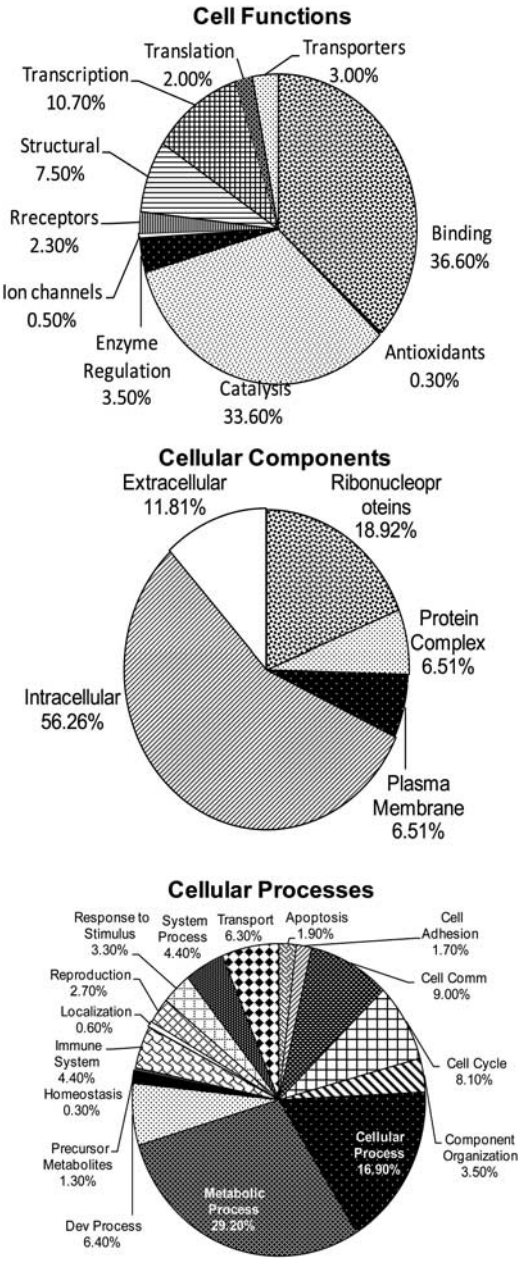

Figure 3 Functional gene ontology of identified annotated sequences represented in the library categorized as 'Cellular functions' 'Cellular components' and 'Biological process'.

Table 4 *List of the EST (18 unknown and 1 novel) with >200 bp ORF predicted*

GenBank Accession no.	ORF size (bp)	Category
GW840946	255	Unknown
GW813166	297	Unknown
GW813167	375	Unknown
GW813169	387	Unknown
GW840949	306	Unknown
GW840933	324	Unknown
GW813178	336	Unknown
GW813181	372	Unknown
GW813182	225	Unknown
GW813183	339	Unknown
GW813184	318	Unknown
HO000099	204	Unknown
HO000111	213	Unknown
HO000102	354	Unknown
HO000107	534	Unknown
HO000112	264	Unknown
HO004898	255	Unknown
HS514055	315	Unknown
HO000106	324	Novel

EST = expressed sequence tags; ORF = open reading frame.

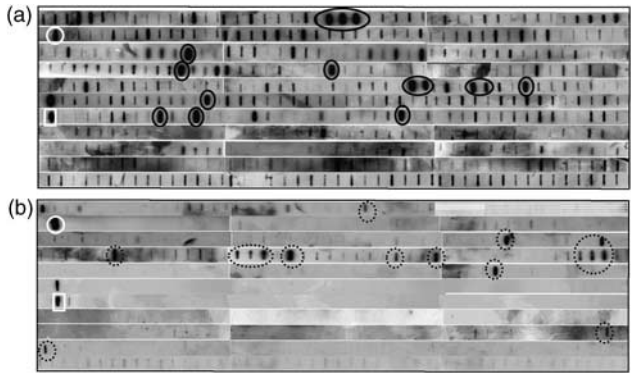

Figure 4 Hybridization signals from selected clones using probes made from the pool of oocytes (a) and the pool of somatic tissues (b). The housekeeping gene *GAPDH* is in the white circle and *ACTB* in the white square. False positives are represented by black dotted circles and the genes highly abundant in oocyte are in black circles.

least 10-fold higher expression in oocytes than other tissues. These 24 genes included 12 uncharacterized, 11 unknown and the isolated novel gene. The comparative expression profile of these ESTs in oocytes *v.* somatic tissues confirmed high specificity of these ESTs in oocytes (Figure 5). As a strategy to completely negate the possibility of false-positive interpretation about their oocyte-specific expression, we

Table 5 *Summary of the known ESTs showing more than 5-fold preferential expression in oocyte* v. *somatic tissue*

GenBank Accession no.	Frequency in the library	Blast identity (*e*-value: 10^{-2} to 10^{-5})
GW863794	6	Bb growth differentiation factor 9 B(Gdf9B)
GW841063	1	Bt growth differentiation factor 3 (Gdf3)
GW841063	1	Bt bone morphogenetic protein 15 (BMP15)
GW863647	12	Bt Nucleoplasmin 2(Npm2)
GW841014	8	Hs B-cell translocation gene 4(BTG4)
GW841060	5	Bt CyclinB1
GW863667	5	Bt THAP domain containing 9 (THAP9)
GW863820	10	Bt zona pellucida glycoprotein 4 (ZP4)
GW863673	5	Bt zona pellucida glycoprotein 2 (ZP2)
GW841050	2	Bt zona pellucida glycoprotein 3 (ZP3)
GW863710	5	Bt oocyte-secreted protein 1 (OOSP1)
HO004863	1	Bt retinoblastoma binding protein 7 (RBBP7),
GW863639	15	Bt BRISC complex subunit Abro1 (FAM175B)
HO004873	1	Bt transforming growth factor, beta receptor 1 (TGFBR1)
GW996805	2	Bt developmental pluripotency associated 3 (DPPA3)
GW840929	2	Hs developmental pluripotency associated 4 (DPPA4)
GW996802	1	Bt LIM homeodomain 3 protein b isoform (Lhx3)
GW863811	1	Bt cleavage stimulation factor, 3′ pre-RNA, subunit 3, 77 kDa (CSTF3)
GW863657	6	Bt yippee-like 5 (Drosophila; YPEL5)
GW863652	5	Bt transmembrane protein 87A (TMEM87A)
GW841074	5	Bt aurora kinase A (AURKA)
GW863720	2	Bt JY1
GW996762	2	Bt LSM7 homolog, U6 small nuclear RNA associated (*Saccharomyces cerevisiae*; LSM7)
GW863804	2	Bb myogenic factor 5 (Myf-5)
GW996748	3	Bt cadherin 1, type 1, E-cadherin
HO004883	2	Bt Culin1
GW841064	1	Bt F-box protein 8 (FBXO8)
GW841071	2	Bt M-phase phosphoprotein 6
GW863829	1	Bt nucleoporin 54 kDa (NUP54)
GW840915	6	Bt similar to PTTG1 protein (PTTG1)
GW841073	1	Bt similar to splicing factor, arginine/serine-rich 12, transcript variant 2 (SFRS12)
GW863779	1	Bt mitogen-activated protein kinase kinase 1 (MAP2K1)
GW996745	2	Bt F-box protein 4 (FBXO4)
GW841069	3	Bt Netrin receptor UNC5C precursor (Protein unc-5 homolog C) (Unc-5 homolog 3)
GW863669	1	Bt eukaryotic translation elongation factor 1 alpha 1, transcript variant 1
GW863671	1	Bt heterogeneous nuclear ribonucleoprotein D
GW863672	1	Bt mediator complex subunit 18 (MED18),
GW863702	5	Bt fragile histidine triad isoform 4 (FHIT4)
GW863674	1	Bt maternal embryonic leucine zipper kinase (MELK),
GW863750	2	Bt tripartite motif-containing 49 (TRIM49),
GW996725	2	Bt breast cancer anti-estrogen resistance 4 transcript variant 2 (BCAR4) mRNA
GW863843	3	*Bos taurus* Y-box binding protein 1 (YBX1)

EST = expressed sequence tags; Bb = *Bubalus bubalis*; Bt = *Bos taurus*; Hs = *Homo sapiens*.

used 10-fold excess cDNA from somatic tissues than that from oocytes while setting up the RT-PCR assay. Results revealed that at least eight ESTs (five uncharacterized and three unknown) were expressed exclusively in oocytes and the remaining ones were barely detectable in a pool of somatic tissues. Interestingly, the novel gene (BOE283N) was found to be expressed only in oocytes and testis and no other somatic tissues tested (Figure 6).

Discussion

Over the last decade, vast insight has been gained about molecular events involved in oocyte function in several mammalian species, by identifying genes preferentially expressed in oocytes (Bettegowda *et al*., 2007; Gaska *et al*., 2007; Baillet *et al*., 2008; Parrish *et al*., 2011). In spite of their economic importance in dairy production, however,

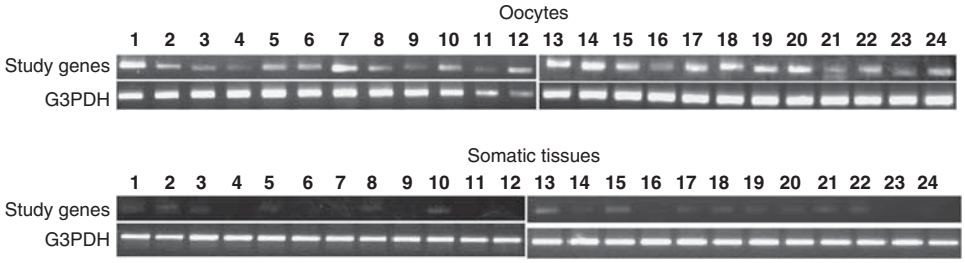

Figure 5 Differential expression status of 24 unknown genes in oocytes *v.* somatic tissues. Lane 1: identified novel gene, Lanes 2 to 12: uncharacterized genes and Lanes 13 to 24: unknown genes.

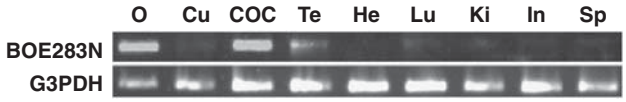

Figure 6 Expression pattern of the novel expressed sequence tags (BOE283N) in different somatic and reproductive tissues. O = oocytes; Cu = cumulus; COC = cumulus oocyte complex; Te = testis; He = heart; Lu = lung; Ki = kidney; In = intestine; and Sp = spleen.

buffalo remains a relatively unexplored species particularly in terms of explaining their reproduction disadvantages. The genomic and EST database of this species also remains grossly incomplete. The present study targeted isolation and identification of oocyte-expressed transcripts in buffalo, which might help in understanding the typical molecular mechanism of oocyte maturation and to discover new oocyte-specific genes, which will enrich knowledge about the reasons for reproduction constraints in this species.

Different strategies have been used in the past to isolate differentially expressed tissue-specific genes under given biological situations. In recent times, several high-throughput sequencing methods have evolved, which allow reading millions of nucleic acid sequences in short time, thereby expanding the scope of gene discovery beyond what was possible with standard dye terminator methods (Schuster 2008; Tucker *et al*., 2009). However, SSH as a method was chosen in the current study as it does not require *a priori* sequence knowledge and can be used with a low quantity of RNA obtained from less abundant tissues such as oocytes and embryos. To increase the possibility of oocyte-specific gene discovery, cDNA library was generated by using tester mRNA isolated from a carefully selected pool of oocytes representing all morphological types and physiological status, and it was subtracted by the driver mRNA prepared from a pool of somatic tissues representing diverse cell types. Testis was purposely not included among the driver population so that ESTs specific to germ cells of both sexes are represented in our library. In the analyzed sequence repository, we observed that the chance of obtaining a new independent clone was more than 50% even after sequencing of 800 clones. This confirmed the efficiency of cDNA subtraction strategy used in the present study, which was better than reported earlier (Hennebold *et al*., 2000). The custom macroarray approach used in the present study represents a quick and simple method of high-throughput screening of tissue-specific cDNA libraries. It was observed that the intensity of the hybridization signals of individual ESTs was positively correlated with its level of redundancy reflecting the fidelity of subtracted cDNA library construction (Sahu and Shaw, 2009). Large-scale sequence analysis data lead us to suggest that the subtraction method followed in generating the SSH library was potent enough to exclude common genes in oocytes and other somatic tissues. False-positive cases in the library were only about 2% (16 out of 645).

In this study, we identified the 496 known, 103 uncharacterized, 43 unknown and 1 novel EST in our library, which are not previously reported to be expressed in buffalo oocytes and submitted these to the EST database (Table 2). On the basis of the GO annotation (of the identified known ESTs), it was determined that ~94% of transcripts detected in our library were involved in crucial molecular events associated with oocyte and preimplantation embryonic development in different mammalian species such as bovine (Dalbie's-Tran and Mermillod 2003; Fair *et al*., 2007; Ghanem *et al*., 2007; Salilew-Wondim *et al*., 2007), ovine (Baillet *et al*., 2008), human (Assou *et al*., 2006) and mice (Su *et al*., 2007). Identified genes were also classified as oocyte-specific genes involved in folliculogenesis, oogenesis, fertilization, zygotic genome activation and preimplantation embryo development (Table 3). Sequence analysis of identified unknown and novel ESTs revealed that at least 18 of these were matching with the intergenic part of the bovine chromosomes and presented a poly-A tail with potential ORF, which qualified them as the protein coding transcripts. The remaining unknown ESTs with poly-A tail of more than 200 bp size but having no apparent coding clue can be speculated as potential oocyte-expressed long noncoding RNA, which have recently been suggested as the key regulators of embryogenesis (Pauli *et al*., 2011). Detailed sequence analysis of the identified novel EST revealed 324 bp continuous ORF encoding a 108-amino acid polypeptide that was found to have similarity with mammalian transmembrane proteins and thus could qualify to be a potentially important oocyte functional gene. This will, however, require further functional validation. To our knowledge, this is the first report on molecular description of buffalo oocytes at the mRNA level. This would help to understand the apparently unresolved problems with buffalo reproduction shortcomings and compromised ART results in this species (Singh *et al*., 2009).

Conclusion

The present study reports isolation of 44 new oocyte transcripts identified from an SSH cDNA library prepared from buffalo oocytes taken from diverse follicular milieu. Functional identity of the associated genes/ESTs is yet to be established in any mammalian species. It was interesting to observe that some of these genes are expressed exclusively in oocytes and some were detected preferentially in oocytes to the extent of more than 80-fold over a variety of somatic tissues tested. The current cDNA library was validated with more than one approach to confirm its high oocyte specificity. However, further characterization of identified transcripts by full-length cDNA amplification and functional annotation is required. The next order of validating their functional relevance would be to gather experimental evidence for loss of function using knockout or gene-silencing models. There is a possibility that some of these identified genes might emerge as important candidate genes involved in oogenesis and early follicle development, which could subsequently lead to the discovery of new potential oocyte competence markers. To our knowledge, this is the first report providing detailed information on the transcriptional control of buffalo oocytes, a species with acknowledged reproductive problems.

Acknowledgment

This work was supported under World bank NAIP project C4056 to the corresponding author.

Supplementary material

For supplementary material referred to in this article, please visit http://dx.doi.org/10.1017/S1751731112001620

References

Armstrong DT 2001. Effects of maternal age on oocyte developmental competence. Theriogenology 55, 1303–1322.

Assou S, Anahory T, Pantesco V, Le Carrour T, Pellestor F, Klein B, Reyftmann L, Dechaud H, De Vos J and Hamamah S 2006. The human cumulus-oocyte complex gene-expression profile. Human Reproduction 21, 1705–1719.

Baillet A, Pépin BM, Cabau C, Poumerol E, Pailhoux E and Cotinot C 2008. Identification of transcripts involved in meiosis and follicle formation during ovine ovary development. BMC Genomics 9, 436–438.

Bettegowda A, Yao J, Sen A, Li Q, Lee KB, Kobayashi Y, Patel OV, Coussens PM, Ireland JJ and Smith GW 2007. JY-1, an oocyte-specific gene, regulates granulosa cell function and early embryonic development in cattle. Proceedings of the National Academy of Sciences of the United States of America 104, 17602–17607.

Chomczynski P and Sacchi N 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. Analytical Biochemistry 162, 156–159.

Dalbiès-Tran R and Mermillod P 2003. Use of heterologous complementary DNA array screening to analyze bovine oocyte transcriptome and its evolution during in vitro maturation. Biology of Reproduction 68, 252–261.

Drost M 2007. Advanced reproductive technology in the water buffalo. Theriogenology 68, 450–453.

Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T and Wilson A 2007. Geneious, v3.5. Retrieved April 2011 from http://www.geneious.com

Fair T, Carter F, Park S, Evans ACO and Lonergan P 2007. Global gene expression analysis during bovine oocyte in vitro maturation. Theriogenology 68, 91–97.

Gaska S, Pellestor F, Assou S, Loup V, Anahory T, Dechaud H, De Vos J and Hamamah S 2007. Identifying new human oocyte marker genes: a microarray approach. Reproductive Biomedicine Online 14, 175–183.

Ghanem N, Hölker M, Rings F, Jennen D, Tholen E, Sirard MA, Torner H, Kanitz W, Schellander K and Tesfaye D 2007. Alterations in transcript abundance of bovine oocytes recovered at growth and dominance phases of the first follicular wave. BMC Developmental Biology 7, 90–108.

Hennebold JD, Tanaka M, Saito J, Hanson BR and Adashi EY 2000. Ovary-selective genes I: the generation and characterization of an ovary-selective complementary deoxyribonucleic acid library. Endocrinology 141, 2725–2734.

Klatsky PC, Wessel GM and Carson SA 2010. Detection and quantification of mRNA in single human polar bodies: a minimally invasive test of gene expression during oogenesis. Molecular Human Reproduction 16, 938–943.

Kumar P, Verma A, Roy B, Rajput S, Ojha S, Anand S, Yadav P, Arora J, De S, Goswami SL and Datta TK 2011. Effect of varying glucose concentrations during in vitro maturation and embryo culture on efficiency of in vitro embryo production in buffalo. Reproduction in Domestic Animals 47, 269–273.

Luo S, Kleemann GA, Ashraf JM, Shaw WM and Murphy CT 2010. TGF-β and insulin signaling regulate reproductive aging via oocyte and germline quality maintenance. Cell 143, 299–312.

Nanda AS and Nakao T 2003. Role of buffalo in the socioeconomic development of rural Asia: Current status and future prospectus. Animal Science Journal 74, 443–455.

Nandi S, Raghu HM, Ravindranatha BM and Chauhan MS 2002. Production of buffalo (*Bubalus bubalis*) embryos in vitro: premises and promises. Reproduction in Domestic Animals 37, 65–74.

Parrish E, Siletz A, Xu M, Woodruff TK and Shea L 2011. Gene expression in mouse ovarian follicle development in vivo versus an ex vivo alginate culture system. Reproduction 142, 309–318.

Pauli A, Rinn JL and Schier AF 2011. Non-coding RNAs as regulators of embryogenesis. Nature Reviews Genetics 12, 136–149.

Robert C, Barnes FL, Hue I and Sirard MA 2000. Subtractive hybridization used to identify mRNA associated with the maturation of bovine oocytes. Molecular Reproduction and Development 57, 167–175.

Sahu BB and Shaw BP 2009. Isolation, identification and expression analysis of salt-induced genes in *Suaeda maritima*, a natural halophyte, using PCR-based suppression subtractive hybridization. BMC Plant Biology 9, 69–94.

Salilew-Wondim D, Rings F, Holker M, Gilles M, Jennen D, Tholen E, Havlicek V, Besenfelder U, Sukhorukov VL, Zimmermann U, Endter JM, Sirard M-A, Schellander K and Tesfaye D 2007. Dielectrophoretic behavior of in vitro-derived bovine metaphase II oocytes and zygotes and its relation to in vitro embryonic developmental competence and mRNA expression pattern. Reproduction 133, 931–946.

Sambrook J, Fritsch T and Maniatis T 1989. Molecular cloning: a laboratory manual, 2nd edition. Cold Spring Harbor Press, New York, NY, USA.

Singh B, Chauhan MS, Singla SK, Gautam SK, Verma V, Manik RS, Singh AK, Sodhi M and Mukesh M 2009. Reproductive biotechniques in buffaloes (*Bubalus bubalis*): status, prospects and challenges. Reproduction, Fertility and Development 21, 499–510.

Schuster SC 2008. Next-generation sequencing transforms today's biology. Nature Methods 5, 16–18.

Su YQ, Sugiura K, Woo Y, Wigglesworth K, Kamdar S, Affourtit J and Eppig JJ 2007. Selective degradation of transcripts during meiotic maturation of mouse oocytes. Developmental Biology 302, 104–117.

Tucker T, Marra M and Friedman JM 2009. Massively parallel sequencing: the next big thing in genetic medicine. American Journal of Human Genetics 85, 142–154.

Tomek W, Torner H and Kanitz W 2002. Comparative analysis of protein synthesis, transcription and cytoplasmic polyadenylation of mRNA during maturation of bovine oocytes in vitro. Reproduction in Domestic Animals 37, 86–91.

Wells D, Alfarawati S and Fragouli E 2008. Use of comprehensive chromosomal screening for embryo assessment: microarrays and CGH. Molecular Human Reproduction 14, 703–710.

Yamashita M, Mita K, Yoshida N and Kondo T 2000. Molecular mechanisms of the initiation of oocyte maturation: general and species-specific aspects. Progress in Cell Cycle Research 4, 115–129.

Zeng F and Schultz RM 2003. Gene expression in mouse oocytes and preimplantation embryos: use of suppression subtractive hybridization to identify oocyte and embryo-specific genes. Biology of Reproduction 68, 31–39.

Zuccotti M, Merico V, Cecconi S, Redi CA and Garagna S 2011. What does it take to make a developmentally competent mammalian egg? Human Reproduction Update 17, 525–540.